

# What do humans perceive in asset returns?\*

Jasmina Hasanhodzic,<sup>†</sup> Andrew W. Lo,<sup>‡</sup> and Emanuele Viola<sup>§</sup>

March 16, 2014

## Abstract

Both policymakers and academics have recently been studying whether graphical, as opposed to numerical, presentation of financial information to investors affects investors' decisions. Of particular interest is the usefulness of graphical representations of assets' performance via temporal charts of asset prices. Such charts are prevalent in financial media and investment disclosures, and are studied routinely, with the naked eye, by both casual and professional investors. This brings to the forefront a fundamental question: just what information can human beings extract from such charts?

We run several experiments to test if and how human subjects can differentiate time series of actual asset returns from time series that are generated “synthetically” via various processes. Specifically, we consider time series obtained by permuting at random the samples of actual ones, and also those arising from first-order autoregressive (AR1) models.

In contrast with previous anecdotal evidence, we find statistical evidence that subjects can distinguish between actual and synthetic time series. These results show that temporal charts of asset prices convey to investors information that cannot be reproduced by summary statistics. They also provide a first refutation of a strong form of the efficient-market hypothesis that is based entirely on human perception as opposed to computational statistics.

Our experiments are implemented via an online video game (<http://arora.ccs.neu.edu>). We also investigate whether subjects improve their performance while playing.

---

\*We would like to thank Zvi Bodie for pointing out to us Hung, Heinberg, and Yoong (2010), Michael Coen for helpful discussions on a preliminary version of this work, especially on  $p$ -values, and Jason Chen, François Gourio, and Jawwad Noor for helpful feedback and discussions. We are also grateful to Hamid Jahanjou for his help with Amazon Mechanical Turk, and to the anonymous referees for useful feedback.

<sup>†</sup>Boston University Department of Economics, jah@bu.edu.

<sup>‡</sup>MIT Sloan School of Management and AlphaSimplex Group, alo@mit.edu.

<sup>§</sup>Northeastern University College of Computer and Information Science, viola@ccs.neu.edu. Partially supported by a grant from MIT and NSF grants CCF-0845003 and CCF-1319206.

# 1 Introduction

One of the most important and complex decisions individuals face is how to save and invest. Choices they make affect not only their own quality of life, but may have an impact on the economy by creating dependencies on government-sponsored benefits. However, it is well noted that when it comes to investing, individuals are not well positioned to make sound decisions. Several reasons have been proposed in the literature, ranging from overload of information about investment products to choose from, marketing strategies designed to mislead, behavioral biases, and financial illiteracy; see, for example, Bazerman (2001), Bodie (2007), Choi, Laibson, and Madrian (2010), and the references therein. The problem of inadequate individual investment decisions is especially acute in the case of retirement savings, where the recent shift from defined benefit pension plans to privatized 401(k) plans has forced individuals to, in effect, manage their own money. As a result, much debate among policymakers and academics has taken place about improving the quality and presentation of data available to investors. For example Bazerman (2001) and Kozup et al. (2008) call for research on investors perceptions of investment products, and ways of making the information about those products easy to access and comprehend.

An example of recent academic work in this direction is Hung, Heinberg, and Yoong (2010), who evaluate versions of the Department of Labor's proposed Model Comparative Chart, which provides a standard simplified disclosure format for investment information. They conduct an online experiment where subjects are asked to allocate \$10,000 among different funds based on funds performance disclosure. In one version of the disclosure, past returns are presented as a numerical table. In another version, in addition to the numerical table, the disclosure shows a graphical representation of returns over a 10-year period, as a bar chart. For completeness, a relevant figure from their work is reproduced in Figure A.2 in the Appendix at the end of this paper. Perhaps surprisingly, the authors find that the two disclosures have a statistically significant effect on the retirement investment allocation, although the effect may not be practically significant in terms of investment outcomes.

Together with the prevalence of temporal charts of asset returns in financial media such as Yahoo! Finance, and their widespread use by both casual and professional investors, the above brings to the forefront a fundamental question: Just what information can human beings extract from charts of financial returns? This question has several ramifications. For example: Are there any patterns in financial asset returns that humans can actually extract by looking at such charts? Is seeing a chart more informative than just having a few parameters like, say, average and variance? Could Yahoo! and numerous other websites that display charts save space by getting rid of them altogether, with no harm to investors? In Hung, Heinberg, and Yoong's (2010) experiment, is the mere *presence* of some chart biasing the subjects, or are subjects actually gathering information from the contents of the chart?

In this paper we report the results of several experiments designed to test if and how human subjects can differentiate time series of actual asset returns from time series that are generated "synthetically" via various processes. Specifically, we consider time series obtained by permuting at random the samples of actual returns, and also those arising from first-order autoregressive (AR1) models. Our experiments are implemented via an online video game

(<http://arora.ccs.neu.edu>).

The main finding of this paper is that humans can distinguish actual time series from synthetic ones. The results related to random permutations indicate that subjects perceive the temporal order of financial data. The results related to AR1 indicate that subjects are employing more than just first-order autocovariance to differentiate the two time series.

Our findings are in contrast with previous anecdotal evidence. Specifically, it was argued that humans cannot tell price charts from “random,” such as charts generated by a random walk. For example, in an experiment (Malkiel 1973, p. 143) students were asked to generate returns (i.e., price differences) by tossing fair coins, and it was argued that those yielded observations that were indistinguishable from market returns to human subjects observing corresponding price charts. For similar arguments in the finance literature see, for example, Roberts (1959), Kroll, Levy, and Rapoport (1988), DeBondt (1993), Wärneryd (2001), and Swedroe (2005). Such anecdotal evidence has also been collected in the computer science literature. For example, Keogh and Kasetty (2003) “asked 12 professors at UCRs Anderson Graduate School of Management to look at Figure A.1 (included at the end of this paper) and determine which three sequences are random walk, and which three are real S&P500 stocks.” They find that “the accuracy of the humans was 55.6%, which does not differ significantly from random guessing.”

Our results are also of interest in light of the Efficient Market Hypothesis according to which “prices fully reflect all available information” and hence must be unforecastable; see, for example, Samuelson (1965), Fama (1965a), Fama (1965b), and Fama (1970). A strong form of this hypothesis presumes asset returns to be independent and identically distributed, see, e.g., Fama (1970). In this case, it would be impossible to distinguish actual asset returns from a random shuffling of them. But, again, we show that humans can do that.

Note that works such as Lo and MacKinlay (1988, 1999) and Lo, Mamayski, and Wang (2000), provide compelling evidence that markets are not efficient, i.e. price data does possess statistical properties that noticeably deviate from random models. In fact, they show that autocorrelation is such a property. However, we point out that the data analysis in all these works is *computer*, not *human*-based. Consequently, the works leave open the question of whether markets look efficient *to human beings*. Our work appears to be the first to provide such an answer.

We also investigate whether subjects improve their performance while playing. For some contexts, our results mildly indicate that subjects indeed do so.

We note that the idea of testing the ability of human subjects to distinguish random vs. real data using graphical representations is not new. Indeed, this has been studied in depth in the Information Visualization literature, see for example the works by Heer, Kong, and Agrawala (2009), and Wickham, Cook, Hofmann, and Buja (2010), and the references therein. However, we are unaware of any previous work where this idea has been used in a financial setting.

Similarly, we do not view the video game we developed as a main contribution of this paper. This game displays data in a fashion similar to commonly used trading platforms; and similar tools are for example reviewed in the Information Visualization papers just cited. Instead, implementing the experiment as a video game is intended to make the process fun and engaging for the subjects, so that they do not get tired, bored, or frustrated in a way that might affect their behavior. Moreover, the game allows the subjects to make their choices quickly, allowing us to get a large amount of

data efficiently, with as little cost to subjects as possible.

This paper is organized as follows. In Section 2 we describe our experiment and the data we used. In Section 3 we describe the synthetic processes and our results. Section 4 investigates whether subjects improve their performance while playing. We conclude in Section 5.

## 2 Experiment Design

We develop a simple web-based video-game, available at <http://arora.ccs.neu.edu>. In this game, subjects are shown two dynamic price series (i.e., moving charts) side by side—both of which display price graphs evolving in real time (a new price realized roughly each second)—but only one of which is a “replay” of actual historical price series. The other series is constructed via a synthetic process. See Figures 1 and 2. Subjects are asked to press a button indicating their selection of the actual price series, and are informed immediately whether they were correct or incorrect, see Figures 3 and 4, after which the next pair of price series begins being displayed. Note that the charts are moving, so at any point in time there is a certain number of observations present on the screen for each time series, which is a subset of the total number of observations subjects see on a moving chart before having to make a guess (these parameters are reported for each dataset later in the paper). Subjects do not have to wait until the entire moving chart is completed being displayed before making their choice, but can guess at any time prior to its completion (an omnipresent counter informs them of the time left). They have a counter telling them how many seconds they have remaining before the moving chart is done. The game is fast-paced: subjects can observe the charts for 10 to 25 seconds (depending on the dataset) before having to make a guess.<sup>1</sup>

For the actual time series we used eight datasets consisting of returns of commonly traded financial assets. These datasets were arbitrarily named after animals, so that users had no knowledge of the specific financial assets used in the experiment. Figure 5 summarizes the data used. It also reports how many charts were shown to each subject, how many data points constitute a chart, and, since charts are moving, how many points of the chart fit onto the screen in any given time. The Dow Jones Corporate Bond Price Index was obtained from the Global Financial Database, while all other data series were obtained from Bloomberg.

Subjects were recruited with Amazon Mechanical Turk. To register, a subject has to fill a short demographic questionnaire where they were asked to select one of prespecified categories describing occupation (academic, finance, student, other), sex (male, female), education (high school, undergraduate, MS, PhD), age, and country.

After registration, a subject can participate in trials from eight different contests, each consisting of the same game applied to different datasets.

Participating in a trial consists of the following task. The subject is shown two dynamic price charts on a computer screen, one above the other (Figures 1 and 2). Each graph evolves through time—similar to those appearing in computer trading platforms—plotting the price at that point in time as well as the trailing prices over a fixed time window over the most recent past. Prices

---

<sup>1</sup>Jumping ahead, we used 8 different datasets named after animals. Subjects are given 11 seconds to guess in the Bull contest, 15 in Bear, Elk, and Raindeer, 20 in Lynx and Mandrill, 22 in Seal, and 24 in Beaver.

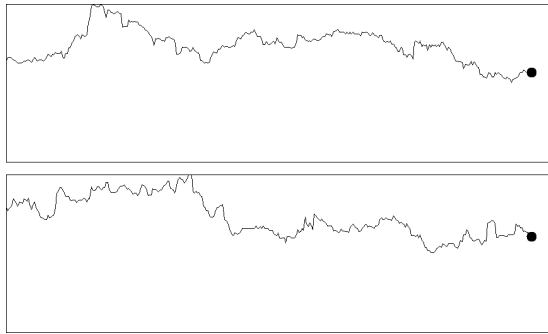


Figure 1: Reindeer (real data in top panel).

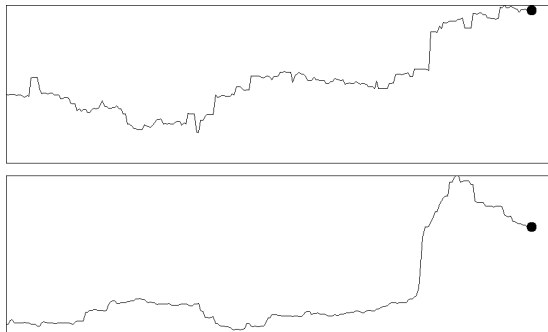


Figure 2: Bear (real data in bottom panel).



Figure 3: Wrong choice in Beaver contest.



Figure 4: Correct choice in Elk contest.

| Contest  | data  | points<br>on screen | charts<br>per subject | points<br>per chart |
|----------|---|---------------------|-----------------------|---------------------|
| Reindeer | Gold Spot Price<br>tick data (1-60 sec.)<br>Jun-Oct 2009                | 350                 | 40                    | 500                 |
| Elk      | US Dollar Index<br>tick data (about 5 sec.)<br>May-Sep 2009             | 350                 | 40                    | 500                 |
| Lynx     | Canada / U.S. Foreign Exchange Rate<br>daily data<br>Aug 1978-Apr 2009  | 90                  | 35                    | 190                 |
| Mandrill | S&P GSCI Corn Index Spot<br>daily data<br>Jun 1982-Oct 2009             | 38                  | 50                    | 125                 |
| Bull     | Russell 2000 Index<br>tick data (about 10 sec.)<br>May-Dec 2009         | 110                 | 31                    | 153                 |
| Bear     | Nasdaq Composite Index<br>tick data (about 1 sec.)<br>May-Jul 2009      | 250                 | 37                    | 400                 |
| Beaver   | Dow Jones Industrial Average<br>daily data<br>Sep 1926-May 2009         | 300                 | 36                    | 500                 |
| Seal     | Dow Jones Corporate Bond Price Index<br>daily data<br>Jan 1941-Apr 2009 | 250                 | 39                    | 400                 |

Figure 5: Datasets

are defined as the cumulative sum of a sequence of returns. Of the two moving charts, only one corresponds to the sequence of market returns from the actual dataset; we call this graph the “real” chart or  $\{p_t\}$ . The other corresponds to a “synthetic” sequence of returns, as described in Section 3. We call this graph the “synthetic chart” or  $\{p_t^*\}$ . The computer chooses at random which of the two graphs is placed at the top or the bottom.

The subject is asked to decide which of the two moving charts is the real one by clicking on it. The game registers the subject’s choice, and informs the subject immediately whether his/her guess is correct or incorrect, see Figures 3 and 4. For each dataset, the user is shown approximately 35 pairs of moving charts and asked to make as many choices. The subject is also free to refrain from choosing. This happened rarely, and to err on the conservative side, we recorded the absence of a guess as an incorrect choice for that trial.

To evaluate the robustness of our experimental design, we varied various parameters of the experiment across datasets, as indicated in Figure 5. In addition, we presented subjects with data charts in two different ways. For half of the datasets corresponding to transaction-by-transaction (or “tick”) data, each subject was shown a fresh set of charts, based on a sequence of returns disjoint from the sequences shown to any other subjects. For the other half of the data, corresponding to daily data, the charts shown to each subject were based on the same sequence of returns.<sup>2</sup>

Finally, for each dataset, subjects were offered the opportunity to train on a disjoint set of data.

### 3 Synthetic processes

In this section we describe the various synthetic processes we considered, and the corresponding results. We start with random shuffling and then AR1.

#### 3.1 Random shuffling

Here we want to test the null hypothesis  $H$  that human subjects cannot distinguish between actual time series and a time series that is obtained by permuting at random the entries of the actual one. Details follow.

We begin with a time series of actual historical prices  $\{p_0, p_1, p_2, \dots, p_T\}$  and compute the logarithmic returns  $\{r_t\}$ ,

$$r_t \equiv \log(p_t) - \log(p_{t-1}). \quad (1)$$

From this, we construct a randomly generated price series  $\{p_0^*, p_1^*, \dots, p_T^*\}$  by cumulating randomly permuted returns:

$$p_{t+1}^* \equiv p_t^* \cdot e^{r_{t+1}} \quad , \quad p_0^* \equiv 1 \quad , \\ \pi(k) : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$$

where  $\pi(k)$  is a uniform permutation of the set of time indexes  $\{1, \dots, T\}$ . A random permutation of the actual returns does not alter the marginal distribution of the returns, but it does destroy

---

<sup>2</sup>However, the data was shifted by a random amount for security reasons, i.e., to avoid the possibility that two subjects could coordinate their guesses, for example by simultaneously playing the same charts on two nearby machines.

| Contest  | subjects | p-value | correct guesses per subject  |
|----------|----------|---------|--|
| Mandrill | 56       | 0.00484 | 25 35 24 27 29 24 32 23 17 27 19 20 25 24 25 30 28 25 27<br>32 31 20 24 32 23 31 24 23 34 28 26 28 31 28 26 19 27 29<br>18 29 32 20 31 23 27 28 34 23 26 25 27 33 30 21 28 30          |
| Bear     | 55       | 0.00000 | 29 30 32 27 28 24 28 34 29 16 21 28 28 29 33 17 35 32 20<br>21 28 17 22 27 30 32 24 32 27 31 17 31 23 29 17 30 32 22<br>25 24 33 30 34 30 16 34 36 24 19 18 16 36 19 21 25             |
| Lynx     | 56       | 0.00022 | 11 20 21 17 21 20 21 20 19 17 21 15 17 19 23 18 17 21 17<br>17 16 18 17 24 19 18 21 17 19 22 20 19 19 16 21 18 20 24<br>19 10 19 23 21 19 22 20 23 13 20 19 23 18 17 16 17 20          |
| Reindeer | 56       | 0.00000 | 21 25 30 37 38 18 26 34 25 34 31 15 25 26 27 14 19 21 28<br>18 25 22 29 31 24 28 31 19 24 28 37 28 20 28 18 25 29 26<br>25 24 32 30 21 24 24 29 17 17 25 17 23 38 28 21 24 22          |
| Beaver   | 58       | 0.00007 | 21 21 23 19 26 17 24 19 15 16 15 22 15 23 20 19 13 22 15<br>24 23 19 20 16 20 26 19 26 18 26 18 20 19 23 20 17 18 17<br>20 27 16 18 27 28 27 19 21 22 16 18 18 20 25 14 21 14 25<br>15 |
| Bull     | 57       | 0.00000 | 25 30 26 30 31 25 26 31 20 19 28 31 28 26 31 31 31 23 12<br>17 30 15 29 31 9 16 30 19 17 31 30 30 31 31 31 25 30 31<br>28 22 31 31 29 30 31 21 31 30 11 16 12 19 28 23 9 31 30         |
| Elk      | 58       | 0.00000 | 34 29 27 28 27 27 24 33 19 36 31 33 32 21 34 18 38 16 14<br>33 12 32 37 19 30 32 28 18 29 24 33 26 33 33 35 18 33 38<br>40 37 30 31 36 38 31 33 32 32 24 21 20 22 37 29 27 14 34<br>33 |
| Seal     | 55       | 0.00000 | 30 28 33 21 32 25 25 33 33 24 25 31 15 33 36 23 21 24 30<br>15 17 33 33 25 34 34 18 29 28 21 31 33 32 27 35 35 33 26<br>32 29 33 31 31 25 33 33 27 20 20 22 33 25 23 28 35             |

Figure 6: Results for distinguishing price charts from their permutation

the time-series structure of the original series, including any temporal patterns contained in the data. Therefore, the randomly permuted returns will have the same mean, standard deviation, and moments of higher order as the actual return series, but will not contain any time-series patterns that may be used for prediction. This construction will allow us to test specifically for the ability of human subjects to detect temporal dependencies in financial data.

The results are reported in Table 6. The null hypothesis is refuted for all eight datasets:  $p$ -value is always less than 0.5%.



| Contest  | subjects | p-value |  |
|----------|----------|---------|--|
| Mandrill | 17       | 0.0577  | 19 21 23 23 24 25 25 26 27 29 29 30 30 32 32 33 35   |
| Bear     | 29       | 0.0000  | 37 26 29 37 34 6 33 32 35 29 37 32 36 31 30 35 37 30 26<br>27 20 27 27 34 29 33 33 19 21                             |
| Lynx     | 26       | 0.0016  | 15 15 16 16 16 16 17 17 18 18 19 19 20 20 20 21 21 22 22<br>22 22 22 22 22 23 25                                     |
| Reindeer | 22       | 0.0000  | 40 16 27 38 30 29 32 27 31 38 31 25 30 27 30 21 25 29 31<br>34 21 15   |
| Beaver   | 23       | 0.0033  | 12 13 15 18 18 19 19 19 19 20 20 21 23 23 23 23 23 24<br>24 25 29 30   |
| Bull     | 32       | 0.0000  | 31 29 29 31 25 31 30 30 26 29 30 28 31 31 29 30 31 31 31<br>26 30 30 31 29 30 31 29 31 31 31 30 14                   |
| Elk      | 25       | 0.0000  | 40 18 31 35 36 35 22 30 37 36 33 40 33 36 36 35 40 36 38<br>36 35 37 40 34 21  |
| Seal     | 38       | 0.0000  | 16 17 18 19 21 23 23 23 24 24 25 25 25 26 26 27 27 28 29<br>29 29 30 30 31 31 31 31 32 32 32 32 33 33 34 35 35 36 39 |

Figure 7: Results for distinguishing price charts from their permutation – variant.

### 3.2 A variant

To evaluate the robustness of the results we also considered the following variant of the process, where returns are simply obtained via price differences:

$$r_t \equiv p_t - p_{t-1}.$$

From this, we construct a randomly generated price series  $\{p_0^*, p_1^*, \dots, p_T^*\}$  by cumulating randomly permuted returns:

$$p_t^* \equiv \sum_{k=1}^t r_{\pi(k)} \quad , \quad p_0^* \equiv p_0 \quad ,$$

$$\pi(k) : \{1, \dots, T\} \rightarrow \{1, \dots, T\}.$$

To recruit subjects for this variant, an announcement was emailed to Northeastern computer science students, MIT Sloan MBA students in the Fall section of 15.970, members of the American Association of Individual Investors mailing list, Market Technicians Association mailing list, the MTA Educational Foundation mailing list, and the staff and Twitter followers of TraderPsyches.

Results for this variant are reported in Table 7. The  $p$ -value is less than 5% for all but eight datasets. We attribute the slightly less decisive outcome for this variant to the smaller number of subjects.

### 3.3 AR1

Here we want to test the null hypothesis  $H$  that human subjects cannot distinguish between an actual time series  $S$  and a time series that is generated by an AR1 process that is calibrated to

match mean, variance, and (first-order) autocovariance of  $S$ . Details follow. We refer to Section 3.4 of Hamilton (1994) for background on AR1 processes.

Again we begin with a time series of actual historical prices  $\{p_0, p_1, p_2, \dots, p_T\}$  and compute the logarithmic returns  $\{r_t\}$ ,

$$r_t \equiv \log(p_t) - \log(p_{t-1}). \quad (2)$$

Then we compute the sample mean  $\mu$ , variance  $v$ , and (first-order) autocovariance  $\alpha$  of the series  $r$ . This defines an AR1 process

$$y_t := c + \phi \cdot y_{t-1} + \epsilon_t,$$

where  $\epsilon_t$  are i.i.d. normal distributions with mean 0 and variance  $\sigma^2$ , as follows:

$$\begin{aligned} \phi &= \alpha/v, \\ c &= \mu(1 - \phi), \text{ and} \\ \sigma^2 &= v(1 - \phi^2). \end{aligned}$$

The starting point  $y_0$  of the AR1 process is taken to be  $r_h$  for an index  $h$  chosen uniformly at random.

And finally we set

$$p_{t+1}^* \equiv p_t^* \cdot e^{y_{t+1}} \quad , \quad p_0^* \equiv 1.$$

The results are reported in Table 8. We obtain a  $p$ -value less than 0.5% for 5 of our 8 data sets, and higher for the other three.

## 4 Learning

As our game provides feedback, we investigate whether subjects improve their performance while playing. We do so by comparing the subjects' performance in the first and the last part of each contest. Specifically, for each contest, we consider the subset consisting of the first  $\alpha = 1/5$  fraction of guesses, and that consisting of the last  $\alpha$  fraction. For each subset, we add up the number of correct guesses across subjects and divide that sum by the total number of guesses in the subset times the number of subjects. We call this the fraction of correct guesses made by the combined pool of subjects.

Figure 9 reports the results for the shuffling process. In all but one dataset the average number of correct guesses increases. In some cases, as for Mandrill and Elk, by a seemingly significant amount.

Figure 10 reports the results for the AR1 process. Again, in some cases, like Reindeer or Bull, the difference in the averages seems significant.

The results are robust with respect to other choices of  $\alpha$ . Recall also that in our experiment subjects are required to practice before entering a contest. This makes the results in this section less prone to be influenced by extraneous factors such as becoming comfortable with the interface.

| Contest  | subjects | p-value | correct guesses per subject   |
|----------|----------|---------|---|
| Mandrill | 36       | 0.27891 | 26 23 20 22 24 27 29 18 29 32 24 28 30 22 18 25 21 28 25<br>29 26 28 25 25 23 30 22 28 24 27 24 29 25 22 27 27                |
| Bear     | 40       | 0.00000 | 37 22 32 36 18 23 37 20 10 27 20 33 27 34 18 29 22 29 33<br>33 35 26 23 34 27 36 35 16 32 35 20 37 28 21 34 33 28 34<br>36 37 |
| Lynx     | 38       | 0.07185 | 19 21 18 15 11 16 14 20 16 22 20 19 18 20 14 21 20 23 21<br>20 17 23 13 15 13 20 13 18 19 24 21 22 18 19 18 18 16 19          |
| Reindeer | 39       | 0.00016 | 23 19 36 23 25 22 17 23 21 25 28 18 16 17 19 18 18 17 18<br>27 19 18 36 23 28 22 17 23 36 17 28 31 33 29 28 22 32 28<br>26    |
| Beaver   | 37       | 0.05353 | 18 21 14 26 17 23 25 25 21 17 20 20 16 15 20 20 22 16 19<br>21 16 17 17 25 19 19 18 21 10 16 19 12 12 21 20 24 22             |
| Bull     | 37       | 0.00000 | 24 20 20 24 24 17 25 18 13 24 25 30 25 27 17 19 30 30 27<br>15 28 23 19 27 18 28 22 17 24 20 17 22 22 27 26 25 27             |
| Elk      | 36       | 0.00249 | 21 22 25 18 16 25 19 17 25 19 19 25 18 19 19 27 19 25 35<br>24 18 23 25 28 19 26 23 22 24 15 20 22 19 21 30 23                |
| Seal     | 38       | 0.00000 | 26 20 23 26 23 27 25 18 23 26 33 27 31 22 22 19 27 32 23<br>31 24 13 32 17 31 29 24 30 26 19 24 22 31 27 30 29 23 31          |

Figure 8: Results for distinguishing price charts from AR1

| Contest  | number of guesses | first   | last    |
|----------|-------------------|---------|---------|
| Mandrill | 557               | 0.50628 | 0.58887 |
| Bear     | 384               | 0.69271 | 0.73438 |
| Lynx     | 391               | 0.57289 | 0.58824 |
| Reindeer | 445               | 0.62022 | 0.65618 |
| Elk      | 461               | 0.68980 | 0.75488 |
| Bull     | 342               | 0.79825 | 0.82456 |
| Beaver   | 406               | 0.55911 | 0.59360 |
| Seal     | 385               | 0.78701 | 0.59221 |

Figure 9: Performance improvement with the shuffling process. For each contest, the column “first” reports the fraction of correct guesses made by the combined pool of subjects in the first one-fifth of guesses. The column “last” reports the corresponding value for the last one-fifth of guesses. The column “number of guesses” is the denominator in the calculation of these fractions of correct guesses.

| Contest  | number of guesses | correct first | correct last |
|----------|-------------------|---------------|--------------|
| Mandrill | 359               | 0.49582       | 0.48189      |
| Bear     | 280               | 0.73929       | 0.83214      |
| Lynx     | 265               | 0.52453       | 0.53585      |
| Reindeer | 312               | 0.50962       | 0.63462      |
| Elk      | 287               | 0.54355       | 0.56794      |
| Bull     | 221               | 0.67421       | 0.81900      |
| Beaver   | 258               | 0.54264       | 0.50000      |
| Seal     | 266               | 0.67669       | 0.51880      |

Figure 10: Performance improvement with the AR1 process. For the meaning of the columns see the caption of Figure 9.

## 5 Conclusion

A natural question that arises is how were the subjects able to perform so well in seven out of eight datasets? Casual inspection of Figures 1–4 shows that distinguishing real data from synthetic data is challenging; for some datasets the real chart tends to be smoother, as in Figure 2, while for other datasets the opposite is true, the real chart tends to be spikier, as in Figure 4. What complicates the matter further is that, as is evident from the data, the “smoothness” of actual data varies with time. Still, feedback from just a few trials seems sufficient for the user to extract characteristics of the data to be used in classifying charts in the near future.

The importance of feedback is supported by the information about winning strategies that some of the subjects volunteered to share with us (anonymously). For example, a subject wrote:

Admittedly, when first viewing the two datasets in the practice mode, it is impossible to tell which one is real, and which one is random, however, there is a pattern that quickly emerges and then the game becomes simple and the human eye can easily pick out the real array (often in under 1 second of time).

For some contests, our results mildly suggest that indeed subjects improve while playing.

Having shown that human beings can distinguish price charts from charts obtained by permuting the returns at random, an important next step is to understand what properties of the data the subjects exploited. Our results related to AR1 processes indicate that subjects use more than just first-order autocovariance of the data. Another interesting direction is to compare humans’ performance in our experiment against the performance of computers. The human eye—as opposed to a computer algorithm—may have a crucial advantage. It is well known that computers still struggle with many image-recognition and classification tasks that are trivial for humans. The same may be the case for distinguishing asset returns from synthetic processes.

Given the recent regulatory push towards ensuring that “consumers have the information they need to choose the consumer financial products and services that are best for them,”<sup>3</sup> the study of

---

<sup>3</sup>The Consumer Financial Protection Bureau, <http://www.consumerfinance.gov/protecting-you/>.

optimal ways to present financial data to investors is of current interest. Our paper is a contribution to the growing body of literature on the usefulness of temporal charts in evaluation of financial asset performance.

## References

- [1] M. H. Bazerman. Consumer research for consumers. *Journal of Consumer Research*, 27(4):499–504, 2001.
- [2] Z. Bodie. The challenge of investor education. In Z. Bodie, D. McLeavey, and L. B. Siegel, editors, *The Future of Life-Cycle Saving and Investing*, pages 169–171. Research Foundation of the CFA Institute, February 2008.
- [3] J. Choi, D. Laibson, and B. Madrian. \$100 bills on the sidewalk: Suboptimal saving in 401(k) plans. *Review of Economics and Statistics*, 2010.
- [4] W. P. M. De Bondt. Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting*, 9(3):355–371, 1993.
- [5] E. Fama. The behavior of stock market prices. *Journal of Business*, 38(1):34–105, 1965.
- [6] E. Fama. Random walks in stock market prices. *Financial Analysts Journal*, 21:55–59, 1965.
- [7] E. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.
- [8] J. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.
- [9] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *27th International Conference on Human Factors in Computing Systems (CHI)*, pages 1303–1312, 2009.
- [10] A. A. Hung, A. Heinberg, and J. K. Yoong. Do risk disclosures affect investment choice? Technical report, RAND Labor and Population, September 2010.
- [11] E. J. Keogh and S. Kasetty. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371, 2003.
- [12] J. Kozup, E. Howlett, and M. Pagano. The effects of summary information on consumer perceptions of mutual fund characteristics. *Journal of Consumer Affairs*, 42(1):37–59, Spring 2008.
- [13] Y. Kroll, H. Levy, and A. Rapoport. Experimental tests of the mean-variance model for portfolio selection. *Organizational Behavior and Human Decision Processes*, 42(3):388–410, 1988.

- [14] A. Lo, H. Mamaysky, and J. Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, LV(4):1705–1765, August 2000.
- [15] A. W. Lo and A. C. MacKinlay. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1):41–66, 1988.
- [16] A. W. Lo and A. C. MacKinlay. *A Non-Random Walk Down Wall Street*. Princeton University Press, Princeton, NJ, 1999.
- [17] B. G. Malkiel. *A Random Walk Down Wall Street*. W. W. Norton & Company, 2003.
- [18] H. V. Roberts. Stock-market ‘patterns’ and financial analysis: Methodological suggestions. *The Journal of Finance*, 14(1):1–10, 1959.
- [19] P. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49, 1965.
- [20] L. E. Swedroe. *The Only Guide to a Winning Investment Strategy You’ll Ever Need*. St. Martin’s Press, 2005.
- [21] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.
- [22] K. E. Wärneryd. *Stock-market psychology: How people value and trade stocks*. Edward Elgar, 2001.
- [23] H. Wickham, D. Cook, H. Hofmann, and A. Buja. Graphical inference for infovis. *IEEE Trans. Vis. Comput. Graph.*, 16(6):973–979, 2010.

# A Appendix

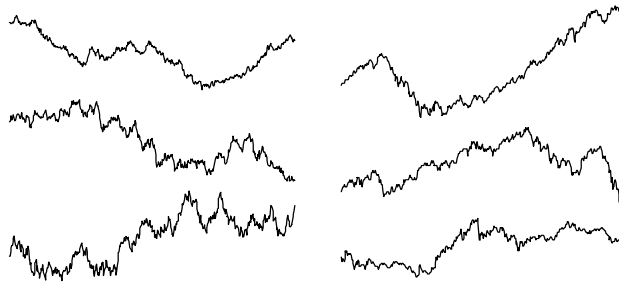


Figure 1. Six time series, three are random walk data, and three are real S&P500 stocks. Experiments suggest that humans cannot tell real and synthetic stock data apart (all the sequences on the right are real).

Table 1. The confusion matrix for human experts in attempting to differentiate between random walk data and stock market data.


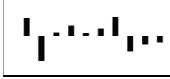

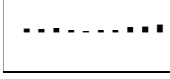
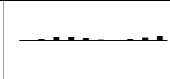

|        |             | Predicted |             |
|--------|-------------|-----------|-------------|
|        |             | S&P Stock | Random Walk |
| Actual | S&P Stock   | 20        | 16          |
|        | Random Walk | 16        | 20          |

Figure A.1: From Keogh and Kasetty (2003).

**Modified Comparative Chart, Version 2  
(Bar Chart of Annual Total Returns)**

**Performance Information**

This table describes investment options that provide variable rates of return. This table shows each option's performance over several time periods and compares the performance with a recognized benchmark. For options with returns that vary over time, past performance does not guarantee how your investment in the option will perform in the future; your investment in these options could lose money.

| Table--Variable Return Investments |   |   |       |       |                    |                                     |       |       |                                     |
|------------------------------------|---|---|-------|-------|--------------------|-------------------------------------|-------|-------|-------------------------------------|
| Name/<br>Type of Option            | Graph:<br>2000-2009<br>Year-End Total<br>Returns*                                   | Average Annual Total Return<br>as of 06/30/2010 |       |       |                    | Benchmark/Index<br>as of 06/30/2010 |       |       |                                     |
|                                    |   | 1yr.  | 5yr.  | 10yr. | Since<br>Inception | 1yr.                                | 5yr.  | 10yr. |                                     |
| <b>Stock Funds</b>                 |   |   |       |       |                    |                                     |       |       |                                     |
| Small Cap Stock<br>Index<br>Fund A |    | 22.7%   | 2.2%  | 3.7%  | 5.0%<br>05/21/98   | 22.9%                               | 2.2%  | 3.6%  | Spliced Small Cap Stock<br>Index    |
| Stock Market<br>Index<br>Fund B    |   | 14.3%   | -0.9% | -1.7% | 10.1%<br>08/31/76  | 14.4%                               | -0.8% | -1.6% | S&P 500<br>Index                    |
| Global Equity<br>Fund C            |  | 14.5 %  | 0.2%  | 4.8%  | 7.3%<br>08/14/95   | 11.8%                               | 1.3%  | 0.0%  | Spliced Global Equity<br>Index      |
| <b>Bond Funds</b>                  |   |   |       |       |                    |                                     |       |       |                                     |
| Bond Market<br>Index<br>Fund D     |  | 9.3%  | 5.5%  | 6.2%  | 6.9%<br>12/11/86   | 9.5%                                | 5.5%  | 6.5%  | Barclays US Aggregate<br>Bond Index |
| <b>Other</b>                       |   |   |       |       |                    |                                     |       |       |                                     |
| Money Market<br>Fund E             |  | 0.00%   | 2.7%  | 2.6%  | 3.5%<br>12/14/92   | 0.0%                                | 2.0%  | 2.1%  | Avg. Money Market<br>Treasury Fund  |
| Balanced Index<br>Fund F           |  | 13.6%   | 2.4%  | 2.4%  | 7.3%<br>11/9/92    | 13.8%                               | 2.4%  | 2.5%  | Balanced Composite<br>Index         |

\* The bar charts show the changes in each fund's performance from year to year.

Figure A.2: From Hung, Heinberg, and Yoong (2010).