Rational Recency:

Does the Recency Effect Violate Bayes' Rule? *

STEVEN J. JORDAN†

Yale School of Management

February 7, 2007

Abstract

Based on experimental results from the 1970's psychological literature, De Bondt and Thaler (1985) claim: "It has now been well-established that Bayes' rule is not an apt characterization of how individuals actually respond to new data... In revising their beliefs, individuals tend to overweight recent information..." I analyze the conditions under which such a claim might be supported by the data. For financial markets, I show that the conditions necessary to support this claim require implausible parameter values.

JEL subject classifications: G12, G13, C14, G22

Key words: Bayes' rule, Recency Bias, behavioral finance, heuristics, noise, market efficiency, overconfidence, signal extraction.

*I would like to thank Zhiwu Chen, Catherine Eckel, Uzma Khan, Matthew Spiegel and Shyam Sunder for comments on an earlier draft of this paper.

†Ph.D. Candidate, Yale School of Management, Yale University, 135 Prospect Street, New Haven, CT 06520; email: `steven.jordan@yale.edu`.

1 Introduction

The field of finance does not exist without noise and uncertainty. Underlying all models is an assumption of a specific form of market efficiency, which states the mechanism of how information is embedded into prices. The standard Noisy Rational Expectations Equilibrium paradigm is that information is utilized in an optimal manner and it is the formation of expectations that sets prices in equilibrium. Within a noisy environment, to utilize information optimally requires some sort of signal extraction mechanism. The standard methodology employed is for an agent to apply Bayes' rule.

Given that market prices are at best noisy realizations of the true underlying value of the asset, Bayes' rule is one of the most widely used paradigms in Finance today.[1] This specific benchmark has proved tremendously useful in studying financial markets and developing important insights within a simple and tractable framework. Currently, there is increasing evidence that questions the relevance and applicability of Bayes' rule as a decision making model. Is Bayesian updating still a valid paradigm?

The answer is not obvious. Outside of finance, there is growing evidence to the contrary. Bayes' rule has been under direct attack for decades in the psychological literature. Tversky and Kahneman (1974) conducted extensive experiments to demonstrate the inadequacy of Bayesian updating. Their main conclusion is that agents do not use Bayes' rule, rather they use heuristics, i.e. simplifying rules, in order to make decisions in complex situations. The main ammunition underpinning these claims is that agents overweight the most recent data, termed the Recency Bias (RB). The main evidence that Bayes' rule is violated is the wide spread existence of RB in individual decision making situations. De Bondt and Thaler (1985, 1987) review this literature.

[1] A simple search on JSTOR under "Bayesian" matched 3,992 articles, while a search for one of the most common models, "random walk", retrieved only about 68% as many hits at 2,746 articles and the standard for market efficiency, "martingale", only had 24% as many hits at 989 articles.

More recent psychological research seriously challenges the methodology used in the heuristics literature and comes to very different conclusions.[2] Sedlmeier and Gigerenzar (2001) also make a credible defense of Bayes' rule. They present evidence that people (and even animals) do utilize a simplifying heuristic to gather and analyze data, however this simple heuristic approximates Bayes' rule. Their claim is that agents use frequencies as opposed to the probability distributions used in most psychological experiments. Jamal and Sunder (1996) use computer simulations to examine asset markets with imperfect information populated by either Bayesian or heuristic (representativeness and anchor-and-adjust) agents. All markets converge to the same Bayesian equilibrium even in the absence of profit maximization, natural selection, arbitrage, and mutual cancellation of random actions. They conclude that rationality of the market emerges

[2] My paper does not critique the psychological literature on base-rate fallacy, which underlies the claim in finance that Bayes' rule is an invalid paradigm for financial markets. However, there is an ongoing challenge within the psychological literature that calls into question the initial findings of a base-rate fallacy. Koehler (1996) reviews the base-rate fallacy literature and comes to the conclusion that this fallacy is not supported by the empirical facts and has been "oversold." Hertwig and Ortmann (2001) challenge the methodological procedures used in many of the heuristics studies, in particular they challenge the methodology used in the base-rate fallacy literature. They claim that the wide variance of results is due to lack of formal methodological protocol. Interestingly, they consider the subgroup of psychological studies that would pass the protocol utilized in economic experiments and find this subset of experiments, although small in number, support the diametrically opposite conclusions of the general base-rate usage literature. Another objection to the heuristics literature is discussed in Krueger and Funder (2004). They debate the validity of the negative perspective that pervades psychological research. Their argument is nostalgic of the data-snooping argument in finance and economics, in that only negative results are publishable and this sets the incentives to find and report such results. Finally, Gigerenzer (1999) points out that animals, like humans, use simplified heuristics to make decisions and, like humans, can be fooled within experimental design. However, in environments in which adaptation has likely occurred, that is in naturally occurring decision environments, the simple heuristics used closely approximate optimal Bayesian decisions. Along this line of reasoning, in environments that are stable enough for adaptation to occur, such as many financial settings, a Bayesian outcome may be a reasonable expectation.

as a consequence of the market structure, and not from the rationality of individuals. Jamal and Sunder (2001) show that in their simple markets an adaptive dividend adjustment process is sufficient to cause the markets with different agents to converge to the same equilibrium. Likewise, Sandroni (2005) demonstrates that in a general equilibrium setting both learnable and unlearnable assets are eventually priced according to the Bayesian paradigm. He concludes: "Asset prices will eventually be as close to the rational expectations ideal as permitted by the Bayesian updating of the available information."

It is healthy and necessary for a discipline to reexamine its foundations. Recently, many financial economists have utilized the early psychological findings of a RB to do just this. They claim that RB is exhibited by agents in financial markets as well. However, it seems difficult to reconcile the above two lines of literature. In finance, the underlying theoretical model is much better defined than in psychology or human cognition. Market efficiency, which is almost exclusively a martingale, is the underlying assumption. It is this difference that allows the financial economist to make stronger and more exact statements, at least with respect to accepted theory. I exploit this advantage.

Before discussing the details of these papers, one "obvious," but heretofore never explicitly stated, point should be noted first. There are environments when a good Bayesian would give greater weight to recent events, in fact to only the very last observation. For example, if there was no uncertainty (zero measurement error) and the process followed a random walk, then a Bayesian updater would give 100% weight to the most recent observation and 0% to past observations. Does this have anything to do with the RB? Clearly, as soon as new information is observed, prices would jump immediately to its new true value. Under perfect observability, price would follow a random walk. Thus, it is clear that with a random walk model of market efficieny that RB cannot occur. Any use of past data would instead "under-weight" the most recent data. This popular financial model of immediate and perfect adjustment to new information cannot be the model that the Bayesian critics claim to reject by observing decisions based mainly on a few past observations.

Obviously, real markets have measurement error or noise. De Long et al. (1989) summarize this literature and conclude "...that asset prices respond not only to news but also to ... 'noise trading'." That is, there is noise in prices. It is this modified model of market efficiency for prices, random walk (perfect observability) perturbed by noise, that the Bayesian critics refer to, as this process is a more realistic underlying process of a financial market. The question explored in this paper is:[3]

> Do financial markets contain enough noise to reject Bayes' rule as a valid heuristic based on the experimental literature's findings that agents weight heavily the most recent data observations?

To answer this question, I explore an illustrative model that explicitly accounts for noise in observed data. My model is a standard model. It has been used over the past 50 years to model financial markets, so there can be no claim of "model snooping." The structure of the model is simple enough that I am able to invert the variance-covariance matrix and thus I am able to obtain a closed-form solution. Within a Bayesian framework I use the closed form solution to solve for an agent's updating problem for a simplified economy where information is convolved with noise.[4] The closed form solution allows me to simulate the connection between market prices and agent expectations in order to determine whether observed market anomalies contradict or are consistent with the

[3]Although some papers refer to the property that agents use only the last few observations in their decision as RB, this is not accurate. There is no comparison with an optimal Bayesian update. To do so would require knowledge of: (1) the level of noise in the observed data, and (2) the agent's subjective probability distribution. Of course, neither is known and both are difficult to measure.

[4]I develop a simple random walk model of the unobserved state variable, e.g. the true underlying value of a stock. I explicitly model the effect that unobservability implies by modeling an observed proxy (fundamentals) to the true state variable. This proxy has noise (measurement error) superimposed on the true state variable and thus fundamentals (or observable data) follow a random walk with noise.

concept of market efficiency. For various levels of noise, I can determine how many past observations should bear heavily on the agent's decision.

My main result of this section is that, in the presence of realistic levels of noise, the updating solution emphasizes the most recent data observations.[5] Thus, empirically observed weights mostly on recent data is consistent with a Bayesian framework in a financial setting.[6] I demonstrate that one of two interpretations of my results are possible. The first interpretation is that market data consists mainly of noise. Under this interpretation, most anomaly studies would be due solely to chance and data snooping, and thus must be rejected as unreliable. This first interpretation gives strong support to the claims made in Fama (98). The second interpretation is that the empirically documented use of past returns is consistent with Bayesian updating. Given Bayesian agents heavily weight only the few most recent data observations, they should trade upon the arrival of new information. The empirical literature documenting the high correlation between recent data and agents' decisions does not support the claim that agents are heuristic and exhibit RB. By directly demonstrating the emphasis on recent data within the standard financial markets paradigm, I also demonstrate that it is not always feasible to directly transfer results from the less structured field of psychology to that of finance and economics. Results surmised from contrived experiments on individuals, where the structure is very different than that faced in real financial markets, may have little resemblance to what is actually observed in a very structured and complex market

[5]The fact that Bayes' rule heavily emphasizes recent data opens up the possibility to rationalize many of the documented anomalies in the financial literature. Jordan (2004a) demonstrates that a model similar to that used here can explain the excess volatility puzzle. Jordan (2004b) shows with a simplified version of this model that long-term reversals, termed the overreaction hypothesis, can be accounted for. Levy and Levy (forthcoming) argue that heavily weighting recent data has implications for mutual fund flows, momentum, and the fact that empirical tests of the CAPM have failed.

[6]Although not stressed in the paper, my results have implications for the literature on trading volume. In one of the few instances where the profession can agree, there seems to be a puzzle as to why trading volume is so high, see Odean (1999). Of course, this intuition is given with the caveat that there are no market frictions in my model.

setting. More to the point is that every model is wrong before the ink dries (otherwise it would not be a "model"). However, a model is judged by how well it predicts reality and on how useful it is in guiding intuition. Bayes' rule passes both these standard tests of usefulness in regards to the observed pattern of data usage in financial markets.

Two of the most widely used assumptions in economics and finance are Bayes' Law and market efficiency. Therefore, the results obtained in this paper may have implications for a wide range of problems studied in the literature. Much of the information literature uses the assumption that agents act as Bayesians. For example, all agents in Grossman (1976) and the vast subsequent literature are assumed to be Bayesian updaters. Today, the Bayesian paradigm is utilized to provide important insights in many different fields from asset allocation to monetary policy.[7] It seems too early to throw in the towel on Bayes' rule and so much of the insight and understanding it has led to in both finance and economics. In contrast to much of the behavioral literature, I find that Bayes rule is still an interesting and valid model. I also find that much of the insight gained in poorly understood environments common in psychology are not necessarily applicable in models with well understood assumptions, as is typically the case for financial models.

The paper proceeds as follows. In section 2, I review the current literature on the RB and some implications of the above results are discussed. In section 3 the assumptions made and the model set up are reviewed. This is followed, in section 4, by the solution to the signal extraction problem. Simulations are performed in section 5 and it is shown that emphasizing recent data does not contradict the use of Bayes' rule. I point out the causal relationship between overconfidence and RB in section 6. In section 7, I make some simple estimates from observed data to calibrate the random walk with noise model

[7]Shanken and Tamayo (2001) use a Bayesian methodology to extend the work on asset allocation and dividend yield by Kandel and Stambaugh (1996) to accommodate variation in risk as well as expected return. Feldstein (2003) discusses the nature of the uncertainty faced by central banks and considers the asymmetric nature of these risks utilizing a Bayesian approach.

and derive some estimates of the level of noise to information in market data. These estimates act as lower bounds on the level of noise that must exist in the data for these empirical observations to be truly classified as anomalies. Finally, I conclude and make some closing remarks in section 8.

2 Recency Bias - An Overview

Tversky and Kahneman (1974) document the RB and representative heuristics in detail. They argue that cognitive limitations force people to use heuristics to make complex calculations and decisions. Shiller (2002) makes a strong case that attention is also a fundamental aspect of human intelligence. Limitations in attention lead naturally to a RB. De Bondt and Thaler (1985, 1987) argue that overreaction and systematic price reversals lend strong support for RB and are inconsistent with Bayes' rule.[8]

Short-term time series extrapolation is used in models of positive-feedback trading (Shleifer and Summers (1990) and De Long, et al (1990)). Positive-feedback trading is a trend following strategy that emphasizes recent observations. Shefrin and Statman (1994) model some traders as true Bayesians and others as non-Bayesians in that they "...form forecasts by overweighting recent events and underweighting more distant events..."

Sirri and Tufano (1998) provide evidence that flows into mutual funds are concentrated among those funds that have recently had extraordinarily high performance.

[8] It is important to note here that some papers confuse RB with price reversals. However, the name "Recency" Bias clearly identifies its meaning, that is people's observed actions or decisions seem to be driven mostly by the last few "recent" observations in violation of the standard optimal decision, which is taken to be Bayesian updating. Note that neither price continuation nor price reversals necessarily contradict Bayes' rule. Either result can obtain depending on whether the agent's prior considers ranking or sorting. Bayes' rule cannot be rejected based on a claim that the agent uses one prior verse another prior.

Karceski (2002) shows that recent market returns also have a large economic impact on subsequent aggregate mutual fund flows. Brown, Harlow and Starks (1996) and Chevalier and Ellison (1997) find that mutual funds alter their risk taking behavior in response to this flow performance relationship. Consistent with the mutual fund evidence, Gruber (1996) finds that individual investors chase mutual fund performance through time based on the most recent past returns.[9]

Technical trading is a popular technique on Wall Street and with individual investors, as evidenced by the quantity of related news and published books. Technical trading uses recent past data, in the form of charts, to forecast future price movements. Experimentally, Andreassen and Kraus (1988) verify this tendency by documenting that people see the same trends and patterns in artificially generated data and in real stock price data. Other trading strategies widely used that are based (at least implicitly) on the expectation of recent trends to continue are stop-loss orders and range-break-out strategies.

IPOs tend to trade for high returns when they first open, subsequently returns mean revert. Derrien and Womack (2003) document that these initial IPO returns depend on the past 3 months of the market return. Shiller and Pound (1989) find that, consistent with the RB heuristic, institutional investors are also attracted to recent past price increases. Frankel and Froot (1990) find that professional forecasters of exchange rates expect recent price trends to continue in the short run even when they expect a long run price reversal. Thus, claims of RB have been empirically documented across all classes of investors, both individual and institutional. This body of evidence is generally interpreted as corroborating the conclusion that Bayes' rule is an inadequate model of investors' decision making process.

[9]Sirri and Tufano (1998) suggest that the relation between fund flows and returns are related to advertising activities. However, Gallaher, Kaniel and Starks (2005) show that the relation between advertising and fund flows is independent of the relation between past performance and fund flows.

3 Model Setup

In the previous section, most of the literature documenting the relation between decisions and recent observations in financial markets either (1) utilized mental experiments of posing past prices and soliciting future expectations or (2) utilized regression techniques. Both methods take the past decision variable as given or exogenous and correlate future decisions or actions with the assumed exogenous or independent variables. In order to adequately model these empirical studies in a theoretical framework, my model explicitly assumes the use of an exogenous decision variable.

1. *The decision variable.* My initial model assumes a single decision variable, which is dynamically observed over time. I limit the decision process to one variable for simplicity. As in the empirical literature, I assume the decision variable is given exogenously. This assumption implies that the decision maker is a price taker. In order to facilitate the intuition of the setup, I will interchangeably refer to the decision variable as the observed or realized log market price of a firm and I will refer to the true underlying state variable as the true log value of the firm. However, the analysis holds for any Bayesian agent who must use a noisy proxy for a state variable that itself follows a random walk.

2. *A Bayesian agent.* There is an agent who is Bayesian by assumption. The agent observes the realizations of the observed decision variable (the log market price of the firm), and forms expectations for the true state variable (the true underlying log value of the firm). The agent maximizes utility, or equivalently minimizes a loss function, to form optimal expectations.[10]

3. *Observed and unobserved variables.* To motivate the specific model I use for the

[10] Simon (1956) shows that expectations are optimal when "certainty equivalence" applies, which applies in discrete time when the objective function is quadratic and the process is linear in the unobservable state variables. Both conditions are satisfied in my setup.

exogenous decision variable (e.g. past log prices), I only need to ascertain from the reader that observed market data is imprecise (or contains a temporary component.) I assume there is a true state value for every time period. If we lived in a perfect information world, then we would observe this true state value to be a random walk, that is, the true state would adjust immediately and instantaneously to its new true value upon the arrival of perfectly observed new information. The fact that we cannot observe this true state motivates the addition of a noise or temporary component to the true state value in order to model the observed proxy (e.g. the observed log market price).

I assume two types of innovations. The permanent innovation v_t drives the innovation of the true underlying state variable which follows a random walk. In contrast, the temporary innovation, or measurement error, e_t represents the error that is contained in the observed decision variable. The state variable is the true log value of the risky asset, s_t, which is *unobserved* or imperfectly observed. The agent only sees a noisy proxy x_t, the log price of the asset, and infers the state variable from the realization of the current and past changes in the observed decision variable (changes in log price). The temporary shock, which generates the measurement error in the state variable, is assumed to persist for one period. The following system describes the model:

$$
\begin{aligned}
s_t &= s_{t-1} + v_t \\
x_t &= s_t + e_t \\
v_t &\sim N(0, \sigma_v^2) \\
e_t &\sim N(0, \sigma_e^2).
\end{aligned}
\tag{1}
$$

The measurement error and information innovation are assumed independent, this implies that $\text{cov}(v_t, e_\tau) = 0 \;\forall\; t, \tau$. Note that the measurement error, e_t, confounds the price discovery because agents cannot separate mispricing from information at each point in time.[11]

[11] The random walk with noise model is an established model in the finance literature. Muth (1960)

4. *Information.* The information set, I_t includes observed variable, e.g. log prices, $x_t, x_{t-1}, x_{t-2}, \ldots$, which is a noisy proxy for the true variable. Equivalently within a scale factor, the agent can use observed changes $\Delta x_t, \Delta x_{t-1}, \Delta x_{t-2}, \ldots$

5. *Forecast of the state variable.* The central problem is to study the evolution of the expectation of the state variable, $E[s_{t+1}|I_t]$. I will show how a Bayesian agent updates this conditional expectation upon the arrival of a new observation. This optimal update exhibits an endogenous "recency" effect, in that only a few recent data points are weighted heavily.

4 The Bayesian Updating Solution

4.1 Variance-Covariance Matrix and its Inverse

Before I begin to discuss the model and its implications, there is a small but important detail that needs to be addressed first. For all asset pricing models the variance-covariance matrix ultimately comes under consideration. What is more important is that in order to perform a Bayesian update, inverting the variance-covariance matrix is necessary. Typically this is an immensely difficult task if complete generality is desired. In the above model, we are fortunate to have some structure that makes this daunting task doable. The inversion is still difficult and involved. So as not to ruin the continuity or the intuition in the following sections, I address invertability in Appendix A.

Since I have an explicit solution to the inverse of the variance-covariance matrix, it is now possible to solve the updating problem explicitly. As shown in Appendix A, the inverse variance-covariance matrix can be summarized by the following functional form

popularized the model, while slight variations have been used in various empirical studies, e.g. Summers (1986), Fama and French (1988) and Conrad, Gultekin and Kaul (1997).

(see equation (A.2) and equation (A.3)):

$$f_t(n) = \sum_{k=0}^{n} \binom{n+1+k}{n-k} \sigma_e^{2(t-1-k)} \sigma_v^{2k} \tag{2}$$

where the i^{th}, j^{th} element of the inverse variance-covariance matrix, Σ_t^{-1}, can be written as:

$$\Sigma_t^{-1}(i,j) = \sum_{k=0}^{t-i} f_t(2k + j - 1 - t + i)$$

It is worth pointing out at this time the structure of the function $f(n)$ is similar to a binomial polynomial in terms of σ_v^2 and σ_e^2. This structure is embedded in the inverse variance-corvariance matrix Σ_t^{-1} by $f(n)$ and this binomial structure has important implications for the behavioral traits that a Bayesian agent will exhibit.

4.2 Information Extraction

The next theorem answers the question of how a Bayesian agent would extract information from the observed noisy proxy, e.g. log market price, and how a Bayesian would update expectations of the state variable, e.g. the true underlying log value of a firm, over time. It is important to note that when conducting an experiment that any observed action by the agent is driven by the update to the prior distribution. Any information already incorporated into the prior will not influence an agents observed actions nor will it change the variable of interest. Thus, the following theorem determines the change in expectations upon the arrival of new information. It is this change in expectations that is related to an agent's observed action or observed response to new information.

Theorem 1 *For a quadratic loss function and the economy with measurement error represented by equation (1), the optimal Bayes' update is given by:*

$$h(I_t) = E[s_{t+1}|I_t] - E[s_t|I_{t-1}] \tag{3}$$

where $I_t = \mathbf{\Delta x}_t = (\Delta x_1, \Delta x_2, \ldots, \Delta x_t)'$ is the vector of changes in the observed variable up to time t. $h(I_t)$ is the Bayesian update to the time-$(t-1)$ expectation of the state *parameter, s_t, on the arrival of a new observation, Δx_t. The optimal update function, $h(I_t)$ is given by:*

$$h(I_t) = k\,\mathbf{w}_t^{'} \cdot \mathbf{z}_t \tag{4}$$

where $\mathbf{w}_t$ is a weighting vector given by:

$$\mathbf{w}_t^{'} = [(\sigma_e^2)^{t-1},\, (\sigma_e^2)^{t-2}(\sigma_v^2),\, (\sigma_e^2)^{t-3}(\sigma_v^2)^2, \cdots, (\sigma_e^2)(\sigma_v^2)^{t-2},\, (\sigma_v^2)^{t-1}] \tag{5}$$

$\mathbf{z}_t = \mathbf{z}_t(I_t)$ is the vector function of past observations, which is given by:

$$\mathbf{z}_t = \begin{bmatrix} \binom{1}{0}\Delta x_1 + \binom{2}{1}\Delta x_2 + \binom{3}{2}\Delta x_3 + \cdots + \binom{t-1}{t-2}\Delta x_{t-1} + \binom{t}{t-1}\Delta x_t \\ \binom{3}{0}\Delta x_2 + \binom{4}{1}\Delta x_3 + \cdots + \binom{t}{t-3}\Delta x_{t-1} + \binom{t+1}{t-2}\Delta x_t \\ \binom{5}{0}\Delta x_3 + \cdots + \binom{t+1}{t-4}\Delta x_{t-1} + \binom{t+2}{t-3}\Delta x_t \\ \ddots \\ \binom{2t-3}{0}\Delta x_{t-1} + \binom{2t-2}{1}\Delta x_t \\ \binom{2t-1}{0}\Delta x_t \end{bmatrix} \tag{6}$$

and k is a constant given by:

$$k = \frac{\sigma_v^2}{f_{t+1}(t) - \sigma_e^2 f_t(t-1) + \sigma_e^{2t}}$$

Proof of Theorem 1[12]

See Appendix B.

[12]The use of the squared loss function is justifiable for several reasons. First, the empirically observed use of recent data is documented predominately utilizing least squares regressions. It makes intuitive sense that a theoretical model that aims to explain the empirically observed regularities should make similar assumptions. On a more theoretical footing (see Berger (1993)), squared-error loss occurs naturally in certain settings. For example, if the agent is risk-neutral, the state of the economy is perceived to be uniformly distributed (i.e. no good predictive rules), and the variable of interest depends on the action taken in a linear manner, then the decision maker will exhibit a squared error

The signal extraction problem is well studied in many fields. Muth (1960) is probably the most noted reference. However, my proof is the first finite time horizon solution to my knowledge. There are two main results derived from this theorem. The first result is the exponential nature of decay for the weights applied to past data typical of the optimal signal extraction. This is not a new insight, but it is comforting to find that the decay occurs quite quickly and is a strong property of the finite time optimal solution as well. The second result is specific to the finite time solution. I am able to decompose the decay structure of the weights applied to past data into two main components, each reflecting a different cause for the heavy emphasis on recent data inherent to Bayesian updating. The first component is due to the noise level of observed data. This is reflected in the weighting vector, $\mathbf{w}_t$, of Theorem 1. The more noise there is in observed data, the more weight will be placed on past data. The second component reflects the inherent "recency" effect of the linear forecasting rules that constitute an optimal Bayes' update. This is reflected in the $\mathbf{z}_t$-vector of Theorem 1. The mere fact that an observation occurs more recently implies these linear forecast rules will weight it more heavily. This decomposition is new and it is explored in detail next.

The first attribute of the update function $h(I_t)$ is that there is a weighting vector $\mathbf{w}_t$. This weighting vector is a binomial product of the two variance terms. σ_v^2, the permanent shock variance, represents the accuracy of information while σ_e^2, the temporary shock variance, represents the degree of measurement error or uncertainty. The ratio, $q = S/N = \sigma_v^2/\sigma_e^2$, is the signal to noise ratio for the economy. The higher S/N is the more informative the observed signals. An important symmetry exists in $\mathbf{w}_t$. For a higher S/N economy, the elements in $\mathbf{w}_t$ will shift the weight of an agent's decision more and more

loss. Another justification can be made locally. If the decision is a local decision, then the error of the forecast will be small on average. A second order Taylor expansion of the true utility function will yield a squared error loss. Although not perfect, a squared error loss is tractable and does seem consistent with the empirical methodologies used in this literature. Other commonly used techniques found in the finance and economics literature, such as the Kalman Filter, also depend critically on the squared-loss assumption.

toward the later elements of $\mathbf{z}_t$. In the limit, for $S/N \to \infty$, the weighting vector weights the last observation in the information set infinitely more than all other observations. Thus, the more informative a signal is or the more informative an agent *thinks* a signal is, as prices are a function of subjective beliefs, the more the weighting vector will weight those elements of $\mathbf{z}_t$ that are functions of the most recent data. In the limit of large S/N, the agent will essentially only consider the last observation in the information set, Δx_t.

The second attribute of the update function $h(I_t)$ that is worthy of note is the pattern of the coefficients in the sums that constitute each element of $\mathbf{z}_t$. Each element, $z_{t,i}$, is a sum of the past $n-(i-1)$ observations. As will be proved below, the coefficients of each term in the sum grows with the recency of the observation. Thus in each row of $\mathbf{z}_t$ the most recent observation has the largest weighting.

An important attribute of the $t \times 1$ update vector $\mathbf{z}_t$ is that each element (row) is a linear forecasting rule over past observations in the information set. Each element, $z_{t,i}$, is a sum of past observations. If $m > n$, then row $z_{t,m}$ will use fewer past observations than row $z_{t,n}$. The observations used in each row are the most recent $t-(m-1)$ observations for row m and the most $t-(n-1)$ observations for row n. As will be proved below, the coefficients of each term in the sum grows with the recency of the observation. Thus in each row of $\mathbf{z}_t$ the most recent observation has the largest weighting. Summing all update rows with appropriate weights from $\mathbf{w}_t$ creates a weighted average of all these linear forecasts.

The fact that a Bayesian agent in my model uses any observation other than the most recent observation is strictly due to the fact that noise is present in the economy. This is easily seen in either of two ways. The easiest way is to consider the weighting vector, $\mathbf{w}_t$. If $\sigma_e^2 \to 0$, then $\mathbf{w}_t \to [0, 0, \ldots, 0, 1]\sigma_v^{2(t-1)}$. Thus, from equation (2), equation (4) and equation (B.4), $h(I_t) \to \Delta x_t$. This also can be seen from equations (A.1), (B.5) and (3). In the absence of noise, the variance-covariance matrix reduces to, $\Sigma_t = \sigma_v^2 I$, where

I is the identity matrix. The inverse is now trivially $\Sigma_t^{-1} = \sigma_v^{-2} I$. Now equation (B.5) reduces to $E[s_{t+1}|I_t] = \sum_{j=0}^{t} \Delta x_j$. Plugging this into equation (3), we again obtain our result.

As pointed out in footnote 2 in the introduction, if the benchmark model is a Bayesian model under a random walk, then the use of the description "over", as in overweight the most recent data, is inaccurate. In the absence of noise, any use of observed data other than using the most recent data actually underweights the use of the last observation and overweights any other past data that is used in the updating algorithm. With the presence of noise, that is under a random walk with noise economy, it is possible for a Bayesian to project onto past data. However, as I shown next, the preponderance of the weight is still on relatively few data points.

5 Recency & Bayes' Rule

5.1 Using a Proxy with Bayes' Rule

Theorem 2 *For a quadratic loss function and the economy with measurement error represented by equation (1), a Bayesian agent will make decisions that will heavily emphasize the most recent data for any reasonable level of noise in the market.*

Proof of Theorem 2

See Appendix C.

By ignoring the weighting vector $\mathbf{w}_t$ we can make very specific statements about the relative amount of use a Bayesian agent will make of respective data in the information set. As shown in the noise estimation section below, the calculations here are probably lower bounds on the weight put on recent data and thus on the conservative side. Real agents most likely weight recent data points more heavily than is proved in the above theorem.[13]

To gain some intuition and feel for the degree to which recent data is emphasized, it is useful to work through a few examples.

Example 1 I first compare the adjacent coefficients in a specific row of $\mathbf{z}_t$ and show that the weighting of any new data increases over the weight of any prior data, including the most highly weighted of the prior data. First note, in the vector $\mathbf{z}_t$, the coefficient of the j^{th}-row linear forecast rule and the k^{th}-information observation, Δx_k, within this j^{th}-row is:

$$c_{j,k} = \binom{k-1+j}{k-j} \tag{7}$$

Note that by the structure of the updating vector $\mathbf{z}_t$, $k > j$ always. Therefore, the coefficient in each row grows in magnitude across adjacent columns as:

$$\Delta c_{j,k} = c_{j,k+1} - c_{j,k} = \binom{k+j}{k+1-j} - \binom{k-1+j}{k-j} \tag{8}$$

A little algebra reduces this to:

$$\Delta c_{j,k} = \frac{2j-1}{k+1-j}\binom{k-1+j}{k-j} \tag{9}$$

Now it is obvious that the second term of the product on the right is greater than one as it is a binomial coefficient. The first term of the product is positive as $2j > 1$ always and by construction $k > j$ thus $k + 1 - j > 0$. Thus $\Delta c_{j,k} > 0$ as claimed. The coefficients in the Bayesian updating vector increase with the recency of the observation. As this result holds for each row of $\mathbf{z}_t$, we can conclude that each new observation is universally given the most weight by a Bayesian, irrespective of the value of $\mathbf{w}_t$.

[13] The weighting of the most recent observed data holds for any finite observation set and holds in the limit as an asymptotic result. The asymptotic result utilizes the tail of an infinite sequence, which is not directly applicable for our purposes. However, this result is useful as it allows us to see how the weights change as new data arrives.

Example 2 It is worth highlighting a preliminary result that is proved in the derivation of Theorem 2. For any finite time t, the sum of all weights for the k^{th} observation is just the sum of all the coefficients for the k^{th} observation in each row of $\mathbf{z}_t$. This sum can be written:[14]

$$\omega_{t,k} = \sum_{j=1}^{k} c_{j,k} = \sum_{j=1}^{k} \binom{k-1+j}{k-j} \tag{10}$$

It was shown that $\omega_{t,k+1}/\omega_{t,k} > 1$ for any finite k. This shows that the weight on the $(k+1)^{st}$ observation is increasing at a faster rate than any previous observation, including the k^{th} observation. This can be written in a more intuitive form as, $\omega_{t,k+1} > \omega_{t,k}$, which implies the same thing. Weighting recent data more heavily is prevalent in the Bayesian update formula.

Similar to example 2, we can also compare the k^{th} coefficient with the $(k+n)^{th}$ coefficient. As shown in the proof of theorem 2, this ratio, $\omega_{t,k+n}/\omega_{t,k} \to \infty$ as $n \to \infty$. This last result is an asymptotic result. However, as I show below the weight on recent data exhibited by a Bayesian agent is very strong in finite samples as well. In fact, the convergence of the above result is extremely fast.

5.2 Simulations

As the result in Theorem 2 is an asymptotic result, it is interesting to investigate the finite sample properties of the Bayesian update solution. I simulate the weights on the respective data for a range of reasonable past horizon lengths and for reasonable values of the signal-to-noise ratio, $q = \sigma_v^2/\sigma_e^2$. I assume that monthly data is used and calculate the how many past data points are given the bulk of the decision weighting for forecasts based on histories of 12, 60, 240 and 480 months. Results for longer periods of data are

[14] When $q \neq 1$, $\omega_{t,k} = \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}$.

shown in order to demonstrate the robustness of the results. I use the following values of the signal-to-noise ratio: $q = \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, \frac{3}{2}, 2$. The level of noise in a market is an empirical issue that seems to have attracted little attention in the academic literature, most likely due to intractability.

I simulate the updating coefficients for each observation that a Bayesian agent facing measurement error will use. Equation (10) is the coefficient equation for each observation's total weighting. I then calculate the ratio of the weights in the last N observations to the total weights on all observations at time-t. The lowest number N such that $P(N) > 0.50$ for the 50% number and such that $P(N) > 0.90$ for the 90% number is calculated. The calculated number shown in the tables is the number of observations that accounts for 50% and 90% of the total weight respectively for a Bayesian agent. The ratio of the last N observation weights to the total weight for the whole history of observations is given by:

$$P(N) = \frac{\sum_{k=t-N+1}^{t} \omega_{t,k}}{\sum_{k=1}^{t} \omega_{t,k}} = \frac{\sum_{k=t-N+1}^{t} \sum_{j=1}^{k} c_{j,k}}{\sum_{k=1}^{t} \sum_{j=1}^{k} c_{j,k}} = \frac{\sum_{k=t-N+1}^{t} \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}}{\sum_{k=1}^{t} \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}}$$

The lowest number of observations necessary to account for $100x\%$ of the total weighting used in the Bayesian update vector is given by:

$$P(N) > x \qquad \text{where } x = 0.50,\ 0.90$$

Table A summarizes the findings for the benchmark case of $q = 1$ and clearly demonstrates the strong emphasis on recent data present in the optimal Bayesian updating formula.

[**Table A about here**]

The numbers in Table A are simulations of the total number of observations that are used to account for 50% and 90%, respectively, of the total weights on all the observations in the information set. As the table demonstrates, a Bayesian making a decision with price uncertainty heavily emphasizes recent data and uses very few data points to make the bulk of the decision. For $q = 1$ and using 480 observation periods (e.g. 40 years) of historic data to forecast, the Bayesian only weighs two monthly observations with more than 90% of the total wieghting. This table is a vivid demonstration of how fast the asymptotic convergence in Theorem 2 really is.

The first set of numbers in Table A ($q = 1$) are simulated ignoring the weighting vector $\mathbf{w}_t$. As $\mathbf{w}_t$ can be written as a vector of powers in terms of the signal-to-noise ratio q (see Equation (D.1)), the N most recent data points that account for $100x\%$ of the total decision wieghts calculated in Table A could be quite different. If $q << 1$, i.e. the market is extremely noisy, then heavy weighting on only a few past observations will be mitigated. The question is how much will noise mitigate the heavy emphasis on only the most recent data that is inherent in the Bayesian updating paradigm?

Table B calculates N for high noise economies, i.e. for $q = 1/16, 1/8, 1/4, 1/2$. A value of $q = 1/16$ demonstrates the effect of substantial noise in a market on a Bayesian's weighting of past data. For this example, the overall market's subjective estimate of the level of noise in observed data is 16 times that of the market's subjective estimate of the level of information. This is indeed a very noisy market. Yet, in this case, only 7 to 9 of the most recent data points account for more than 90% of the total decision weights. For an economy with a subjective level of $q = 1/4$, which still constitutes high levels of noise, a mere 4 recent data points are used in a Bayesian updating decision. Table B again vividly demonstrates the strong and robust emphasis on only the most recent data inherent in the Bayesian framework when only a noisy proxy of the true state variable is available.

[Table B about here]

One feature is interesting to point out at this time. For very short historical data sets, the number of points used in Bayesian updating is sensitive to acquiring more data. However, as is apparent in all three tables, once a minimum level of historical data is available, then the decision of a Bayesian becomes rather robust to lengthening the data set further. It seems for any reasonable parameterization of an economy's noise and information content that no more than 1 year of data and certainly no more than 5 years of data are required to obtain near optimal estimates. Even if the large literature that documents the possibility of time-varying distributions for economic variables is ignored, the results in this paper imply that only a short data set is necessary for a Bayesian constrained by the accuracy of the observables. Thus, large data sets are not necessarily going to yield substantial improvements over estimates that are made using 60 observations.[15]

If $q \cong 1$, the above results, i.e. an optimal Bayes' forecast weighting heavily only the most recent observations, in Table A are accurate. In fact, for a noisy economy with a reasonable value of the signal-to-noise ratio, $q < 1$, heavily weighting recent data is also predicted. The simulations show that the mitigating effects of noise are not substantial. While the level of noise grows exponentially, the number of relevant data points, N, grows slightly faster than linearly. Thus, there would have to be unreasonable subjective levels of noise to that of information for the emphasis on recent data exhibited in the tables to disappear.

If $q > 1$, say because the agent is overconfident or because the observed data is informative, then the heavy weight on recent data is enhanced. Given the above results, an overconfident Bayesian agent or a Bayesian agent using data that is informative will only use the very last observation. Thus, if a Bayesian agent is overconfident, then RB cannot be observed. In fact, Table C shows that it takes very little enhancement of the perceived quality of information in observations to make the last observed data point

[15] In the context of stock market predictability, the results of the noisy Bayesian model imply that the search for better results via collection of longer and longer data sets may in fact be a quest for the Fountain of Youth.

dominate the decision process. It must have been a Bayesian who coined the phrase: "A little information can go a long way."

[Table C about here]

One thing that all three tables indicate is that by relying on so few observations, a Bayesian with noisy observations or proxies will update and change predictions (and thus decisions) frequently. Unless the data generating process contains significant noise levels, then positions should be updated frequently as well.[16] Thus, if investors are Bayesian and if market prices are noisy proxies of the true underlying asset value (both reasonable assumptions), then investors should trade frequently. More importantly from the perspective of the RB debate, a Bayesian agent's observed actions should be mainly driven by the last few observations.

5.3 Bayes' Rule and the Recency Bias – Discussion

In simple, single-agent, static environments, Tversky and Kahneman (1974) show that base-rates are not adequately used. Although this claim is being reconsidered in the psychological literature, this view has been adopted as fact in the economics and finance literature. De Bondt and Thaler (1985) ask: "Does the stock market overreact?" They present some empirics that overreaction exists. They claim that Bayes' rule is not an apt characterization of how individuals actually respond to new data. The reasons given are: "In revising their beliefs, individuals tend to overweight recent information and underweight prior (or base rate) data." As I demonstrate in Theorem 2, this claim is not true and the fault lies in the fact that the level of noise is not credibly high enough in order to overcome the inherent heavy weighting of recent data that exists in a Bayesian updating solution. What I show in this paper is that it is not enough to document the

[16] Of course, transaction costs will mitigate the degree to which this statement is true.

number of data points that affect an agent's decision to prove that RB exists. All this gives is a lower bound on the level of noise in the data. To prove RB, it is also necessary to document the true level of noise.[17]

Shefrin and Statman (1994) model some traders as true Bayesians and others as non-Bayesians in that they weight recent observations too heavily. The financial empirical literature does not support the claim that agents "over-weight" recent data. This literature only confirms that recent data is weighted heavily. Again, this is precisely how Bayesian agents who face price uncertainty behave. Thus I find the assumption of poor- or non-Bayesian agents not supported by the experimental observations that decision makers utilize only recent data. Following De Bondt and Thaler (1987), who argue that RB leads to overreaction, it also appears that it is possible for a rational Bayesian agent to exhibit "overreaction" as well.

The anti-Bayesian literature has even started taking some of these psychological observations as a definition of Bayesian decision making, which of course makes the definition self-fulfilling. A definition in Daniel and Titman (1999) states that Bayes' rule "specifies that the weights placed on the different pieces of information should be proportional to their respective precision." As demonstrated above, Bayes' rule is not so simplistic. It is a weighted average of past data or more precisely a weighted average of a set of linear projection models. These weights rely on several factors that are not specifically related to the precision of the data. Other important determining factors for how specific observations are weighted are: (1) the recency of the data and (2) the relative informative properties of the overall market.

To conclude, when it is necessary to use a proxy in place of the true state variable

[17] Of course, within the Bayesian framework, the assumption of overconfidence by De Bondt and Thaler will make the Bayesian model an even more appropriate model for agents who emphasize only a few data points, as what little effect noise could have extracted will be systematically ignored. This also makes clear that the subjective probability of the agent must also be measured and compared with the objective probability before any claims that RB exist can be made. (see Section 6).

of interest, Bayes' rule stands the test of much of the empirically observed behavior of decision makers. Relying mostly on a few data observations and overreaction are consequences of the use of Bayes' rule, as implausible levels of noise would have to exist in observed market data to significantly change these conclusions. As shown below, if overconfidence exists, as much the same behavioral literature strongly claims, then it is certain that any decision maker using Bayesian updating is going to exhibit strong "recency" effects.

6 Overconfidence and Recency

Given the vast theoretical and empirical evidence that agents are overconfident, it is interesting to explore how such overconfidence affects the fact that agents utilizing Bayes' rule utilize recent data when unobservability is present. Robustness checks of results to the existence of other possible effects is a very desirable property for any theoretical model.

The weighting vector, $\mathbf{w}_t$, in the Bayesian updating formula has an important influence on the observed recency of data used. Specifically, under the assumption of overconfidence, we can determine very definite directional effects on the weights. Thus, it is possible to be precise about how the assumption of overconfidence will impact on a Bayesian decision maker's distribution of weights over past observations.

The following theorem summarizes the relevant findings.

Theorem 3 *For a quadratic loss function and the economy with measurement error represented by equation (1), an overconfident Bayesian agent will continue to make decisions that heavily weight the most recent data. An overconfident Bayesian will exhibit an enhanced use of recent data compared to a rational Bayesian. Extreme overconfidence implies that the latest or most recent observation will dominate all observable decisions* *by the agent.*

Remark 1 The most important insights to be gained from this theorem are (1) under the assumption of overconfidence, the solution of any theoretical model should also exhibit extreme weighting of recent data to be credible, (2) if a model assumes overconfidence, then the credible degrees of freedom on the assumptions concerning how agents use information is reduced to one degree, that is agents will utilize less data than if the agent were not overconfident, and (3) if agents are overconfident and an experimenter or empiricist ignores this fact, then the observed results will appear to exhibit RB. That is, RB is a consequence of overconfidence and thus is not an independent heuristic.

Proof of Theorem 3

See Appendix D.

In my model, overconfidence manifests itself by the agent misjudging observed data to be more accurate than it really is. This can easily be represented mathematically as the signal-to-noise ratio, q, being perceived to be larger than it actually is. If the market is perceived to be informative, then it is reasonable to assume that the perceived magnitude of the variability in noise is smaller. Thus, it is reasonable to assume that $q = \sigma_v^2/\sigma_e^2$ is greater than reality and strong overconfidence should quickly make $q > 1$. Table C demonstrates the enhancement of the recency effect when the Bayesian decision maker is overconfident, i.e. when $q > 1$.

If a market has measurement error, as is true for any market where a noisy proxy must be used for the true underlying state variable, then a Bayesian agent will tend to be reactive to new information. This follows directly from the heavy weight placed on recent observations exhibited by the updating formula. If an agent makes decisions based primarily on only the last few observations, then his decisions will change whenever new data is received. There is little to no smoothing of information by a Bayesian who is forced to make decisions even if substantial noise exists in the proxy used for the true state variable. Overconfidence only strengthens these conclusions.

7 Aggregate Market Signal-to-Noise Ratio Estimate

Let's state what we know as a profession. First, observed market data, including returns, is noisy. Second, a decision maker in a noisy environment will desire to extract information from observables. Whether optimal or heuristic, past data will be useful.

To put the anomaly literature and the endogenous "recency" effect inherent in Bayes' rule into perspective, we will consider two anomalies. IPOs tend to trade for high returns when they first open, subsequently returns mean revert. Derrien and Womack (2003) show that the initial returns on IPOs in France over the 1992 to 1998 range were predictable using the prior 3 months market returns. As returns reflect expectations of the market, on average, agents utilized the 3 most recent past returns in forming expectations. De Bondt (1985) suggests that agents, by concentrating their decision weights on the 3 most recent observations, utilize too few observations and exhibit RB, in that past returns are weighted too little. As many of the empirically documented use of the most recent past data is in the mutual fund literature, we take this as our second example. Although most funds flow studies utilize the prior years returns, Warther (1995) investigates fund flows with respect to past monthly returns. He finds statistically and economically significant use of the prior three months returns. However, returns due to months prior to the first three are not significant. Both studies could be used to claim that agents' decision making process violate market efficiency and optimality. Neither study controls for the level of noise in the observed data, nor is there any control for agents' subjective probabilities. With my model, we can consider both these studies within a unifying framework and see that these studies report consistent results.

The idea behind the analysis is simple. First, to set up a benchmark case, we assume the observed use of the data is optimal according to Bayes' rule. If we define the relevant range of data for making decisions as those points that contain at least 90% of the decision weight, then the above calculations allow us to estimate the aggregate market subjective signal-to-noise ratio, q_M^*. Once we can identify how many past observations an optimal agent weighs heavily, we can use the simulation results to back out the level of noise in the data, assuming the agent is Bayesian. For the claimed anomaly to truly be RB, it is then necessary for the level of noise in the data to be greater than we estimated under optimality. That is, we can determine an upper bound on the signal-to-noise ratio in the market data for RB to be credible. Finally, we need to make a value judgment whether lower signal-to-noise ratios (or higher levels of noise) than our upper bound (given by our estimate under optimality) are reasonable or not. Depending on the reader's value judgment, RB will be validated or dismissed. Ultimately, it is an empirical issue as to how much noise is in the data and how overconfident an agent is. The less noise in the data and the more overconfidence that is assumed, the less credible is the claim that RB exists.

Applying the above idea to back out the signal-to-noise ratio of the market data, associated with the use of the past three observations, gives us a value of $q_M^* = 0.5$ from Table B. That is, we arrive at the estimate: $\sigma_e^2 \approx 2\sigma_v^2$. If we assume the decision maker acts optimally, then this must be the subjective level of noise in the data. However, if we assume the agent exhibits RB, then $q_M^* < 0.5$, that is $q_M^* = 0.25$ or $q_M^* = 0.125$ or even less![18] Depending on our perspective, we can now conclude either, but not both, of the following:

1. **Irrational Pricing or RB:** The signal-to-noise ratio is smaller than or bounded above by 0.5.

2. **Optimal Pricing:** The signal-to-noise ratio is approximately 0.5, as information is used optimally.

[18]Recall that returns are given by: $\Delta x_t = v_t + e_t - e_{t-1}$. Thus, under the RB assumption, returns will consist almost totally of noise.

The above literature supports the hypothesis that there is extreme levels of noise in observed market data. Taking this as true, then it is unreasonable to make the assumption that a decision maker uses observed data as the true value. The data suggests that real decision makers truly are deciding as though the observed decision variable is a noisy proxy for the true underlying state variable of interest. Thus, if any model of financial decision making does not model measurement error, then large model misspecification error will be introduced with certainty. The extreme level of noise in the market demands that its effects need to be accounted for specifically. This is true regardless of one's view concerning the use of optimal or heuristic decision rules.

The vast body of tests of market efficiency are based on the random walk hypothesis. To date, this literature has had disappointing results in its attempt to prove that markets are efficient. The results in this paper suggest that any model of market efficiency that is based on a model that ignores the large effects of noise will introduce very large model misspecification error. Fama (1998) discussed in detail the dire effects of the "bad [asset-pricing] model problem" which arises from model misspecification error introduced by using the wrong asset pricing model. It seems that many of Fama's concerns are as relevant in the case of the random walk model of market efficiency. The "bad market-efficiency model problem" must now be added to the list of reasonable explanations as to why empiricists should *expect* to find anomalies when using historical data with current models for financial markets that ignore noise.

8 Summary

Thirty years ago, the psychological literature proposed that agents exhibit RB, that is agents utilize fewer observations than required by an optimal Bayesian update. Although the psychological literature in the past ten years has rejected and overturned this proposal (see footnote 2), the finance literature continues to accept this as dogma.

Interestingly, in nearly a quarter century since this claim was brought to the finance profession, no paper carefully analyzes, in an imperfect and noisy environment, the relation between the empirically observed fact that agents use only a few recent data observations and the optimal data use with respect to Bayes' rule, which is the norm in financial markets. This paper fills this void.

My results echo the concerns discussed in Levy and Levy (forthcoming), who demonstrate that the static one-shot lotteries utilized in many of the individual behavioral experiments are not adequate to model investor decisions in a time series setting. They demonstrate that providing data in a one-shot time-series representation alters individual decisions compared to those made under a static lottery representation. Real markets present decision makers with a dynamically updating time series of returns. The data observed by real market participants is drastically different in representation than data used in either the heuristic literature or Levy and Levy. I compliment Levy and Levy (forthcoming) in that I demonstrate that in such dynamic environments with imperfect information, that is when noise is convolved with information, heavily weighting of recent data does not necessarily contradict Bayes' rule, as Bayes' rule can generate any level of weighting on recent data depending on the quality of information in the observed data. I conclude that reasonable estimates of the level of noise in financial data imply consistency with Bayes' rule and the observed use of data by decision makers in the empirical literature.

The Bayesian paradigm has been an important and useful tool utilized by finance and economic academics for decades. Some very basic and important intuitions in both fields have been gained due to the tractability of the standard Bayesian updating problem. Recently, the finance and economic literature has been espousing the non-viability of Bayes' rule. In this paper, I demonstrate that this position is unfounded. I also identify the cause of this misperception of Bayes' rule as invalid to the simple fact that implausible levels of noise would have to exist in the proxies for the true underlying decision variable to be able to validate RB. When Bayes' rule is adjusted for the presence of a noisy proxy many previous anomalies are predicted decision norms. Thus, the vast body of psychological and experimental literature on decision making supports (rather than violates, as claimed) the validity of the Bayesian paradigm.

A secondary, but still important, implication of the results of this paper is that finance and economic researchers should be cautious of adopting results directly from the psychological literature. Standards should be set, as a profession, to ensure that the structure of the psychological experiments reflect the structure of the environment faced by an investor, both institutionally and observationally, as previous literature suggests both are important determinants of the end results. Further, experiments, typically conducted on individuals utilizing standards and procedures much different from accepted experimental practices in economics, should be questioned and stress tested to ensure the results hold up to the standards required in our own profession. Lastly, we should ensure that results are generalizable from individual to aggregate settings, as ultimately a market is an aggregate entity.

Finally, to adequately construct an economic or market experiment that can distinguish between Bayes' rule and RB, the experimenter must control for (1) the level of noise perceived in the data, and (2) the level of overconfidence inherent in the agent. Both represent formidable challenges to the psychological and experimental literature. To date, not one paper has performed adequate controls to conclusively conclude that Bayes' rule is an invalid decision making paradigm. As argued in this paper, many of the empirically documented facts in the finance literature appear to be consistent with the assumption of a Bayesian decision maker.

Appendix A: Var-Cov Matrix and its Inverse

For all asset pricing models the variance-covariance matrix ultimately comes under consideration. What is more important is that in order to perform a multivariate Bayesian update, inverting the variance-covariance matrix is necessary. Typically this is an immensely difficult task if complete generality is desired. In the above models (equation (1)), we are fortunate to have some structure that makes this daunting task doable. The inversion is still difficult and involved. So as not to ruin the continuity or the intuition in the paper, I address invertability in this appendix.

Define the information set $I_t = \mathbf{\Delta x}_t = (\Delta x_1, \Delta x_2, \ldots, \Delta x_t)'$. For simplicity of notation, define $\Sigma_t = \Sigma_t^{\Delta x, \Delta x} = \mathrm{cov}(\mathbf{\Delta x}_t, \mathbf{\Delta x}_t)$. We can calculate the variance-covariance matrix of $\Delta x_t = v_t + e_t - e_{t-1}$ as follows:

$$\Sigma_t = \begin{bmatrix} 2\sigma_e^2+\sigma_v^2 & -\sigma_e^2 & 0 & 0 & \ldots & 0 & 0 & 0 \\ -\sigma_e^2 & 2\sigma_e^2+\sigma_v^2 & -\sigma_e^2 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -\sigma_e^2 & 2\sigma_e^2+\sigma_v^2 & -\sigma_e^2 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -\sigma_e^2 & 2\sigma_e^2+\sigma_v^2 & -\sigma_e^2 \\ 0 & 0 & 0 & 0 & \ldots & 0 & -\sigma_e^2 & 2\sigma_e^2+\sigma_v^2 \end{bmatrix} \tag{A.1}$$

This variance-covariance matrix has many desirable properties. First and foremost, it is a tridiagonal matrix. A band matrix is always easier to work with than a full matrix. Second, the matrix is both symmetric (⧅ or mathematically $A(i,j) = A(j,i)$) and persymmetric, (⧄ or mathematically $A(i,j) = A(n+1-j, n+1-i)$). Finally, the matrix has only one repeated row vector. However, this row is shifted to the right[19] one column each time the row number is increased by one. Such a matrix is referred to

[19] This means that a zero is dropped off the right side of the row vector and a zero is added to the left side of the row vector.

as a Toeplitz matrix.[20]

These three special properties of our variance-covariance matrix allow us to calculate its inverse. This allows us to derive a closed form solution for the Bayesian update problem in the next section. This closed-form solution then allows us to study the specifics of how a Bayesian agent updates expectations on the arrival of new information.

The symmetric and persymmetric properties of the variance-covariance matrix are preserved in the inversion process, so the inverse matrix has these properties as well. Thus, we only need to calculate one quadrant of the inverse matrix as symmetry gives the right quadrant and persymmetry gives the remainder of the matrix. If t is odd, then we need to calculate the following elements of the inverse variance-covariance matrix:

$$\text{For } t \text{ Odd} = \begin{bmatrix} & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & X & & & \\ & & & \vdots & & & \\ & & X & \dots & X & & \\ & X & X & \dots & X & X & \\ X & X & X & \dots & X & X & X \end{bmatrix}$$

If t is even, then we need to calculate the following elements of the inverse variance-

[20] The general form of a 4x4 Toeplitz matrix, for example, is:

$$T = \begin{bmatrix} r_0 & r_1 & r_2 & r_3 \\ r_{-1} & r_0 & r_1 & r_2 \\ r_{-2} & r_{-1} & r_0 & r_1 \\ r_{-3} & r_{-2} & r_{-1} & r_0 \end{bmatrix}$$

In our case, $r_{-1} = r_1 = -\sigma_e^2$, $r_0 = 2\sigma_e^2 + \sigma_v^2$, and all other $r_i = 0$ for $|i| > 1$.

covariance matrix:

$$\text{For } t \text{ Even} = \begin{bmatrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & X & X & & & \\ & & & \vdots & \vdots & & & \\ & & X & \dots & \dots & X & & \\ & X & X & \dots & \dots & X & X & \\ X & X & X & \dots & \dots & X & X & X \end{bmatrix}$$

Theorem A *Given the variance-covariance matrix Σ_t of the Random Walk with Noise economy represented by equations (1), the inverse, Σ_t^{-1}, can be calculated.*
For t odd, Σ_t^{-1} equals:

$$\frac{1}{f_{t+1}(t)}\begin{bmatrix} & & & & & & \\ & & & & & & \\ & & & \sum_{j=0}^{\frac{t-1}{2}} f_t(2j) & & & \\ & & & \vdots & & & \\ & & \sum_{j=0}^{2} f_t(2j) & \dots & \sum_{j=0}^{2} f_t(2j+t-5) & & \\ & \sum_{j=0}^{1} f_t(2j) & \sum_{j=0}^{1} f_t(2j+1) & \dots & \sum_{j=0}^{1} f_t(2j+t-4) & \sum_{j=0}^{1} f_t(2j+t-3) & \\ f_t(0) & f_t(1) & f_t(2) & \dots & f_t(t-3) & f_t(t-2) & f_t(t-1) \end{bmatrix}$$

For t even, Σ_t^{-1} equals:

$$\frac{1}{f_{t+1}(t)}\begin{bmatrix} & & & & & & & \\ & & & & & & & \\ & & & \sum\limits_{j=0}^{\frac{t-2}{2}} f_t(2j) & \sum\limits_{j=0}^{\frac{t-2}{2}} f_t(2j+1) & & & \\ & & & \vdots & \vdots & & & \\ & & \sum\limits_{j=0}^{2} f_t(2j) & \cdots & \cdots & \sum\limits_{j=0}^{2} f_t(2j+t-5) & & \\ & \sum\limits_{j=0}^{1} f_t(2j) & \sum\limits_{j=0}^{1} f_t(2j+1) & \cdots & \cdots & \sum\limits_{j=0}^{1} f_t(2j+t-4) & \sum\limits_{j=0}^{1} f_t(2j+t-3) & \\ f_t(0) & f_t(1) & f_t(2) & \cdots & \cdots & f_t(t-3) & f_t(t-2) & f_t(t-1) \end{bmatrix}$$

Where,

$$f_t(n) = \sum_{k=0}^{n} \binom{n+1+k}{n-k} \sigma_e^{2(t-1-k)} \sigma_v^{2k} \tag{A.2}$$

and the i^{th}, j^{th} element of Σ_t^{-1} can be written as:

$$\Sigma_t^{-1}(i,j) = \sum_{k=0}^{t-i} f_t(2k+j-1-t+i) \tag{A.3}$$

Proof of Theorem A

For $t = 2$ and $t = 3$ the inverse is calculated directly from Gauss elimination as:

$$\Sigma_2^{-1} = \frac{1}{3\sigma_e^4 + 4\sigma_e^2\sigma_v^2 + \sigma_v^4} \begin{bmatrix} 2\sigma_e^2+\sigma_v^2 & \sigma_e^2 \\ \sigma_e^2 & 2\sigma_e^2+\sigma_v^2 \end{bmatrix}$$

$$\Sigma_3^{-1} = \frac{1}{4\sigma_e^6+10\sigma_e^4\sigma_v^2+6\sigma_e^2\sigma_v^4+\sigma_v^6} \begin{bmatrix} 3\sigma_e^4+4\sigma_e^2\sigma_v^2+\sigma_v^4 & 2\sigma_e^4+\sigma_e^2\sigma_v^2 & \sigma_e^4 \\ 2\sigma_e^4+\sigma_e^2\sigma_v^2 & 4\sigma_e^4+4\sigma_e^2\sigma_v^2+\sigma_v^4 & 2\sigma_e^4+\sigma_e^2\sigma_v^2 \\ \sigma_e^4 & 2\sigma_e^4+\sigma_e^2\sigma_v^2 & 3\sigma_e^4+4\sigma_e^2\sigma_v^2+\sigma_v^4 \end{bmatrix}$$

Calculating $f_t(n)$ for the following combinations of t and n, we directly verify the validity of the inverse matrix for $t = 2$ and $t = 3$.

$$f_2(0) = \sigma_e^2$$

$$f_2(1) = 2\sigma_e^2 + \sigma_v^2$$

and

$$f_3(0) = \sigma_e^4$$

$$f_3(1) = 2\sigma_e^4 + \sigma_e^2\sigma_v^2$$

$$f_3(2) = 3\sigma_e^4 + 4\sigma_e^2\sigma_v^2 + \sigma_v^4$$

and

$$f_4(3) = 4\sigma_e^6 + 10\sigma_e^4\sigma_v^2 + 6\sigma_e^2\sigma_v^4 + \sigma_v^6$$

Now make the induction step and assume that the inverse formula holds for $t = T-1$. It is only necessary to show the validity for $t = T$ and we am done. By a simple application of Gauss elimination, we obtain the inverse of the following partitioned matrix:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1}(I + A_{12}FA_{21}A_{11}^{-1}) & -A_{11}^{-1}A_{12}F \\ -FA_{21}A_{11}^{-1} & F \end{bmatrix}$$

Where

$$F = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$$

Since we have

$$\Sigma_T = \begin{bmatrix} \Sigma_{T-1} & A_{21}' \\ A_{21} & 2\sigma_e^2 + \sigma_v^2 \end{bmatrix}$$

If we let a prime ($'$) signify matrix transposition, then $A_{11} = \Sigma_{T-1}$, $A_{22} = (2\sigma_e^2 + \sigma_v^2)$, $A_{12} = A_{21}'$ and A_{21} is the following $1 \times (T-1)$ matrix:

$$A_{12} = [0, 0, \ldots, 0, -\sigma_e^2]$$

Therefore we immediately get the inverse of Σ_T from the formula for the inverse of a partitioned matrix above:

(A.4) $$F = (A_{22} - A_{12}' \cdot A_{11}^{-1} \cdot A_{12})^{-1} = [(2\sigma_e^2 + \sigma_v^2) - A_{12}' \cdot \Sigma_{T-1}^{-1} A_{12}]^{-1}$$

$$\Sigma_T^{-1} = \begin{bmatrix} \Sigma_{T-1}^{-1}(I + A_{12} \cdot FA_{12}' \cdot \Sigma_{T-1}^{-1}) & -\Sigma_{T-1}^{-1}A_{12}F \\ -FA_{12}' \cdot \Sigma_{T-1}^{-1} & F \end{bmatrix} \tag{A.5}$$

Note that in our application, since A_{22} is a scalar, F is a scalar and can be moved in and/or out of the matrices as desired. Also note that A_{12} pulls the last column out of Σ_{T-1}^{-1} and then A_{12}' pulls the last row or element out of $\Sigma_{T-1}^{-1}A_{12}$. We thus get a simple formula for F as follows:

(A.6)

$$F = \left[(2\sigma_e^2 + \sigma_v^2) - \sigma_e^4 \frac{f_{T-1}(T-2)}{f_T(T-1)}\right]^{-1} = \left[\frac{f_T(T-1)}{(2\sigma_e^2 + \sigma_v^2)f_T(T-1) - \sigma_e^4 f_{T-1}(T-2)}\right]$$

By extending the definition of the binomial coefficient to the cases $j < 0$ and $n < j$, such that:

$$\binom{n}{j} = 0 \tag{A.7}$$

it is possible to handle the calculation of $F \; \forall \, n$, rather than on a case by case basis. We will need an elementary analytical result for binomial coefficients:[21]

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1} \tag{A.8}$$

To finish the derivation of F, we need to show that the denominator in Equation (A.6) is $f_{T+1}(T)$. This will then show that in fact $F = (\Sigma_T)_{T,T}$ according to the formulas in Theorem 1. Utilizing the definition of $f_t(n)$ in Equation (A.2):

$$\begin{aligned} (2\sigma_e^2 + \sigma_v^2)f_T(T-1) - f_{T-1}(T-2) = \quad & 2\sum_{k=0}^{T-1} \binom{T+k}{T-1-k} \sigma_e^{2(T-k)} \sigma_v^{2k} \\ & + \sum_{k=0}^{T-1} \binom{T+k}{T-1-k} \sigma_e^{2(T-1-k)} \sigma_v^{2(k+1)} \\ & + \sum_{k=0}^{T-2} \binom{T-1+k}{T-2-k} \sigma_e^{2(T-k)} \sigma_v^{2k} \end{aligned} \tag{A.9}$$

[21] Equation (3.1.4) in Abramowitz and Stegun (1972).

$$
\begin{aligned}
(2\sigma_e^2+\sigma_v^2)f_T(T-1)-f_{T-1}(T-2) = \quad & 2\sum_{k=0}^{T-1}\binom{T+k}{T-1-k}\sigma_e^{2(T-k)}\sigma_v^{2k} \\
& + \sum_{k=1}^{T}\binom{T-1+k}{T-k}\sigma_e^{2(T-k)}\sigma_v^{2k} \\
& + \sum_{k=0}^{T-2}\binom{T-1+k}{T-2-k}\sigma_e^{2(T-k)}\sigma_v^{2k}
\end{aligned}
\tag{A.10}
$$

Utilizing the extended definition of the binomial coefficient Equation (A.7), the coefficient of $\sigma_e^{2(T-k)}\sigma_v^{2k}$ can be written for all $k = 0 \ldots T$ as:

$$
c_k = 2\binom{T+k}{T-1-k} + \binom{T-1+k}{T-k} - \binom{T-1+k}{T-2-k} \tag{A.11}
$$

Repeatedly applying Equation (A.8):

$$
c_k = \binom{T+k}{T-1-k} + \binom{T-1+k}{T-k} + \underbrace{\binom{T+k}{T-1-k} - \binom{T-1+k}{T-2-k}}_{use\ \ Equation\,(A.8)} \tag{A.12}
$$

$$
c_k = \binom{T+k}{T-1-k} + \underbrace{\binom{T-1+k}{T-k} + \binom{T-1+k}{T-1-k}}_{use\ \ Equation\,(A.8)} \tag{A.13}
$$

$$
c_k = \underbrace{\binom{T+k}{T-1-k} + \binom{T+k}{T-k}}_{use\ \ Equation\,(A.8)} \tag{A.14}
$$

$$
c_k = \binom{T+1+k}{T-k} = \binom{(T+1)+k}{(T+1)-1-k} \tag{A.15}
$$

Thus, plugging in c_k and calculating the denominator of F, we can then write f as:

$$
F = \frac{f_T(T-1)}{\sum_{k=0}^{T} c_k\sigma_e^{2(T-k)}\sigma_v^{2k}} = \frac{f_T(T-1)}{\sum_{k=0}^{T}\binom{T+1+k}{T-k}\sigma_e^{2(T-k)}\sigma_v^{2k}} = \frac{f_T(T-1)}{f_{T+1}(T)} \tag{A.16}
$$

This proves that F, which is the $(T,T)^{th}$ element of Σ_T^{-1}, is in fact as the formula suggests.

The next step in the construction of Σ_T^{-1} is to computer directly the remainder of the T^{th} row. The formula is given by Equation (A.5) as $R_T = -FA_{12}^{'} \cdot \Sigma_{T-1}^{-1}$. Recalling from above that $A_{12}^{'}$ pulls the last row from Σ_{T-1}^{-1} multiplied by the constant $-\sigma_e^2$, we get:

$$R_T = \sigma_e^2 F \frac{1}{f_T(T-1)} \left[f_{T-1}(0), f_{T-1}(1), \ldots, f_{T-1}(T-2) \right] \tag{A.17}$$

Noting that by definition of $f_t(n)$ that $\sigma_e^2 f_{t-1}(n) = f_t(n) \forall t$ and recalling the formula of F given by Equation (A.16), we obtain:

$$R_T = \frac{1}{f_{T+1}(T)} \left[f_T(0), f_T(1), \ldots, f_T(T-2) \right] \tag{A.18}$$

we have thus constructed the last row of Σ_T^{-1} and demonstrated that the formula holds in this case.

An important property of our matrix already noted is that it is Toeplitz. A Toeplitz matrix preserves the symmetry and persymmetry properties of the matrix upon the operation of inversion. Thus, Σ_t^{-1} is both symmetric and persymmetric for all t. This allows us to calculate the remainder of the lower quarter of the matrix inverse and thus to finish off the proof of the induction step. The symmetry of Σ_T^{-1} gives the last column of the inverse from the last row and the persymmetry property gives the remaining two edges of the inverse matrix, the first row and, most importantly, the first column. we have thus shown:

$$\frac{1}{f_{T+1}(T)} \left[\begin{array}{cccccc|c} f_T(T-1) & f_T(T-2) & f_T(T-3) & \ldots & f_T(2) & f_T(1) & f_T(0) \\ f_T(T-2) & & & & & & f_T(1) \\ f_T(T-3) & & & & & & f_T(2) \\ \vdots & & & & & & \vdots \\ f_T(2) & & & & & & f_T(T-3) \\ f_T(1) & & & & & & f_T(T-2) \\ \hline f_T(0) & f_T(1) & f_T(2) & \ldots & f_T(T-3) & f_T(T-2) & f_T(T-1) \end{array} \right]$$

To finish the proof, we note that all entries of the inverse matrix are determined from the last row of the inverse matrix. The structure of the variance-covariance matrix for

$n = 1 \ldots T - 1$, that is the Toeplitz, symmetric and persymmetric properties, implies the specific structure of the inverse matrix where all entries are determined from the last row in the specified manner. As the variance-covariance matrix for $n = T$ has the same structure and properties, its inverse will be determined once its last row is known and it will possess the exact structure. This symmetry argument completes the proof.

Q.E.D.

Appendix B: Proof of Theorem 1

Theorem 1 *For a quadratic loss function and the economy with measurement error represented by equation (1), the optimal Bayes' update is given by:*

$$h(I_t) = E[s_{t+1}|I_t] - E[s_t|I_{t-1}] \tag{B.1}$$

where $I_t = \mathbf{\Delta x}_t = (\Delta x_1, \Delta x_2, \ldots, \Delta x_t)'$ *is the vector of changes in the observed variable up to time* t. $h(I_t)$ *is the Bayesian update to the time-*$(t-1)$ *expectation of the state parameter,* s_t*, on the arrival of a new observation,* Δx_t*. The optimal update function,* $h(I_t)$ *is given by:*

$$h(I_t) = k\, \mathbf{w}_t' \cdot \mathbf{z}_t \tag{B.2}$$

where $\mathbf{w}_t$ *is a weighting vector given by:*

$$\mathbf{w}_t' = [(\sigma_e^2)^{t-1},\, (\sigma_e^2)^{t-2}(\sigma_v^2),\, (\sigma_e^2)^{t-3}(\sigma_v^2)^2, \cdots, (\sigma_e^2)(\sigma_v^2)^{t-2},\, (\sigma_v^2)^{t-1}] \tag{B.3}$$

$\mathbf{z}_t = \mathbf{z}_t(I_t)$ *is the vector function of past observations, which is given by:*

$$\mathbf{z}_t = \begin{bmatrix} \binom{1}{0}\Delta x_1 + \binom{2}{1}\Delta x_2 + \binom{3}{2}\Delta x_3 + \cdots + \binom{t-1}{t-2}\Delta x_{t-1} + \binom{t}{t-1}\Delta x_t \\ \binom{3}{0}\Delta x_2 + \binom{4}{1}\Delta x_3 + \cdots + \binom{t}{t-3}\Delta x_{t-1} + \binom{t+1}{t-2}\Delta x_t \\ \binom{5}{0}\Delta x_3 + \cdots + \binom{t+1}{t-4}\Delta x_{t-1} + \binom{t+2}{t-3}\Delta x_t \\ \ddots \\ \binom{2t-3}{0}\Delta x_{t-1} + \binom{2t-2}{1}\Delta x_t \\ \binom{2t-1}{0}\Delta x_t \end{bmatrix} \tag{B.4}$$

and k *is a constant given by:*

$$k = \frac{\sigma_v^2}{f_{t+1}(t) - \sigma_e^2 f_t(t-1) + \sigma_e^{2t}}$$

Proof

Define $\mathbf{v}_t = (v_1, v_2, \ldots, v_t)'$ and $I_t = \mathbf{\Delta x}_t = (\Delta x_1, \Delta x_2, \ldots, \Delta x_t)'$. For simplicity of notation, define $\Sigma_t = \Sigma_t^{\Delta x, \Delta x} = \text{cov}(\mathbf{\Delta x}_t, \mathbf{\Delta x}_t)$. First, it is a well known fact that given a quadratic loss function, $L = [s_t - g(\mathbf{\Delta x}_t)]^2$, that the optimal decision function g^* is the mean of the conditional distribution of s_t given the observations $\mathbf{\Delta x}_t$, thus, $g^*(\mathbf{\Delta x}_t) = E[s_t|\mathbf{\Delta x}_t] = E[s_t|I_t]$. Given the assumptions, we can calculate the mean vectors and the variance-covariance matrices as follows:

$$\mu_t^v = \mu_t^{\Delta x} = \mathbf{0}_t \qquad \text{and} \qquad \Sigma_t^{v,v} = \Sigma_t^{v,\Delta x} = \sigma_v^2 I$$

and $\Sigma_t = \Sigma_t^{\Delta x, \Delta x}$ as defined in Equation (A.1). Using standard normal distribution theory results[22] and utilizing the linearity of the expectations operator, we obtain:

$$E[s_t|I_t] = \sigma_v^2 \left(\mathbf{1}' \cdot \Sigma_t^{-1} \cdot \mathbf{\Delta x}_t\right) \tag{B.5}$$

By definition, $h(I_t) = E[s_{t+1}|I_t] - E[s_t|I_{t-1}]$. Using equation (B.5), the inverse variance-covariance matrix, Σ_t^{-1}, as calculated in Theorem A and noting that $E[s_{t+1}t|I_t] = E[s_t|I_t]$, we obtain equation (**??**) for $h(I_t)$.

Q.E.D.

[22] See Greene (2000), page 87.

Appendix C: Proof of Theorem 2

The signal extraction problem is well studied in many fields. Muth (1960) is probably the most noted reference. However, this proof is the first finite time horizon solution to my knowledge. There are two main results derived from this theorem. The first result is the exponential nature of decay for the weights applied to past data typical of the optimal signal extraction. This is not a new insight, but it is comforting to find that the decay occurs quite quickly and is a strong property of the finite time optimal solution as well. The second result is specific to the finite time solution. I am able to decompose the decay structure of the weights applied to past data into two main components, each reflecting a different cause for the inherent "recency" effect. The first component is due to the noise level of observed date. This is reflected in the weighting vector, $\mathbf{w}_t$, of Theorem 1. The more noise there is in observed data, the more weight will be placed on past data. The second component reflects the inherent recency of the linear forecasting rules that constitute an optimal Bayes' update. This is reflected in the $\mathbf{z}_t$-vector of Theorem 1. This decomposition is new.

Theorem 2 *For a quadratic loss function and the economy with measurement error represented by equation (1), a Bayesian agent will make decisions that will heavily emphasize the most recent data for any reasonable level of noise in the market.*

Proof

As in Example 2, I define for any finite time t, the sum of all weights for the k^{th} observation as the sum of all the coefficients in the k^{th}-column. This sum can be written:

$$\omega_{t,k} = \sum_{j=1}^{k} c_{j,k} = \sum_{j=1}^{k} \binom{k-1+j}{k-j} \tag{C.1}$$

I will first prove that at any time t, the weighting on the k^{th} observation is greater than the weighting on the $(k-1)^{st}$ observation. Forming the ratio and expanding the factorials:

$$\frac{\omega_{t,k+1}}{\omega_{t,k}} = \frac{\sum_{j=1}^{k} \frac{(k-1+j)(k-2+j)\cdots(k+1-j)}{(2j-1)!}}{\sum_{j=1}^{k-1} \frac{(k-2+j)(k-3+j)\cdots(k-j)}{(2j-1)!}}$$

$$= \frac{\sum_{j=1}^{k} \frac{(k-1+j)[(k-2+j)\cdots(k+1-j)]}{(2j-1)!}}{\sum_{j=1}^{k-1} \frac{[(k-2+j)\cdots(k+1-j)](k-j)}{(2j-1)!}} \tag{C.2}$$

$$= \underbrace{\frac{\sum_{j=1}^{k-1} \frac{(k-1+j)[(k-2+j)\cdots(k+1-j)]}{(2j-1)!}}{\sum_{j=1}^{k-1} \frac{(k-j)[(k-2+j)\cdots(k+1-j)]}{(2j-1)!}}}_{>1} + \underbrace{\frac{\frac{(2k-1)[(2k-2)\cdots(1)]}{(2k-1)!}}{\sum_{j=1}^{k-1} \frac{[(k-2+j)\cdots(k+1-j)](k-j)}{(2j-1)!}}}_{>0}$$

The first term on the right hand side is greater than one as $(k+j) > (k+1-j)$, $\forall j$ by construction. The second term on the right is greater than zero. This proves the first claim that $\omega_{t,k+1} > \omega_{t,k} \forall t, k$.

Now we want to show that $\omega_{t,k+n}/\omega_{t,k} \to 1$ as $n \to \infty$. Again, forming the ratio and expanding the factorials:

(C.3)

$$\frac{\omega_{t,k+n}}{\omega_{t,k}} = \frac{\sum_{j=1}^{k+n} \frac{(k+n-1+j)(k+n-2+j)\cdots(k+n+1-j)}{(2j-1)!}}{\sum_{j=1}^{k} \frac{(k-1+j)(k-2+j)\cdots(k+1-j)}{(2j-1)!}}$$

$$= \underbrace{\frac{\sum_{j=1}^{k} \frac{(k-1+j+n)(k-2+j+n)\cdots(k+1-j+n)}{(2j-1)!}}{\sum_{j=1}^{k-1} \frac{(k-1+j)(k-2+j)\cdots(k+1-j)}{(2j-1)!}}}_{\to\infty \quad as \quad n\to\infty} + \underbrace{\frac{\sum_{j=k+1}^{k+n} \frac{(k-1+j+n)(k-2+j+n)\cdots(k+1-j+n)}{(2k-1)!}}{\sum_{j=1}^{k-1} \frac{(k-1+j)(k-2+j)\cdots(k+1-j)}{(2j-1)!}}}_{>0}$$

Thus we have shown the finite property of each new observation is more weighted then the most heavily weighted observation to date and that all the weight of the data is eventually concentrated in a section of the last data. This proves the second claim. Simulations are needed to demonstrate finite sample properties and speed of convergence. Please see Tables A, B, and C and the simulation section.

Q.E.D.

Appendix D: Proof of Theorem 3

Theorem 3 *For a quadratic loss function and the economy with measurement error represented by equation (1), an overconfident Bayesian agent will continue to make decisions that heavily weight the most recent data. An overconfident Bayesian will exhibit an enhanced use of recent data compared to a rational Bayesian. Extreme overconfidence implies that the latest or most recent observation will dominate all observable decisions by the agent.*

Proof

Overconfidence means that the agent believes the observed data is more informative than it actually is. This is interpreted as, given the relation of the information effect σ_v^2 to the noise effect σ_e^2, the agent perceives this relation to be more informative than it is. To see this, it is probably easier to rewrite the weighting vector $\mathbf{w_t}$ by first defining the signal-to-noise ratio $q = S/N = \sigma_v^2/\sigma_e^2$. Now it is easy to represent overconfidence, as the subjective signal-to-noise ratio is believed to be larger than the true signal-to-noise ratio, ($q_{subjective} > q_{true}$). Writing $\mathbf{w_t}$ in terms of q we get:

$$\begin{aligned} \mathbf{w}_t^{'} &= (\sigma_e^2)^{t-1}[1,\, \tfrac{\sigma_v^2}{\sigma_e^2},\, \left(\tfrac{\sigma_v^2}{\sigma_e^2}\right)^2, \cdots,\, \left(\tfrac{\sigma_v^2}{\sigma_e^2}\right)^{t-2},\, \left(\tfrac{\sigma_v^2}{\sigma_e^2}\right)^{t-1}] \\ &= (\sigma_e^2)^{t-1}[1,\, q,\, q^2, \cdots,\, q^{t-2},\, q^{t-1}] \end{aligned} \tag{D.1}$$

The growth is exponential and therefore it can be very significant. The leading term, $(\sigma_e^2)^{t-1}$, in $\mathbf{w}_t^{'}$ is a constant that is multiplied to every element in the updating vector. Thus this term will not effect the relative values of the different entries and, for our purposes here, can be ignore. We thus need only consider the vector $W = \frac{1}{(\sigma_e^2)^{t-1}}\mathbf{w}_t^{'}$, where:

$$W = [1,\, q,\, q^2, \cdots,\, q^{t-2},\, q^{t-1}] \tag{D.2}$$

Let $m > n$. If $q > 1$, the exponential growth in q weights the later linear forecasting models, say $\mathbf{z}_{t,m}$ in the $\mathbf{z}_t$ vector of equation (B.4) at the expense of the earlier forecasting models, $\mathbf{z}_{t,n}$. It is precisely these later linear forecasting models that use fewer data points. The data points used in $z_{t,m}$ are also the most recent $t-m-1$ observations. And as proven previously, the coefficients are higher on the more recent data. Thus, the linear forecasts with the most weight on the most recent data are more heavily overweighted by an overconfident decision maker compared to a standard decision maker. If the agent is extremely overconfident such that it is believed that full information can be extracted from observed data, i.e. $q \to \infty$, then only the last linear forecasting model, $\mathbf{z}_{t,t}$ is used to update. This singular forecasting model consists of one data observation, which is the last observation Δx_t. Thus an extremely overconfident agent will only use the very last observed data point and will completely dominate the decision making process.

Q.E.D.

9 References

Abramowitz, Milton and Irene A. Stegun, (1972). *Handbook of Mathematical Functions.* Dover Publications, Inc., New York.

Andreassen, P. and S. Kraus, (1988). Judgmental Prediction by Extrapolation. *Harvard University Working Paper*, Department of Psychology.

Berger, James O., (1993). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer Series in Statistics, Second Edition, Springer-Verlag New York, LLC.

Brown, Keith C., W. V. Harlow and Laura T. Starks, (1996). Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry. *Journal of Finance*, **51 (1):** 85–110.

Chevalier, Judith and Glenn Ellison, (1997). Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*, **105 (6):** 1167–1200.

Conrad Jennifer, Mustafa N. Gultekin and Gautam Kaul, (1997). Profitability of Short-Term Contrarian Strategies: Implications for Market Efficiency. *Journal of Business & Economic Statistics*, **15 (3):** 379–386.

Daniel, Kent D. and Sheridan Titman, (1999). Market Efficiency in an Irrational World. *Financial Analysts Journal*, **55 (6):** 28–40.

De Bondt, Werner F. M. and Richard H. Thaler, (1985). Does the Stock Market Overreact? *The Journal of Finance*, **40 (3):** 793–805.

De Bondt, Werner F. M. and Richard H. Thaler, (1987). Further Evidence on Investor Overreaction and Stock Market Seasonality. *The Journal of Finance*, **42 (3):** 557–581.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers and Robert J. Waldmann, (1989). The Size and Incidence of the Losses from Noise Trading. *The Journal of Finance*, **44 (3):** 681–696

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers and Robert J. Waldmann, (1990). Positive Feedback Investment Strategies and Destabilizing Rational Speculation. *Journal of*

Business, **45 (2):** 379–395.

Derrien, Franois and Kent L. Womack, (2003). Auctions vs. bookbuilding and the control of underpricing in hot IPO markets. *Review of Financial Studies*, **16 (1):** 31–61.

Fama, Eugene, (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *The Journal of Financial Economics*, **49 (3):** 283–306.

Fama, Eugene, and Kenneth French, (1988). Permanent and Temporary Components of Stock-Prices. *Journal of Political Economy*, **96 (2):** 119–137.

Feldstein, Martin, (2003). Monetary Policy in an Uncertain Environment. *NBER Working Paper*, **No. w9969**.

Frankel, Jeffrey A. and Kenneth A. Froot, (1990). Chartists, Fundamentalists, and Trading in the Foreign Exchange Market. *American Economic Review*, **80 (2):** 181–185.

Gallaher, Steven, Ron Kaniel and Laura Starks, (2005). Madison Avenue Meets Wall Street: Mutual Fund Families, Competition and Advertising. *Working Paper.*

Gigerenzer, Gerd, (1999). *Simple Heuristics That Make Us Smart.* Gerd Gigerenzer, Peter M. Todd, and the ABC Research Group. New York : Oxford University Press.

Greene, William, (2000). *Econometric Analysis*, Prentice Hall, 4th Edition.

Grossman, Sanford, (1976). On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information. *Journal of Finance*, **31 (2):** 573–585.

Gruber, Martin J., (1996). Another Puzzle: The Growth in Actively Managed Mutual Funds. *Journal of Finance*, **51 (3):** 783–810.

Hertwig, Ralph and Andreas Ortmann, (2001). Experimental Practices in Economics: A Methodological Challenge for Psychologists? *Behavioral and Brain Sciences*, **24 (3):** 383–403.

Jamal, Karim and Shyam Sunder, (1996). Bayesian Equilibrium in Double Auctions Populated by Biased Heuristic Traders. *Journal of Economic Behavior and Organization*, **31 (2):** 273–291.

Jamal, Karim and Shyam Sunder, (2001). Why Do Biased Heuristics Approximate Bayes Rule in Double Auctions? *Journal of Economic Behavior and Organization*, **46 (4):** 431–435.

Jegadeesh, Narasimhan and Sheridan Titman, (1993). Returns to Buying Winners and Selling Losers - Implications for Stock-Market Efficiency. *The Journal of Finance*, **48 (1):** 65–91.

Jordan, Steven J., (2004a). Noise vs. Information: The Excess Volatility Puzzle. *Yale University Working Paper*, Yale School of Management.

Jordan, Steven J., (2004b). Sorting Out Sorts II: Chance and Long-Term Reversals Revisited. *Yale University Working Paper*, Yale School of Management.

Kandel, Shmuel and Robert F. Stambaugh, (1996). On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance*, **51 (2):** 385–424.

Karceski, Jason, (2002). Returns-Chasing Behavior, Mutual Funds, and Beta's Death. *Journal of Financial and Quantitative Analysis*, **37 (4):** 559–594.

Koehler, Jonathan J., (1996). The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges. *Behavioral and Brain Sciences*, **19: (1):** 1-53.

Krueger, Joachim I. and David C. Funder, (2004). Towards a Balanced Social Psychology: Causes, Consequences, and Cures for the Problem-Seeking Approach to Social Behavior and Cognition. *Behavioral and Brain Sciences*, **27 (3):** 313–327.

Levy, Haim and Moshe Levy, (Forthcoming). R. Zwick and A. Rapoport (Eds.) *Experimental Business Research.* **Vol 3**, Springer.

Muth, John F., (1960). Optimal Properties of Exponentially Weighted Forecasts. *Journal of the American Statistical Association*, **55 (290):** 299–306.

Odean, Terrance, (1999). Do Investors Trade Too Much? *American Economic Review*, **89 (5):** 1279–1298.

Sandroni, Alvaro, (2005). Efficient Markets and Bayes' Rule. *Economic Theory*, **26 (4):** 741–764.

Sedlmeier, Peter and Gerd Gigerenzer, (2001). Teaching Bayesian Reasoning in Less Than Two Hours. *Journal of Experimental Psychology - General*, **130 (3):** 380–400.

Shanken, Jay and Ane Tamayo, (2001). Risk, Mispricing, and Asset Allocation: Conditioning on Dividend Yield. *NBER Working Paper*, **No. w8666**.

Shefrin, Hersh and Meir Statman, (1994). Behavioral Capital Asset Pricing Theory. *Journal of Financial & Quantitative Analysis*, **29 (3):** 323–349.

Shiller, Robert J., (2002). Bubbles, Human Judgment, and Expert Opinion. *Financial Analysts Journal*, **58 (3):** 18–26.

Shiller, Robert J. and John Pound, (1989). Survey Evidence on Diffusion of Interest and Information Among Investors. *Journal of Economic Behavior & Organization*, **12 (1):** 47–66.

Shleifer, Andrei and Lawrence H. Summers, (1990). The Noise Trader Approach to Finance. in Symposium: Bubbles, *The Journal of Economic Perspectives*, **4 (2):** 19–33.

Simon, Herbert A., (1956). Dynamic Programming Under Uncertainty with a Quadratic Criterion Function. *Econometrica*, **24 (1):** 74–81.

Sirri, Erik R. and Peter Tufano, (1998). Costly Search and Mutual Fund Flows. *Journal of Finance*, **53 (5):** 1589–1622.

Summers, Lawrence H., (1986). Does the Stock Market Rationally Reflect Fundamental Values? *Journal of Finance*, **41 (3):**, 591-601.

Tversky, Amos and Daniel Kahneman, (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, **185 (4157):** 1124–1131.

Warther, Vincent A., (1995). Aggregate mutual fund flows and security returns. *The Journal of Financial Economics*, **39 (2-3):** 209–235.

Table A: Simulation Results for Bayesian Recency Effect Benchmark Market - (q = 1)

I simulate the updating coefficients for each observation that a Bayesian agent facing measurement error will use. The following coefficient equation gives each observation's total weighting:

$$\omega_{t,k} = \sum_{j=1}^{k} c_{j,k} = \sum_{j=1}^{k} \binom{k-1+j}{k-j}$$

I then calculate the ratio of the weights in the last N observations to the total weights on all observations at time-t. The lowest number N such that $P(N) > 0.50$ for the 50% number and such that $P(N) > 0.90$ for the 90% number is calculated. The calculated number shown in this table is the number of observations that accounts for 50% and 90% of the total weight respectively for a Bayesian agent. The ratio of the last N observation weights to the total weight is given by:

$$P(N) = \frac{\sum_{k=t-N+1}^{t} \omega_{t,k}}{\sum_{k=1}^{t} \omega_{t,k}} = \frac{\sum_{k=t-N+1}^{t} \sum_{j=1}^{k} c_{j,k}}{\sum_{k=1}^{t} \sum_{j=1}^{k} c_{j,k}} = \frac{\sum_{k=t-N+1}^{t} \sum_{j=1}^{k} \binom{k-1+j}{k-j}}{\sum_{k=1}^{t} \sum_{j=1}^{k} \binom{k-1+j}{k-j}}$$

The lowest number of observations necessary to account for $100x\%$ of the total weighting used in the Bayesian update vector is given by:

$$P(N) > x \qquad \text{where } x = 0.50,\ 0.90$$

The following table gives the calculations of the recency effect for each length of time (1, 5, 20 and 40 years).

Number of observations, N, that has at least 50% or 90% of total weighting

Percent of Weights	12 months (1 year)	60 months (5 years)	240 months (20 years)	480 months (40 years)
q = 1				
50%	1	1	1	1
90%	2	2	2	2

Table B: Simulation Results for Bayesian Recency Effect Noisy Market - (q < 1)

I simulate the updating coefficients for each observation that a Bayesian agent facing measurement error will use. The following coefficient equation gives each observation's total weighting:

$$\omega_{t,k} = \sum_{j=1}^{k} c_{j,k} = \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}$$

I then calculate the ratio of the weights in the last N observations to the total weights on all observations at time-t. The lowest number N such that $P(N) > 0.50$ for the 50% number and such that $P(N) > 0.90$ for the 90% number is calculated. The calculated number shown in this tableis the number of observations that account for 50% and 90% of the total weights respectively for a Bayesian agent. The ratio of the last N observation weights to the total weights is given by:

$$P(N) = \frac{\sum\limits_{k=t-N+1}^{t} \omega_{t,k}}{\sum\limits_{k=1}^{t} \omega_{t,k}} = \frac{\sum\limits_{k=t-N+1}^{t} \sum\limits_{j=1}^{k} c_{j,k}}{\sum\limits_{k=1}^{t} \sum\limits_{j=1}^{k} c_{j,k}} = \frac{\sum\limits_{k=t-N+1}^{t} \sum\limits_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}}{\sum\limits_{k=1}^{t} \sum\limits_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}}$$

The lowest number of observations necessary to account for $100x$% of the total weighting used in the Bayesian update vector is given by:

$$P(N) > x \qquad \text{where } x = 0.50,\, 0.90$$

The following table gives the calculations of the recency effect for each length of time (1, 5, 20 and 40 years).

Number of observations, N, that has at least 50% or 90% of total weighting

Percent of Weights	12 months (1 year)	60 months (5 years)	240 months (20 years)	480 months (40 years)
q = 0.5				
50%	1	1	1	1
90%	3	3	3	3
q = 0.25				
50%	1	1	1	1
90%	4	4	4	4
q = 0.125				
50%	1	1	1	1
90%	5	6	6	6
q = 0.0625				
50%	2	2	2	2
90%	7	9	9	9

Table C: Simulation Results for Bayesian Recency Effect Informative or Overconfident Market - (q > 1)

I simulate the updating coefficients for each observation that a Bayesian agent facing measurement error will use. The following coefficient equation gives each observation's total weighting:

$$\omega_{t,k} = \sum_{j=1}^{k} c_{j,k} = \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}$$

I then calculate the ratio of the weights in the last N observations to the total weights on all observations at time-t. The lowest number N such that $P(N) > 0.50$ for the 50% number and such that $P(N) > 0.90$ for the 90% number is calculated. The calculated number shown in this tableis the number of observations that account for 50% and 90% of the total weights respectively for a Bayesian agent. The ratio of the last N observation weights to the total weights is given by:

$$P(N) = \frac{\sum_{k=t-N+1}^{t} \omega_{t,k}}{\sum_{k=1}^{t} \omega_{t,k}} = \frac{\sum_{k=t-N+1}^{t} \sum_{j=1}^{k} c_{j,k}}{\sum_{k=1}^{t} \sum_{j=1}^{k} c_{j,k}} = \frac{\sum_{k=t-N+1}^{t} \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}}{\sum_{k=1}^{t} \sum_{j=1}^{k} \binom{k-1+j}{k-j} q^{j-1}}$$

The lowest number of observations necessary to account for $100x$% of the total weighting used in the Bayesian update vector is given by:

$$P(N) > x \qquad \text{where } x = 0.50,\ 0.90$$

The following table gives the calculations of the recency effect for each length of time (1, 5, 20 and 40 years).

Number of observations, N, that has at least 50% or 90% of total weighting				
Percent of Weights	12 months (1 year)	60 months (5 years)	240 months (20 years)	480 months (40 years)
q = 1.5				
50%	1	1	1	1
90%	1	1	1	1
q = 2				
50%	1	1	1	1
90%	1	1	1	1