Xiao-Ping (Steven) Zhang
and Fang Wang

Signal Processing for Finance, Economics, and Marketing

Concepts, framework, and big data applications

Economic data and financial markets are intriguing to researchers working on data and quantitative models. With rapid growth of and increasing access to data in digital form, finance, economics, and marketing data are poised to become one of the most important and tangible big data applications, owing not only to the relatively clean organization and structure of the data but also to clear application objectives and market demands. However, data-related economic studies often have different viewpoints from signal processing (SP). Also, many fundamental economics and business problems have been well formulated and studied in both theory and practice. The knowledge of foundational finance and economic theories will help SP and data researchers avoid reinventing the wheel and develop meaningful and useful research in these areas.

This tutorial article intends to introduce the mainstream foundational concepts and framework in finance, economics, and marketing research, elaborate on the relationships between the traditional economic research paradigm and SP methodology, and help SP researchers identify relevant research directions. The article aims to present a refreshing SP perspective of finance, economics, and marketing research as well as in-depth examples on SP applications in these fields. We hope to empower SP researchers to broaden their knowledge beyond their current areas of expertise and quickly grasp the right formulations of the research questions and related evaluation criteria in these fields.

Digital Object Identifier 10.1109/MSP.2017.2663138
Date of publication: 26 April 2017

Introduction

For SP researchers, comparing an economic system, either a financial market or a consumer market, with a familiar physical input–output system with an impulse–response relationship is intriguing. However, it is important to understand the different and unique perspectives of an economic system. An economic system is an open-loop system in which humans are active participants. System models to be used in analysis are not known a priori and for certain. In economic studies, researchers always work with assumptions or hypotheses that cannot be verified using physical or natural laws. In addition, it is difficult to perform controlled experiments as is usually done in SP systems.

Increasing amounts of digital market data are readily available, due to the rapid growth of electronic trading along with the development of broadband Internet. Electronic trading platforms can record every bid and ask as well as every high-frequency transaction. Indeed, finance, economics, and marketing data are poised to become one of the most important and tangible big data applications not only because of the relative clean organization and structure of the data but also because of clear application objectives and market demands.

For example, owing to its familiar appearance, financial time series has become a major interest for SP researchers, as is evident by recent special issues [1]–[3] on related topics.

Many SP researchers are eager to apply SP techniques to stock price prediction or profitable trading strategy by analyzing price time series. However, without economic justifications and rigorous out-of-sample tests, such analysis may easily fall into the category of technical analysis, which is not mainstream in finance academia or among financial professionals.

The knowledge of the major streams of finance theories is very useful for SP researchers to position and formulate their research before applying SP techniques to financial data analysis. Economic behaviors of people are often about individual choices. The foundational theory (or hypothesis) of economic choices is the expected utility theory, which explains the relationship between investment risk and expected return. Almost all asset-pricing models are based on the expected utility theory. To understand the aggregated market behavior, a fundamental and insightful theory in finance and economics research, the efficient market hypothesis (EMH) [4], was developed. The EMH assumes that all information is incorporated into price and that the market reaches equilibrium through rational decisions of market participants. Stock returns are compensation for risks investors take in markets. The fundamental asset-pricing model and portfolio theory are formulated on the basis of the EMH [5]. Overwhelming empirical evidence supports the EMH, suggesting that it holds most of the time.

Although the EMH is a foundational theory in academic finance research, most stock analyses used by financial practitioners, such as technical analysis [6], which is popular among individual investors, and fundamental analysis, which is popular among professional investors, contradict the EMH [4] because the current price does not reflect all available market information. Instead, these analyses share some common ground with behavioral economics theories [7]–[9]. Although powerful, behavioral economics theories rely on empirical psychological and qualitative arguments to explain economic events and have not led to fully developed asset-pricing models. In contrast with the EMH-based models, it is often difficult to develop rigorous mathematical formulations for psychological effects, making SP methodology difficult to apply to behavioral economics studies.

The knowledge of extant data techniques, results, and issues in finance research will greatly help SP researchers to exert useful research effort by identifying correct problem formulations more effectively. For example, SP researchers are often interested in market correlation structures, an area where modern asset-pricing and portfolio theories [5] have well formulated the correlation structure within stock markets under the normality assumption. In contrast with SP research in which controlled experiments or simulations are conducted and ground truth is often the basis for testing and verifying the models, model testing and verification in finance pose many traps and difficulties.

Nevertheless, with the tremendous increase in the amount of economic data in digital form, the demand for applying SP techniques to finance, economics, and marketing research is increasing, presenting huge opportunities to SP researchers. Many economics applications (including finance and marketing)

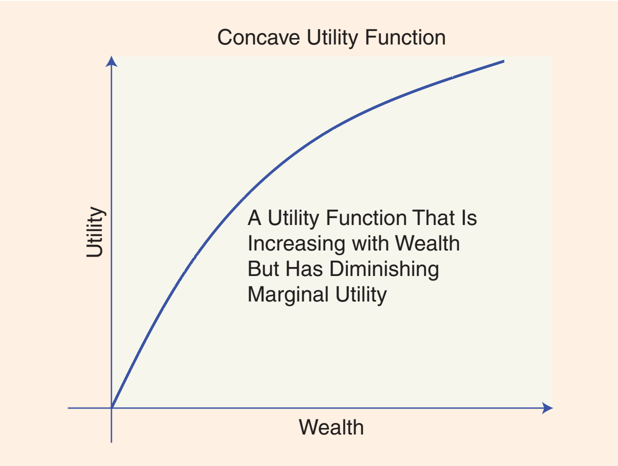

FIGURE 1. The expected utility theory assumes that people make choices according to the expected utility $\mathbb{E}[U(W)]$.

beyond stock price prediction, such as input–output relationships in certain economic systems [10]–[12], are available, and research questions await answers. The recent surge in high-frequency trading (HFT) practices and related theoretical studies [13] provides further opportunities for SP researchers to examine market microstructures and high-frequency system response. Indeed, many SP models and methods share common mathematical grounds with traditional econometric analysis [14] but present different analytical aspects. Therefore they can provide new tools for economic system modeling, analysis, and information extraction for massive finance and economic data.

This tutorial article intends to concisely introduce the mainstream foundational concepts and frameworks in finance, economics, and marketing research to SP researchers, with conceptual elaboration of the relationships between traditional economic research paradigms and SP methodology, and help SP researchers understand economics and business literature and identify relevant research directions with economic significance.

First, we introduce the risks and the foundational economics theory. Then, we formulate the fundamental market equilibrium asset-pricing model in finance within the parameters of modern portfolio theory. Equipped with a basic understanding of the minimum set of economics theories and principles, we introduce an economic viewpoint and basic hypotheses within the context of a free and competitive market—the EMH and competing behavioral economics along with the prospect theory. We elaborate on how different economics theories factor in the individual person—the center of any economic system—his or her choices, decision processes, and emotions. We present a philosophical analysis of the test on economic models with a data–joint hypothesis test. SP perspectives are provided across the article to help readers quickly grasp the differences.

We then move to introduce the fundamental econometrics models and time-series analysis in comparison to the parallel tools known in SP. We focus more on the basic concepts that are rarely present in SP but that are critical in analyzing time-series data in economics and business applications, such as unit roots and causality. We then briefly summarize the relationships between SP and econometric models so that SP researchers can apply their SP knowledge to quickly take advantage of econometrics models. These fundamental economics concepts and methodology that we will introduce have won Nobel Prizes from 1990 to 2013. They are by no means comprehensive, but they encompass a skeleton and basic set of building blocks for data-based economic studies.

We also provide a few detailed state-of-the-art examples in reference to applying SP to economics, finance, and marketing studies. Furthermore, we focus on a few illustrative formulations of economic and business problems. In addition, we give an empirical data analysis example to demonstrate the insight of economic systems that SP is poised to make significant contributions. Throughout this article, we use sidebars to present mathematical formulations and examples to further clarify and illustrate main concepts and ideas.

Risk, risk premium, portfolio optimization, and capital asset pricing

In this section, we introduce fundamental concepts that serve as language and building blocks for economics and finance theory. We begin by introducing the expected utility theory and risk premium (RP).

Expected utility theory and RP

The expected utility theory (or hypothesis) is a cornerstone for economics, game theory, and decision theory and pertains to people's preferences and choice. In economics, a utility function $U(w)$ is defined as a concave function of overall wealth w. An example is shown in Figure 1. Such a utility function assumes that 1) utility is increasing with wealth, i.e., $U(w)$ is monotonically increasing, and 2) wealth has diminishing marginal utility, represented by the concavity of $U(w)$. Therefore, $dU(w)/dw > 0$, and $d^2U(w)/dw^2 < 0$. The expected utility theory further assumes that people make rational choices according to the overall expected utility $\mathbb{E}[U(W)]$. Here, W represents a random variable.

Now, assume that a risk asset (e.g., a stock or an investment) has two possible outcomes of wealth with equal probability in the future, w_0 and w_1 and $w_0 < w_1$. The expected wealth is

$$\mathbb{E}[W] = \frac{1}{2}(w_0 + w_1). \tag{1}$$

The expected utility of this risk asset is

$$\mathbb{E}[U(W)] = \frac{1}{2}(\mathbb{E}[U(w_0)] + \mathbb{E}[U(w_1)]). \tag{2}$$

Given the concavity of the utility function, we have $\mathbb{E}[U(W)] < U(\mathbb{E}[W])$. That is, a rational person is risk averse and thus would prefer a riskless asset, such as cash, with guaranteed wealth of $\mathbb{E}[W]$, to the risk asset with expected wealth of the same $\mathbb{E}[W]$. The certainty equivalent (CE) represents guaranteed wealth whose utility is equivalent to that of the risk asset; i.e.,

$$CE = U^{-1}(\mathbb{E}[U(W)]). \tag{3}$$

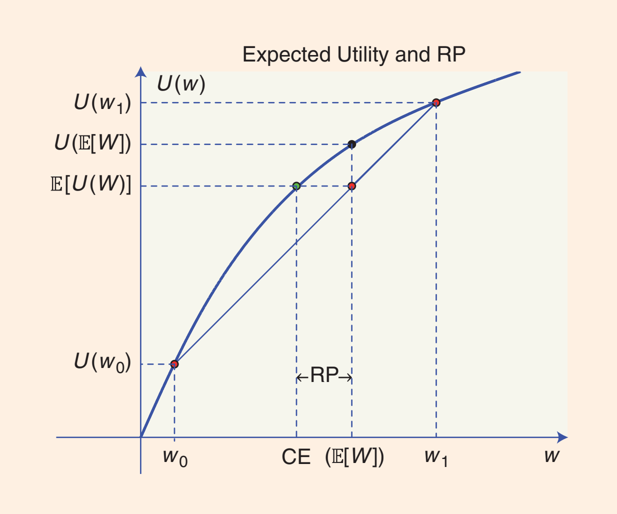

FIGURE 2. The relationship among expected wealth $\mathbb{E}[W]$, expected utility $\mathbb{E}[U(W)]$, CE, and RP.

Take a Bet: Source of the Excess Return

Assume you win the lottery and have the following two choices:

1) the bet: flip a coin (equal probability on both sides); heads, you get US$2,000, tails, you get US$0
2) cash US$1,000.

Would you take the bet or the cash? What if the cash amount were changed to US$900? US$800? Or more? The cash amount for which you are indifferent between choices 1) and 2) is the CE for the risky bet 1). The RP is then 1,000–CE, which is the extra money you need to be compensated for the risk you are taking in 1) or the price you are willing to pay to eliminate the risk in 1) in exchange for the certainty in 2). Are you risk averse? Do you believe the saying that a bird in the hand is worth two in the bush?

The RP is therefore the amount of money that a rational person would demand for taking on a risk asset, or the amount of money that a rational person is willing to pay to eliminate risk. Thus,

$$RP = \mathbb{E}[W] - CE. \tag{4}$$

According to the expected utility theory, the source of the excess return of a risk asset, such as stock, is the RP. The RP is the reason people pay insurance premiums to insurance companies to remove risk, and it is also reflected by the interest rate of a loan. Figure 2 shows the relationship among expected wealth $\mathbb{E}[W]$, expected utility $\mathbb{E}[U(W)]$, CE, and RP. See "Take a Bet: Source of the Excess Return" for a gambling example for the risk-averse behavior and the RP.

Asset correlations and portfolio optimization

For a security/asset i, the net return $R_{i,t}$ at time t is

$$R_{i,t} = \frac{p_{i,t} - p_{i,t-1} + d_{i,t}}{p_{i,t-1}}, \tag{5}$$

where $p_{i,t}$ is the price at time t and $d_{i,t}$ is the dividend during the period $t-1$ to t. [The logarithm of the total return $\log(1+R_{i,t})$ is often used because of asymmetry of the net return. It is easy to show that the log return and the net return are essentially the same when the net return is small. The log return is more commonly used in empirical research.]

Assume that $R_{i,t}$ is stationary and its return is represented by a random variable R_i. The expected return is $\mathbb{E}[R_i] = \mathbb{E}[R_{i,t}]$. If the utility function can be approximated by quadratic form [15], [16], the expected utility maximization becomes a mean-variance investment criterion: maximize the expected return for a given variance, or minimize the variance for a given expected return. The risk of an asset is represented by the variance $\sigma_i^2 = \mathrm{Var}(R_i)$ or the square root of variance σ_i, also called *volatility* in finance.

Using SP concepts, we can rephrase mean-variance criterion: because people are risk averse, they like the signal (expected return) but not the noise (risk). If we have a pool of securities, we can take advantage of the correlations among individual securities and improve the signal-to-noise ratio (SNR) by investing in a portfolio (i.e., a basket of n securities). The portfolio return is $R_p = \sum_{i=1}^{n} x_{ip} R_i$, where x_{ip} is the weight of the ith security, $\sum_i x_{ip} = 1$. The portfolio expected return is

$$\mathbb{E}[R_p] = \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i],$$

and the portfolio variance is

$$\sigma_{R_p}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ip} x_{jp} \sigma_{ij} = \boldsymbol{x}_p^T \boldsymbol{\Sigma} \boldsymbol{x}_p,$$

where $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$ is the covariance of R_i and R_j and Σ is the covariance matrix. We can rewrite

$$\sigma_{R_p}^2 = \sum_{i=1}^{n} x_{ip} \left(\sum_{j=1}^{n} x_{jp} \sigma_{ij} \right) = \sum_{i=1}^{n} x_{ip} \mathrm{Cov}(R_i, R_p).$$

We see that the contribution of security i to the risk or variance of the return on portfolio p is $x_{ip}\mathrm{Cov}(R_i, R_p)$, i.e., the risk of security i in portfolio p or the weighted average of covariances. From the SP perspective, this risk is a projection of R_i on R_p. We can therefore formulate a portfolio optimization problem (see "Mean-Variance Portfolio") to find the best portfolio (weights) [17].

To find the optimal mean-variance portfolios (MVPs), we use the Lagrangian expression:

$$J = \sigma_{R_p}^2 + 2\lambda_e \left(r_e - \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i] \right) + 2\phi_e \left(1 - \sum_i x_{ip} \right), \tag{6}$$

where $2\lambda_e$ and $2\phi_e$ are the Lagrange multipliers. We then take derivative and set it to zero:

$$\frac{\partial \sigma_{R_e}^2}{\partial x_{ie}} - 2\lambda_e \mathbb{E}[R_i] - 2\phi_e = 0, \forall i = 1, \ldots, n,$$

Mean-Variance Portfolio

Mean-variance portfolios (MVPs) are the solution of the following mean-variance optimization (convex quadratic programming) criterion:

$$x_{ip}^* = \arg\min_{x_{ip}} \{J(x_{ip}) = \sigma_{R_p}^2 = \boldsymbol{x}_p^T \boldsymbol{\Sigma} \boldsymbol{x}_p\},$$
$$\text{s.t.} \quad \mathbb{E}[R_p] = \sum_{i=1}^{n} x_{ip}\mathbb{E}[R_i] = x_p^T \mathbb{E}[R] = r_e,$$

where $\sum_i x_{ip} = 1$. Here, r_e is a given target level of expected return. Such portfolios are called *mean-variance efficient portfolios*. All efficient portfolios with different r_e's constitute an efficient frontier.

where the subscript e represents the optimal (efficient) portfolio. Therefore,

$$\sum_{j=1}^{n} x_{je}\sigma_{ij} - \lambda_e \mathbb{E}[R_i] - \phi_e = 0, \forall i = 1 \cdots n. \tag{7}$$

In matrix form, $\boldsymbol{\Sigma x}_e - \lambda_e \mathbb{E}[\boldsymbol{R}] - \phi_e = 0$. Along with $\boldsymbol{x}_e^T \mathbb{E}[\boldsymbol{R}] = r_e$, and $\mathbf{1}^T \boldsymbol{x}_e = 1$, we have

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbb{E}[\boldsymbol{R}] & \mathbf{1} \\ \mathbb{E}[\boldsymbol{R}^T] & 0 & 0 \\ \mathbf{1}^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_e \\ -\lambda_e \\ -\phi_e \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ r_e \\ 1 \end{bmatrix}.$$

The optimal MVP can then be obtained by solving the linear equation.

Capital asset pricing model

Assume that market participants, or at least some of them, are rational in the sense of mean-variance optimal (by maximizing the expected utility). Then all securities in the market should be priced such that the market is in equilibrium. Note that we are indifferent to holding different assets given risk-compensated premia (in "Take a Bet: Source of the Excess Return," it means we are indifferent to taking the cash or the bet). Therefore, the market portfolio (e.g., the basket of all stocks in the U.S. stock market) must be an efficient portfolio e.

Note that according to the envelope theorem [22], the Lagrange multiplier $2\lambda_e$ in (6) is the rate of change (the slope) of the optimal portfolio variance $\sigma_{R_e}^2$ as a function of desired expected portfolio return $\mathbb{E}[R_e]$; i.e.,

$$2\lambda_e = \frac{d\sigma_{R_e}^2}{d\mathbb{E}[R_e]},$$

i.e., the efficient frontier is parabolic, as shown in Figure 3. Figure 3 also shows that the efficient frontier is the boundary of all attainable portfolios and securities (within the efficient frontier). The portfolio B is the minimum-variance portfolio.

For an efficient portfolio A, the slope of the tangent line in Figure 3 is

$$S_e = \frac{d\mathbb{E}[R_e]}{d\sigma_{R_e}} = \frac{\lambda_e}{\sigma_{R_e}}.$$

We know that an efficient portfolio satisfies (7), implying that

$$\sum_{j=1}^{n} x_{je}\sigma_{kj} - \lambda_e \mathbb{E}[R_k] = \sum_{j=1}^{n} x_{je}\sigma_{ij} - \lambda_e \mathbb{E}[R_i].$$

Multiplying both sides by x_{ke} and then summing over k, we get

$$\sigma_{R_e}^2 - \lambda_e r_e = \sum_{j=1}^{n} x_{je}\sigma_{ij} - \lambda_e \mathbb{E}[R_i].$$

Note that the expected return of the efficient portfolio is $\mathbb{E}[R_e] = r_e$ by definition. Rearranging, we get

$$\mathbb{E}[R_i] - \mathbb{E}[R_e] = \frac{1}{\lambda_e}\left(\sum_{j=1}^{n} x_{je}\sigma_{ij} - \sigma_{R_e}^2\right).$$

Therefore,

$$\begin{aligned} \mathbb{E}[R_i] &= (\mathbb{E}[R_e] - S_e\sigma_{R_e}) + \frac{S_e}{\sigma_{R_e}}\mathrm{Cov}(R_i, R_e) \\ &= r_{0e} + \frac{\mathbb{E}[R_e] - r_{0e}}{\sigma_{R_e}}\mathrm{Cov}(R_i, R_e) \\ &= r_{0e} + \beta_{0e}(\mathbb{E}[R_e] - r_{0e}), \end{aligned} \tag{8}$$

where

$$\beta_{0e} = \frac{\mathrm{Cov}(R_i, R_e)}{\sigma_{R_e}^2}, \tag{9}$$

and the quantity

$$r_{0e} \triangleq \mathbb{E}[R_e] - S_e\sigma_{R_e} \tag{10}$$

is the intercept with zero variance and represents a risk-free asset. The relationship (8) must hold in the market equilibrium between the return of a single security and the efficient portfolio. Therefore, there is no expected return reward for the part of security risk that is uncorrelated (zero covariance) with the efficient frontier.

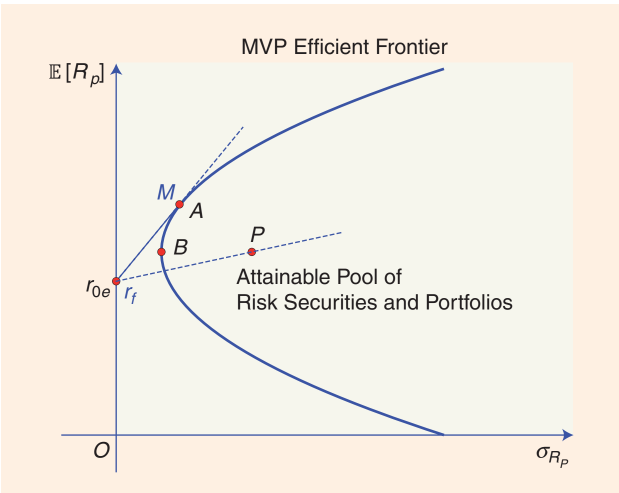

FIGURE 3. The efficient frontier of MVP.

CAPM

The price of a security i (or a portfolio) in a free market can be determined by the CAPM, also called the *Sharpe–Lintner model*, as follows:

$$\mathbb{E}[R_i] = r_f + \beta_{iM}(\mathbb{E}[R_M] - r_f), \forall i,$$

where

$$\beta_{iM} = \frac{\mathrm{Cov}(R_i, R_M)}{\sigma^2(R_M)}.$$

Time-series regressions are used in practice to compute β. The following time-series regression is perhaps the most famous linear regression in the world:

$$R_{i,t} = r_{f,t} + \beta_{iM}(R_{M,t} - r_{f,t}) + \varepsilon_{i,t}.$$

Insights

- The CAPM is a market equilibrium that must hold.
- There is no expected return reward for idiosyncratic risk, the part of security risk that is uncorrelated with the market portfolio M.

The CAPM is the foundation of asset pricing. It is commonly used in corporate finance, such as merger-and-acquisition practice to evaluate a company and the cost of capital.

Terms Related to the CAPM and Asset Pricing

- *Risk asset*: assets that have uncertain returns, e.g., stocks, bonds, real estate, gold, crude oil. Their prices fluctuate over time. Almost all assets are in this category.
- *Risk-free asset*: assets that have fixed (certain) returns. Treasuries (especially T-bills) are considered to be risk free because they are backed by the U.S. government. The return on risk-free assets is very close to the current interest rate.
- *Excess return*: $\mathbb{E}[R_i] - r_f$ is the expected return in excess to the risk-free rate to reward the security risk, i.e., an RP.
- *Market RP*: $\mathbb{E}[R_M] - r_f$ is the excess expected return to reward market risk (approximately 7% annually in the U.S. market).
- *Systematic risk*: market risk that cannot be diversified and is rewarded by the expected return. Represented by β.
- *Idiosyncratic risk*: firm-specific risk that can be diversified, thus not rewarded by the expected return. Represented by $\varepsilon_{i,t}$. Note $\mathbb{E}[\varepsilon_{i,t}] = 0$. It looks like the noise term in SP, but it is not noise in economics.

Evaluating an Active Portfolio Manager—Is There Alpha?

The time-series CAPM regression for an investment portfolio is

$$R_{p,t} - r_{f,t} = \alpha + \beta_{pM}(R_{M,t} - r_{f,t}) + \varepsilon_{p,t},$$

where α is the intercept of the regression. It should be statistically insignificant according to the CAPM, i.e., $\alpha = 0$.

The skill of an investment manager should not be evaluated by the expected return or the excess return over the market portfolio but rather by α.

If the market risk-free rate is r_f, a market participant will hold a combination of the risk-free asset r_f and a risk portfolio or security within or on the efficient frontier. Apparently, the best risk portfolio to hold is on point A because the tangent line provides the best portfolios in terms of the mean-variance criterion. Indeed, the tangent line provides the best SNR from the SP perspective. [The Sharpe ratio of a portfolio p can be considered the SNR defined in finance: $(\mathbb{E}[R_p] - r_f)/(\sigma_{R_p})$.]

The efficient portfolio A is called the *tangency portfolio*. In market equilibrium, the market portfolio M (i.e., the portfolio consisting of all risk securities) must be the tangency portfolio A. A market index, such as the S&P 500 index, is often used as a proxy of the market portfolio M. All securities must have prices that satisfy (8). Such a price relationship (or correlations) among all risk securities in the market is described in the capital asset-pricing model (CAPM) [18]–[21] (see "CAPM" for details).

As we show, the CAPM is consistent with the expected utility theory and portfolio optimization. According to the CAPM, the stock return is an RP and is only related to its correlation with the market portfolio (i.e., the market risk a stock carries). The higher the β, the larger is the expected stock return. The diversifiable (idiosyncratic) risk a stock carries (i.e., the company-specific risk) does not have RP rewards.

Operationally, a higher expected return can always be achieved by leveraging the risk (i.e., by borrowing money to buy market portfolio). In Figure 3, this means moving along the tangency line to the right, thus generating higher risk and higher return with a fixed Sharpe ratio (or SNR).

The CAPM provides a way to evaluate an investment portfolio and attribute its performance (see "Evaluating an Active Portfolio Manager—Is There Alpha?"). "An Example CAPM Regression" shows a typical empirical regression study.

Fama–French three-factor model and multifactor asset pricing

It is difficult to have a significant alpha, i.e., to beat the market. People have long strived to find excess returns over the market premium. Many of the identified excess returns turned out to be statistically chance results, though a few proved to be persistent. The most prominent model identifying excess returns beyond the CAPM is the Fama–French (FF) three-factor

An Example CAPM Regression

Table S1 shows time-series regressions of excess stock portfolio returns (in percent), $R_{pt} - r_{f,t}$, on the excess stock-market return, $R_{M,t} - r_{f,t}$, from July 1963 to December 1991 (342 months) for 25 portfolios using Center for Research in Security Prices monthly return data of all U.S. market stocks. The 25 portfolios are constructed by dividing firms using market caps (size) and book-to-market (B/M) ratios into quintile buckets. The empirical results are extracted from [23]. The t-values and $s(\varepsilon)$ are in parentheses. The residual standard error $s(\varepsilon)$ is the root-mean-square error of the regression. As can be seen, only a few α coefficients are statistically significant, i.e., $|t(\alpha)| > 2$. All of the β coefficients are statistically significant.

Table S1. The time-series regressions from July 1963 through December 1991.

	B/M Ratio				
Size	**Low**	**2**	**3**	**4**	**High**
	α and t-value $t(\alpha)$				
Small	–0.22	0.15	0.30	0.42	0.54
	(–0.90)	(0.73)	(1.54)	(2.19)	(2.53)
2	–0.18	0.17	0.36	0.39	0.53
	(–1.00)	(1.05)	(2.35)	(2.79)	(3.01)
3	–0.16	0.15	0.23	0.39	0.50
	(–1.12)	(1.25)	(1.82)	(3.20)	(3.19)
4	–0.05	–0.14	0.12	0.35	0.57
	(–0.50)	(–1.50)	(1.20)	(2.91)	(3.71)
Big	–0.04	–0.07	–0.07	0.20	0.21
	(–0.49)	(–0.95)	(–0.70)	(1.89)	(1.41)

Table S1. The time-series regressions from July 1963 through December 1991.

	B/M Ratio				
Size	**Low**	**2**	**3**	**4**	**High**
	β and t-value $t(\beta)$				
Small	1.40	1.26	1.11	1.06	1.08
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
2	1.42	1.15	1.12	1.02	1.13
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
3	1.36	1.15	1.04	0.96	1.08
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
4	1.24	1.14	1.03	0.95	1.10
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
Big	1.03	0.99	0.89	0.84	0.89
	(26.33)	(28.12)	(27.01)	(25.03)	(23.01)
	Adjusted R^2 and Residual Standard Error $s(\varepsilon)$				
Small	0.67	0.70	0.68	0.65	0.61
	(–1.46)	(3.76)	(3.55)	(3.56)	(3.92)
2	0.79	0.79	0.76	0.76	0.71
	(3.34)	(2.96)	(2.85)	(2.59)	(3.25)
3	0.84	0.84	0.80	0.79	0.74
	(2.65)	(2.28)	(2.33)	(3.26)	(2.90)
4	0.89	0.90	0.87	0.80	0.76
	(2.01)	(1.73)	(1.84)	(2.21)	(2.83)
Big	0.89	0.91	0.54	0.79	0.69
	(1.66)	(1.35)	(1.73)	(1.95)	(2.69)

FF Three-Factor Model and Cross-Sectional Returns

The FF three-factor model is an empirical asset-pricing model featuring two cross-sectional return factors, the small-minus-big (SMB) factor and the high-minus-low (HML) factor:

$$\mathbb{E}[R - r_f] = \beta_M \mathbb{E}[R_M - r_f] + \beta_s \mathbb{E}[R_{\text{SMB}}] + \beta_h \mathbb{E}[R_{\text{HML}}].$$

- The SMB factor is a firm size factor, represented by the portfolio return of SMB (a portfolio of long small-market-cap stocks and short big-cap stocks. Firm size is measured by firm market capitalization (cap). Small-size firms have excess returns in addition to the market RP.
- The HML is a market-value factor, calculated by the portfolio return of HML (a portfolio of long high-value stocks and short low-value stocks). Firm value is measured by the B/M ratio or the price-to-earnings ratio. High B/M ratios represent value stocks and low ratios growth stocks. Value stocks have excess returns in addition to the market RP.

Insight

SMB (or small-market cap) and HML (or value) portfolios have significant nonzero alpha with respect to the CAPM, i.e., excess returns in addition to the market RP. Note that these two excess returns are predicted by the cross-sectional factors determined by firm accounting features. Cross-sectional means across various stocks/portfolios (sections).

MMV Portfolios and Intertemporal CAPM

Other than expected returns, do people have other tastes? That is, they may like aspects of a firm other than the expected return. For example, a potential explanation for the value factor is that many people like to hold growth stocks (e.g., unicorn firms) for the glory and therefore lower their return expectation.

With multiple tastes/preferences, the MMV portfolio optimization problem can be formulated as follows:

$$x_{ie} = \arg\min_{x_{ip}} \{J(x_{ip}) = \sigma^2_{R_p} = \boldsymbol{x}_p^T \boldsymbol{\Sigma} \boldsymbol{x}_p\},$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_{ip} b_{is} = b_{es}, s = 1, \ldots, S, (i.e., \boldsymbol{x}_p^T \boldsymbol{b}_p = \boldsymbol{b}_e),$$

$$\mathbb{E}[R_p] = \sum_{i=1}^{n} x_{ip} \mathbb{E}[R_i] = \boldsymbol{x}_p^T \mathbb{E}[R] = r_e,$$

$$\sum_i x_{ip} = 1,$$

where r_e is the desired level of expected return. Note that s represents the tastes other than expected returns and b_{es} is the desired level of the sth taste.

The market equilibrium for the MMV criterion is a multifactor intertemporal CAPM [28]:

$$\mathbb{E}[R_i] - r_f = \beta_{ie}(\mathbb{E}[R_M] - r_f) + \sum_{s=1}^{S} \beta_{is}(E[R_s] - r_f), \forall i.$$

That is, there are multiple MMV-efficient market portfolios (factors).

Minimum-Variance Portfolio and Capon Beamformer

Consider the problem of finding the minimum-variance portfolio, i.e., point B in Figure 3, formulated as

$$R_i(t) = a_i \mathbb{E}[R_i] + \varepsilon_i(t), i = 1, \ldots, n.$$

Find weights $\boldsymbol{x}_{mv}$ such that

$$\boldsymbol{x}_{mv} = \arg\min_{\boldsymbol{x}_p} \boldsymbol{x}_p^T \boldsymbol{\Sigma}_R \boldsymbol{x}_p,$$

subject to the weight normalization:

$$\boldsymbol{x}_p^T \boldsymbol{a} = 1.$$

In array SP, this is exactly the minimum-variance distortionless response beamformer, also known as the *Capon beamformer*, to recover the signal sources $E[R_i]$ with minimum noise variance [29]:

$$\boldsymbol{x}_{mv} = \frac{\boldsymbol{\Sigma}_R \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{\Sigma}_R^{-1} \boldsymbol{a}}.$$

Remarks

1) Many portfolio construction/analysis problems have similar forms to array SP problems.
2) The covariance matrix $\boldsymbol{\Sigma}_R$ can be a very large matrix in portfolio analysis because n is large. It is often difficult to estimate such a large matrix reliably. In addition, because of high correlations among assets, this matrix often has a very high conditional number (close to singular), and therefore it is difficult to compute its inverse.

model [24], [23] (see "FF Three-Factor Model and Cross-Sectional Returns"), which suggests that the size and firm value factors can generate excess returns.

Based on the FF three-factor model, prior empirical studies suggest that additional factors persistently produce excess returns. The most notable include the momentum factor, as in the four-factor model by Carhart [25], and the two additional factors, profitability and investment patterns, that Fama and French [26] added in their more recent five-factor model. Theoretically, multiple-factor models can be explained by multifactor minimum-variance (MMV) portfolios [27]. See "MMV Portfolios and Intertemporal CAPM." New research evaluating investment alpha needs to incorporate all known factors in the present multiple-factor models. Historical empirical analysis has shown that, when the four-factor model is considered, no skilled or informed mutual fund portfolio managers can generate persistent investment alpha [25].

The CAPM-related study is a mature area with established statistical methodology. Current studies in this area are mainly empirical research identifying new cross-sectional factors, e.g., [26]. Beta evaluation in traditional study is based on a windowed regression with typically three- to five-year windows, assuming stationarity within the window. The opportunity for SP might lie in the time-varying nonstationary models. The difficulty of any time-varying model is the evaluation criteria given that there are no ground-truth beta parameters.

Portfolio analysis and array SP

Consider the return series of an asset as a noisy signal. A portfolio can be considered a way to use diversity to improve SNR. This is quite similar to many SP formulations. See "Minimum-Variance Portfolio and Capon Beamformer" for the relationships between them in array SP.

Even if we cannot generate investment alpha in portfolio construction, there are other aspects of a portfolio with which we are concerned. For example, different people often have different preferences for systematic market risks represented by various risk factors. These systematic market risks cannot be eliminated by diversification, but investors can choose/analyze the amount of portfolio exposure to those market risks

Example SP Research on Portfolio Optimization and Related Risk Modeling

In [30], a subspace formulation of MVP optimization with risk-factor constraints and related toy examples is given. For example, a market-neutral portfolio requires $\boldsymbol{F}^T \boldsymbol{x}_p = 0$, where $\boldsymbol{F}$ is the loading weights of a single market factor, usually estimated by a time-series regression of all security return time series with a market index, and $\boldsymbol{x}_p$ represents the desired portfolio weights. Torun et al.'s article in *IEEE Signal Processing Magazine* [31] shows an eigenfiltering method to estimate the covariance matrix with a large number of securities. When the security pool is large, the raw covariance matrix is almost singular due to high correlations among securities. Note that the inverse of the covariance matrix, i.e., the precision matrix, is often necessary in MVP optimization. Therefore, a noise regularization term is added on the eigenfiltered covariance matrix to improve the robustness of covariance matrix estimation and its conditional number. The same authors [32] further use an autoregressive model to improve the stability and computational efficiency of estimation of the empirical covariance matrix of highly correlated securities. We also direct readers to an example in [33] that uses Bayesian and regularization methods for the estimation of the unknown observation covariance matrix and the related MVP optimization algorithm. Also, the authors in [34] propose to use smooth and monotone regularization to tackle the high correlation problem in covariance matrix estimation. The work in [35] summarizes the relative performance of different estimation strategies for minimum-variance portfolio optimization problems using the inversion of the estimated covariance matrix or the direct estimation of the precision matrix.

Remarks

We caution SP researchers that the risk and optimal portfolio obtained from the covariance matrix are only from a period of historical data, assuming that there is a stationary joint return distribution among the assets. It is, at most, a rough approximation of the ever-changing financial market. The out-of-sample backtesting is necessary and helpful, but it is by no means comprehensive and bullet proof in terms of representing the future distributions. Therefore, the economic reasoning and understanding of risk factors are always necessary in both practice and economics theoretical development.

(by changing the weight of the risk-free asset). The idiosyncratic risk represented by the regression residual (or noise) is the risk not rewarded (priced) by the market and therefore is the part of the risk that investors need to try to remove by diversification. A portfolio analysis and risk-factor attribution problem usually has the following formulation.

For a set of n portfolios, the time-series model of the portfolio excess return (with risk-free rate subtracted) is

$$R_i(t) = \sum_{s=1}^{K} \beta_{si} R_s(t) + \varepsilon_i(t), i = 1, \ldots, n,$$

where $R_s, s = 1, \ldots, K$ are risk factors. Note that risk factors are also portfolios.

The portfolio analysis or construction problems can be 1) finding common risk portfolios R_s given R_i; 2) given risk factors, which are often represented by different market or section indices, or cross-sectional portfolios, such as those in the FF three- or five-factor models, finding the risk loadings (betas), which reflect the risk exposures of a portfolio; or 3) covariance estimation, portfolio optimization or analysis, often subject to various constraints, such as long/short, borrowing, liquidity, and transaction cost. This is an area to which the SP research can directly contribute due to the similarity in problem formulation in statistical and array SP.

Specifically, practical portfolio construction and optimization usually has following steps.

- Identify portfolio constraints. First, identify the desired expected return or the tolerable risk represented by variance. Note that we can either minimize the variance given the expected return or maximize the expected return given the variance. Second, identify the investable security pool along with investment constraints, including short sell constraints, liquidity constraints, transaction costs, and risk-factor exposure constraints. In practice, these constraints are often determined by the requirements from the investment fund stakeholders, trading systems, and regulatory bodies.
- Identify historical data or proxy data that can be used to, first, estimate covariance and expected returns of a security pool and, second, conduct backtesting.
- Find a robust algorithm to estimate the covariance matrix (or precision matrix) and expected returns.
- Formulate a constrained optimization problem, and find a robust solution.
- Conduct backtesting, and evaluate the portfolio performance against benchmark models, usually known factor models.

In "Example SP Research on Portfolio Optimization and Related Risk Modeling," we show SP research in portfolio optimization. The nature of an economic system is different from that of an SP system. So far, all economic models and theories we introduced are hypotheses. In SP, system and signal models are often rooted in physical laws known to be sufficiently accurate descriptions of system mechanisms. SP researchers tend to be quite confident with their models and believe that the same physical law applies at all times (yesterday, today, and tomorrow).

Therefore, even when the mathematical representation in portfolio analysis, and in economic and business studies in general, is similar to that in SP [36], the treatment can be quite different. For example, in financial empirical studies, when certain risk factors are identified mathematically using the data at hand, it is actually not clear whether the identified risk factors are indeed true risk factors or a result of statistical chance. Two commonly used methods can verify this: 1) conduct statistical significance analysis with out-of-sample data and 2) have reasonable economic explanations for the identified factor. Even then, the results are still confirmatory rather than conclusive because there are no certain physical laws in the market activities, it is unknown whether the same model will hold over time, and controlled experiments cannot be conducted to test the model. Because researchers always operate on hypotheses in economic and business studies, the burden of proof is on their side, especially when results contradict with conventional wisdom or well-recognized theories, such as the EMH.

Efficient market theory and behavioral economics

An efficient capital market is a market that is efficient in processing information: asset prices fully reflect available information. An efficient market does not necessarily mean that the stock price follows the random-walk model or any existing model.

EMH

Because of the similarity to many noisy signals in SP applications, the stock return series is fascinating to many SP researchers. It is tempting to look into the historical stock charts trying to identify money-making patterns and hope that an autoregressive moving average (ARMA) model can fit the chart. It is often all too easy for SP researchers to think that they may be able to predict stock price using sophisticated SP models to make free money.

However, as the expected utility theory suggests, the source of excess return is the risk that investors carry. In contrast with a physical system, two parties are involved in any financial transaction. In a free market, the two parties agree on a (fair) price for the transaction. In market equilibrium, the price should reflect all available information in the market, such that the transaction is merely an RP changing hands. For example, if both seller and buyer predict that the price of an asset will go up tomorrow, the price should go up today. A rational seller will not sell if the price is lower than what he or she predicts. The efficiency of information processing in the market is indeed a natural consequence of competition in a free market.

In the 1960s, Eugene Fama, a 2013 Nobel Laureate in Economics, was one of the first using computers to study stock prices. Based on his empirical study, he proposed that security prices at any time fully reflect all available information [4]. Such a market is called an *efficient market*. See "Efficient Capital Markets—Different Forms."

Efficient Capital Markets—Different Forms

An efficient capital market is a market that is efficient in processing information and asset prices fully reflect available information. Markets have different forms of efficiency.

- Weak form: current prices fully reflect all information in past prices. Technical analysis using past price patterns will not produce profits.
- Semistrong form: current prices fully reflect past prices and all publicly available information. Fundamental analysis (e.g., studying financial statements) will not produce profits.
- Strong form: current prices fully reflect all information, public and private. Insider trading will not produce profits.

Evidence

1) For the weak form, the random-walk stock price model has been empirically validated for short-term stock returns in many tests, i.e., the short-term stock return series is not statistically different from white noise and therefore is unpredictable.
2) For the semistrong form, a large number of event studies have compared the stock returns before and after corporate news events, such as stock split and earning announcements, showing that the information is incorporated into the stock price almost immediately.
3) For the strong form, insider trading makes profit, indicating that the strong-form market efficiency does not hold. However, mutual fund managers who have more private information than the general public do not outperform on a consistent basis.

Holding the utility function constant, we can argue that the efficient market means that the stock price tomorrow is unpredictable, as all available information has been incorporated into today's price. The market only moves by new information, not by known information. See "Fundamental Analysis, Technical Analysis, and Quantitative Analysis" for different types of stock analysis.

Joint hypothesis testing and its implications in SP

To test the market efficiency, we need to have a market equilibrium model, i.e., an asset-pricing model, such as the CAPM, the FF three-factor model, or the random-walk model. We need to test whether the properties of expected returns implied by the market equilibrium asset-pricing model are observed in actual returns. The main difficulty in testing market efficiency indeed lies in the joint hypothesis test: when the test fails—i.e., we find credible anomaly in price behavior—we do not know whether the market is inefficient, a conclusion people often jump to, or the asset-pricing model used is wrong. Note that while the random-walk model is a simple (hypothetical) form for an efficient market (an idea popularized by a bestselling book [37]), an efficient market does not necessarily mean that the stock price follows the random-walk model or any existing model, as many people have assumed.

While all empirical scientific research is more or less subject to the joint hypothesis test, in SP, researchers do not pay much attention to 1) statistical significance and confidence of model estimation, 2) error distribution of the estimation

Fundamental Analysis, Technical Analysis, and Quantitative Analysis

Fundamental analysts analyze financial statements, management and competitive advantages, competitors, and markets to find the intrinsic value of a firm. Fundamental analysis is the mainstream methodology for financial professionals.

- Pick a company with mispriced intrinsic/correct value
- Critique: subjective guesswork is required for future growth and risks.

Technical analysts (chartists) analyze price and volume patterns/charts to predict price movement and direction. Technical analysis is not mainstream among financial professionals.

- Buy low, sell high by analyzing price-chart patterns, e.g., head-and-shoulder or double top/bottom reversal patterns.
- Technical indicators: moving average, relative strength index, moving average convergence divergence, etc.
- Critique: the pattern definition is subjective and without rigorous statistical tests. Different chartists have different opinions.

Quantitative analysts (quants) analyze financial markets using complex mathematical and statistical modeling, measurement, and research of historical data.

- All quantitative risk modeling and algorithm trading, and increasingly more hedge funds, are based on quantitative analysis.
- Critique: many quantitative analyses lack economic justifications. History may not represent the future. The validity of the analysis is subject to various model and data mining risks.

beyond the Cramér–Rao bound, and 3) statistical tests against alternative models, because the confidence in the model used is established from the understanding of physical laws rather than purely from data. For example, to track a flying missile, we have basic understanding of its motion model based on Newton's laws and sources of noise. However, we do not have such understanding of the market price movement other than common sense. The statistical test is often the only quantitative tool on which we can rely.

Thus, for stock price prediction, researchers need to ask themselves, Are we confident that we have done enough tests to bet the money in the market? Are we confident that our price model beats the mainstream asset-pricing models and the collective information processing (crowdsourcing) of all other market participants? Note that market equilibrium is reached by competing arbitrage-seeking efforts in the market. Apparently, there is a large burden of proof against the EMH.

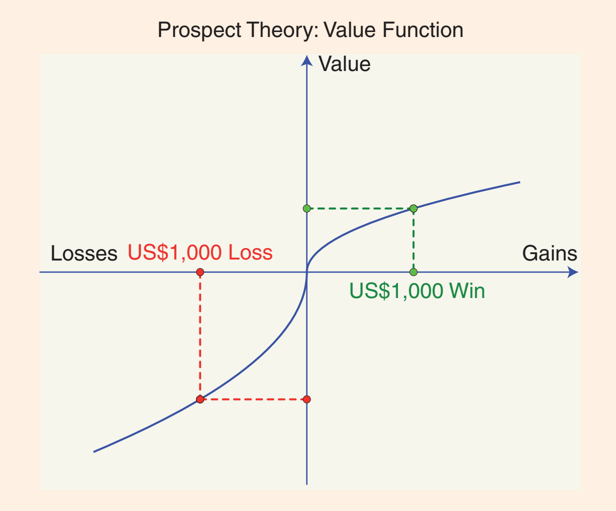

FIGURE 4. The prospect theory assumes that people make choices according to the value function. The amount of pain felt from losing US$1,000 is about double the amount of joy felt from winning the same amount.

Prospect theory and behavioral economics

Both the expected utility theory and the EMH assume that people, or at least some people, act rationally and optimally to exploit risk-free profit in the market to maximize the expected utility. Therefore, the only source of excess return is risk. The expected utility theory and the EMH were dominant before the 1980s in academia.

However, people frequently argue that it is often the irrational or emotional decisions that drive the market, generating bubbles and recessions. Such psychological overreaction or underreaction to information moves the market, generating mispricing and predictable market movement patterns, and thus possible risk-free profitable opportunities. This line of thinking has led to the development of behavioral economics, embraced by many people eager to justify price anomalies and patterns, both academics and (more) practitioners, as it provides a formidable (desirable) alternative to the EMH.

In 1979, the psychologists Daniel Kahneman and Amos Tversky developed prospect theory [7], which has become a foundational theory for behavioral economics. Daniel Kahneman won the 2002 Nobel Memorial Prize in Economic Sciences for his foundational work in behavioral economics (Amos Tversky died in 1996).

In contrast with the expected utility theory, prospect theory states that people make choices based on the value function illustrated in Figure 4. The value function states that people make financial decisions not from the expected utility of the absolute total wealth but from the psychological and emotional perceptions of joy and pain depending on the relative gain or loss. The function is asymmetric in that the amount of pain from losing US$1,000 is about double the amount of joy of winning the same amount. The gain part is concave, while the loss part is convex. A loss-aversion gambling example is shown in "Take a Bet Again—Intolerance of Loss?"

Take a Bet Again—Intolerance of Loss?

Assume you got a traffic ticket and have the following two choices:

1) the bet: flip a coin (equal probability on both sides); heads, you pay a US$2,000 fine, tails, you pay no fine
2) cash: you pay a US$1,000 fine in cash.

Would you take the bet or pay the US$1,000 cash fine? Are you loss averse and risk seeking?

The implication of prospect theory to economics, decision theory, and business management is profound. However, it is also all too easy for researchers to attribute any stock return anomaly found against established asset-pricing models to behavioral factors. Note that for a purely behaviorally caused market anomaly to persist over time, one must assume that even smart investors as a group do not learn over time (or generations) from their psychological mistakes.

It is important to note that most mathematical models in economics are based on the expected utility theory and EMH. For example, the Black–Scholes option pricing model [38] assumes an efficient market, in that stock price follows a random walk and there are no riskless arbitrage opportunities. (Both Robert C. Merton and Myron S. Scholes won the 1997 Nobel Prize in Economics for their contributions to option pricing.) Behavioral factors in the market are still difficult to quantify and have not led to fully developed asset-pricing models that can be tested and potentially rejected.

Long-term stock return predictability—efficient market or behavioral?

Robert Shiller, a prominent behavioral economist who won the 2013 Nobel Prize in Economics the same year as Fama, found that long-term stock return is predictable: a higher stock price (P) to dividend (D) ratio signals a lower expected return, and vice versa [8], [39]. The following time-series regression between the portfolio return and the portfolio firm dividend-to-price (D/P) ratio has significant nonzero coefficient b:

$$R_{t+1} = a + b(D/P)_t + \varepsilon_{t+1}.$$

The efficient market economist John H. Cochrane illustrated this well in an op-ed article in the *Wall Street Journal* at the end of 2008 [67], the height of the financial crisis, shown in Figure 5. In hindsight, the article turned out to be an accurate prediction of the stock market movement.

From the efficient market viewpoint [40], such predictable long-term stock returns are caused by time-varying risk premia. That is, the utility function is time varying. People are more risk averse in recessions because they do not have financial security and demand higher risk premia, i.e., lower stock prices and higher expected returns. By contrast, people are more tolerant of risks in economic booms, and therefore stock prices are high and expected returns (RP) are low. However, behavioral economists like Shiller believe that people overreact to economic situations, resulting in speculative bubbles in booms due to irrational exuberance [41] and recessions due to a lack of psychological confidence. Oftentimes, economists in both camps agree on the findings from data but dispute the cause: whether excess returns are indeed RP or psychological mispricing.

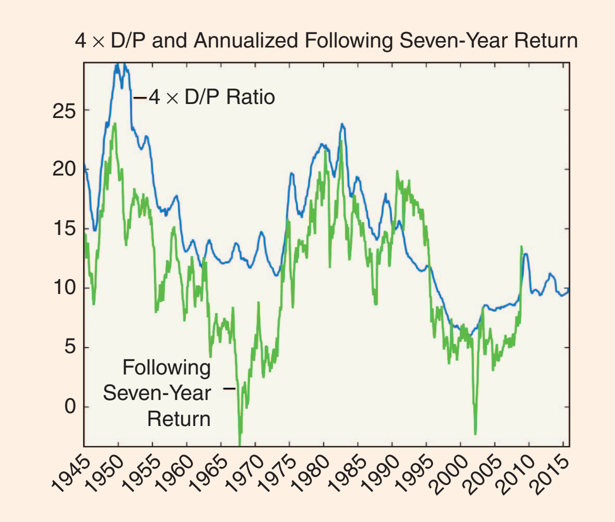

FIGURE 5. The D/P ratio can predict a future seven-year return. (Image courtesy of John H. Cochrane, a former president of the American Finance Association, updated from [40]. An earlier version of this figure also appeared in [67].)

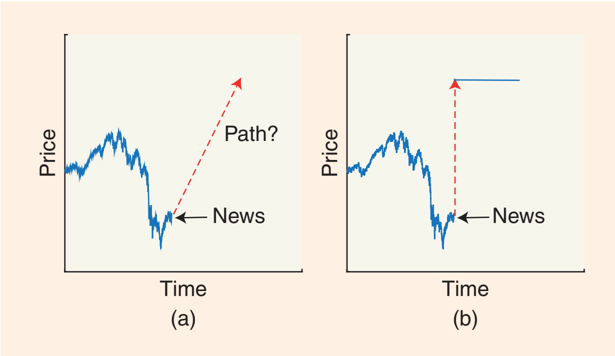

FIGURE 6. The information processing in the market: (a) it takes time for market participants to process complex information and for this information to reflect in the price, whereas in a mostly pure efficient market, information is processed immediately and fully reflected in the price (b). (This figure is modified from a figure in [42]. Image courtesy of John. H. Cochrane.)

Signal and information processing in efficient markets and traps in data

In the most purely efficient market, information is processed immediately and fully reflected in the price, as shown in Figure 6(b). That is, the market moves from one equilibrium to another equilibrium in no time. SP researchers are more familiar with the information-processing process as shown in Figure 6(a). They know that it often takes time (however short) for market participants to process complex information, whether it comes from human behaviors or trading/communication systems. Such transition processes are not well studied quantitatively in finance literature, posing opportunities for SP researchers. It is possible to use SP models and methodologies to find impulse response to certain events, exploring the microstructure of trading systems, especially in HFT systems.

Without ground-truth models and controlled experiments/studies in modeling economic systems, the validity of SP models relies on data as well as economic explanations. We note several traps in dealing with the model and data as follows.

- Traps in data. First, survivorship bias: only data from successful companies or funds are used in analysis. For example, stock data downloaded from popular websites like Yahoo Finance do not include historical delisted companies. Second, forward-looking bias: future (test) data are used to estimate the model or inform the model construction. Third, selection bias: data are selectively reported in a database, e.g., only mutual funds/hedge funds performing well report results.
- Solution: use unbiased data sets, e.g., Center for Research in Security Prices (CRSP) data set, a gold standard in academic research on U.S. market daily stock returns. See [43] for more details.

When evaluating and interpreting empirical results, researchers need to be careful to select meaningful criteria and statistical test methodology.

- Test against alternative models, e.g., can you beat the FF three-factor or the random-walk model?
- Minimum mean-squared errors are not enough. Note that asset-pricing model residuals are often considered idiosyncratic risks or the unknown risks that are not captured by known factors. The important thing may be whether a model properly attributes different types of risks rather than minimizes the residual. The explanatory power of a model should be given more consideration depending on the applications.
- Conduct extensive out-of-sample tests, and prove with statistical significance that your results are not due to chance.

Again, we caution that out-of-sample backtesting is helpful as much as out-of-sample testing in SP and machine-learning applications to check model fitness and robustness for known data over multiple time periods. However, multiple models may have similar fit into a set of data, and the past data may not perfectly represent the future in an open economic system. There is generally no gold criterion to know the ground-truth model without controlled experiments. Indeed, we do not even know whether there exists a ground-truth model. The joint hypothesis problem always exists in discovering a model or a theory. Therefore, economic justifications are always an integral part of economic and business studies.

Basic econometric models, time-series analysis, HFT, and SP

SP shares many similar terms and mathematics with econometrics, even though there are few interactions between the two communities. Understanding basic concepts in econometrics will help propel SP researchers in their work and allow them to appreciate application issues in economics and business.

ARMA

In time-series econometrics [14], [44], the definitions of models have similar forms to those in SP. An autoregressive (AR) process with the first-order, AR(1) process, is defined as

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t,$$

AR(p) processes are defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t,$$

and ARMA(p, q) processes are defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

where ε_t is white noise with zero mean and variance σ^2 and $\mathbb{E}[\varepsilon_t \varepsilon_\tau] = 0$ when $t \neq \tau$. Note that ε_t are indeed unpredictable innovations (shocks).

In econometrics, in place of the Z transform used in SP, a lag operator L is used to represent time shift. Thus, the ARMA(p, q) process can be represented by lag operator polynomials as

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t.$$

When observing an economic system, researchers do not have control of the innovation or the unexpected shock, ε_t, which can be various events, such as a sudden decrease of crude oil prices. Meanwhile, there may be other exogenous variables that are not uncorrelated white noise, such as the advertising investment of a company or interest rates set by central banks. Thus, to analyze an economic system, the ARMA model needs to be generalized to the ARMA processes with the exogenous variables (ARMAX) model. ARMAX(p, q, r) can be defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \sum_{i=1}^{r} \eta_i X_{t-i},$$

where X_t represents exogenous variables. A special case of the ARMAX model is AR distributed lag models (ARDL). An ARDL(p, r) model is defined as

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{r} \eta_i X_{t-i} + \varepsilon_t.$$

A vector AR (VAR) model can be used to examine the interaction of a set of n economic variables. VAR(p) is specified as

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{y}_{t-i} + \varepsilon_t,$$

where $\boldsymbol{y}_t$, $\boldsymbol{c}$, and ε_t are $n \times 1$ vectors and $\boldsymbol{\Phi}_i$ is an $n \times n$ matrix.

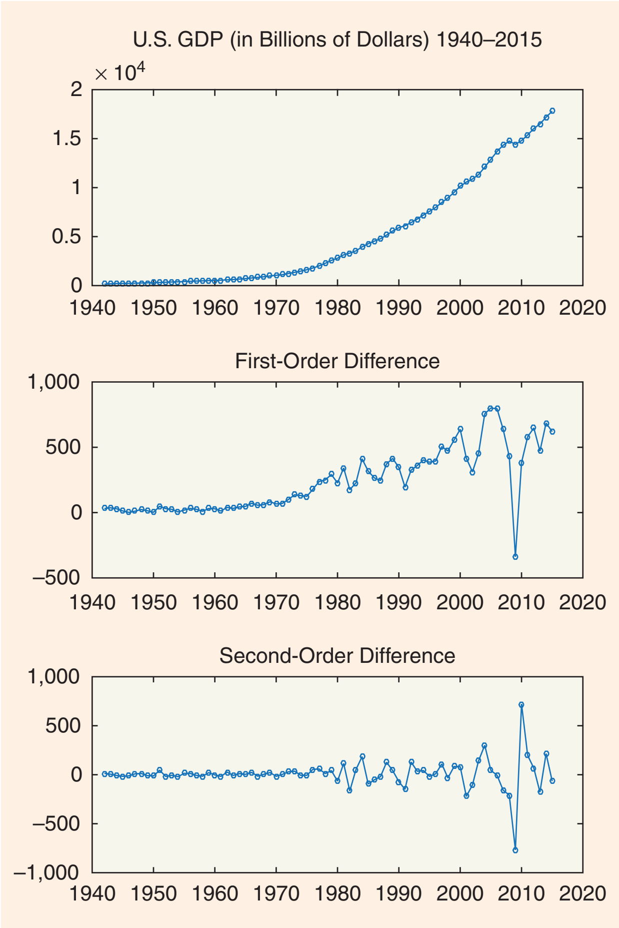

FIGURE 7. The U.S. GDP time series and its differences. This time series is an $I(2)$ process.

Nonstationary, integrated processes, and unit roots

The subtle differences in econometric time-series model definitions from SP reveal the different applications and assumptions in economics. A main difference is that in SP, the distinction between the system/model and the process/signal is clear because researchers can actively control or at least simulate the signal. In economics, the model is often a characteristic of the time series, which oftentimes can only be observed but not controlled or changed. In SP systems, there is no constant term c, as the output in a linear SP system is zero with zero input. In economic time series, the constant term c actually defines the stable status of the variable. For example, for AR(p) time series, we have

$$\mathbb{E}[Y_i] = \frac{c}{1 - \sum_{i=1}^{p} \phi_i}.$$

A more startling difference is that in SP, only stable processes and systems are investigated, i.e., the roots/poles must be within the unit circle. However, economic time series are often nonstable (the econometrics term is *nonstationary* or *evolving*). The nonstationarity (i.e., evolution) is often a desired property for economic variables. For example, a company hopes that its sales grow over time rather than die out over time, as does a country in terms of its gross domestic product (GDP). The analysis of nonstationary processes generates many research problems and tools.

A basic nonstationary evolving process is a random-walk process with drift:

$$Y_t = c + Y_{t-1} + \varepsilon_t.$$

Apparently, this process has one root on the unit circle, i.e., it has a unit root. Its first-order difference is stationary, and thus it is a process of integrated order 1, denoted by $I(1)$. An integrated process with order d, $I(d)$ process, is a nonstationary (unit-root) process whose dth-order difference is stationary. A unit-root process is also called an *evolving process*. See Figure 7 for the U.S. GDP time series and its differences.

The related AR integrated moving-average (ARIMA), ARIMA(p, d, q), model is defined as

$$\Delta^d Y_t = c + \sum_{i=1}^{p} \phi_i \Delta^d Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j},$$

where Δ^d is the dth difference operator.

The first analysis step for an economic time series is often to determine whether it has a unit root, i.e., to conduct the unit-root test; see "Unit-Root Test."

Unit-Root Test

AR(1) unit-root (random-walk) test:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2), \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0, \forall t \neq s.$$

The null hypothesis is $H_0 : \phi = 1$(unit root). There is no analytical closed-form expression for the distribution of ϕ estimate. The Dickey–Fuller test [45] uses empirical distribution of the test statistics:

$$DF_t = \frac{\hat{\phi} - 1}{SE(\hat{\phi})},$$

where $\hat{\phi}$ is the OLS estimate.

The augmented Dickey–Fuller test tests unit root for the AR(p) process:

$$Y_t = c + \phi Y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-i} + \varepsilon_t,$$

for the null hypothesis $H_0 : \phi = 1$(unit root). The test statistics are the same as above.

Cointegration and causality

As we have shown, in contrast with the signals studied in SP, many economic time series are nonstationary integrated processes

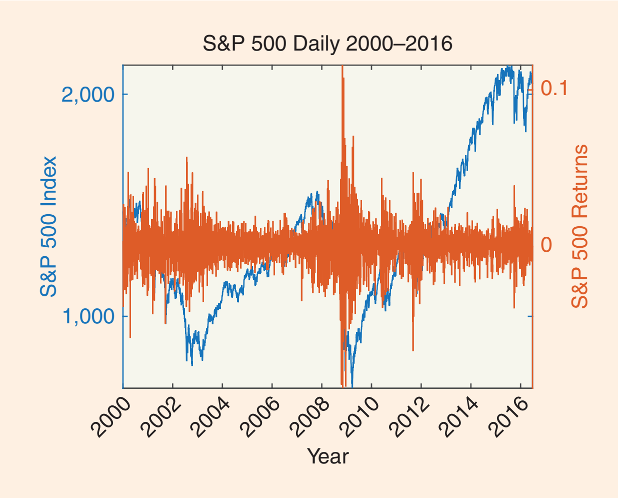

FIGURE 8. The S&P 500 daily prices and returns (2000–2016). When the price is low, the return volatility is high.

with at least one unit root. The integration order is a property of the process and represents a trend of the time series. In SP research, researchers do not care about this trend because they are working on stationary signals. However, a serious problem occurs when examining relationships of multiple nonstationary integrated economic time series.

Consider the following simple linear regression:

$$Y_t = c + \beta X_t + \varepsilon_t.$$

When X_t and Y_t are both unit-root processes with different orders, i.e., when ε_t is also a unit-root process, regression results and statistics become spurious or meaningless. Regression results are only meaningful when X_t and Y_t have a common trend (i.e., the same integration order) or are cointegrated. Clive W.J. Granger's finding of such spurious regressions [46] invalidated many empirical economic studies before the 1970s and, along with his work on cointegration [47], won him the 2003 Nobel Memorial Prize in Economic Sciences. For cointegrated time series, there must exist a linear combination of them that is stationary. The Engle–Granger test [47] applies the Dickey–Fuller unit-root test to examine the cointegration of multiple time series.

The time-series model helps capture correlations among time series but does not find causal relationships. In SP, inputs cause outputs because information flows are obvious in a physical system. However, in analyzing economic data or any data from nonphysical systems, such as a social network, causal relationships are not obvious and cannot be taken for granted. Yet identifying such causal relationships is of great importance to discover information hidden in the data and is necessary for decision making in many big data applications.

The causality relationship is always difficult to define and quantify. Granger gives a definition from the time-series perspective [48], [49]. If we agree that the cause must occur before the effect, the Granger noncausality (or strong exogeneity) can be defined if the following equation holds for the conditional mean:

$$\mathbb{E}[Y_{t+1} | Y_t, X_t, X(t-1), \cdots] = \mathbb{E}[Y_{t+1} | Y_t].$$

If we further agree that for X_t to be the cause of $Y(t+1)$, it must contain unique information about Y_{t+1}, then X_t is said to Granger-cause Y_{t+1} if for some A,

$$P(Y_{t+1} \in A | \Omega_t, X_t) = P(Y_{t+1} \in A | \Omega_t - X_t),$$

where P is the probability and Ω_t represents all knowledge in the universe available at time t. Note that Granger causality is only one of many definitions on causality, but it is statistically testable using a time-series model (VAR or ARDL model), making it instrumental in causality analysis.

Generalized AR conditional heteroskedasticity models

As we already discussed, volatility is a fundamental risk quantity that needs to be estimated in finance, especially in risk modeling and option pricing. As has been observed, the historical volatility of a time series changes over time. See Figure 8 for S&P 500 daily prices and returns.

The time-varying nature of volatility is called *heteroskedasticity*. Robert F. Engle invented the AR conditional heteroskedasticity (ARCH) model to capture the time-varying dynamics of volatility, winning the 2003 Nobel Memorial Prize in Economic Sciences. The qth order ARCH(q) model for a zero-mean normally distributed asset return time series, Y_t, with time-varying volatility σ_t is specified as

$$Y_t \sim N(0, \sigma_t^2),$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2.$$

It is indeed an *MA* model for time-varying variance.

Adding an AR term for σ_t, the generalized ARCH (GARCH) model [50], GARCH(p, q), is defined as

$$Y_t \sim N(0, \sigma_t^2),$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2.$$

Many variations of the GARCH model have been subsequently developed and widely used in risk models, high-frequency volatility models, large-scale multivariate ARCH models, and derivative pricing models [51].

Relationships between SP and econometric models

The time-series analysis has achieved great success in economics, finance, and business studies. It is encouraging for SP researchers, as the time-series models and basic concepts are essentially similar to SP models. Indeed, the Granger causality concept was partly inspired by Nobert Wiener [48], [52]. Many other models used in econometrics [14], [53], including spectrum analysis, Kalman filtering, Markov models, maximum likelihood and Bayes methods, particle filtering, and sequential Monte Carlo methods, are similar to those in SP as well.

However, researchers need to note major differences when applying these models to economic and finance applications.

1) As discussed previously, model specification is more important in economics and finance than in SP applications, because model specification is often not so obvious and testable in economics as it is in SP and the joint hypothesis problem always exists.
2) Statistical analysis and tests are always necessary for model verification and applications. Just an estimated number is not enough. Distributions and confidence intervals of parameter estimates need to be analyzed. Equally important is the distribution of the error terms, which often represent money and risks in financial decisions.
3) In economic applications, researchers often operate on nonstationary time series, and the causal relationships are not obvious and may be difficult to test.
4) In empirical studies, data are noisy and the SNR is low because there are often too many factors (interferences) in an economic system. For example, the Sharpe ratio of the U.S. market S&P 500 index is around 0.4, which roughly translates to an SNR of -8 dB.
5) The data traps discussed previously always exist in economic studies. Application-dependent conceptual justifications, alternative model testing, and out-of-sample tests are always necessary because, theoretically, multiple models could fit finite data well.
6) Although stock price forecasting and trading strategies using SP tools are attractive, these are probably the most difficult tasks because the price itself is, after all, a signal for resource allocation and is sensitive to all market actions, trading strategies, and information. The efficient market concept is a logically good approximation of information processing in the financial market. Conversely, many other economic and business time series and processes, such as production, sales, earnings, investments, and so on, may be easier to forecast and model.

Examples of applying SP to market data

In this section, we provide some examples of applying SP to economics, finance, and marketing studies based mainly on our own experiences. As such, they are only illustrative and by no means comprehensive. These examples, in our view, represent a few broader cutting-edge directions to which SP can contribute significantly.

Market evolution analysis using SP models

In financial markets and general economic systems, market participants can only passively observe the system input–output dynamics but are not able to manage input factors, as we can often do in SP systems. For example, a macroeconomic event, such as unemployment or interest rate change, as a market input can move the financial market in a certain way, but market participants cannot actively design and use such inputs to achieve desired market movement (output) unless they are central bankers or policymakers. Even policymakers are restricted in their power of managing input factors and cannot change many inputs, such as consumer confidence. In contrast, market participants in commercial product markets can actively manage many market inputs, such as a firm's marketing budget, product pricing, and research and development (R&D) investment, to achieve desired market outputs, such as sales targets. SP models are therefore poised to be useful in these types of product marketing research, e.g., to analyze and model market dynamics and optimize marketing strategies and investments. We present a market evolution analysis as an example to demonstrate the use of SP methods in marketing research.

In a commercial product market, when using a time-series model for market dynamics, first we would be able to observe the output of the market response [54]. For a company, among the most important observable market output time series is product sales. Evolving sales are always desirable, i.e., sales are a unit-root process, not a stationary process. However, many causes can lead to evolving sales, and managers need to know how to maintain sales performance. Meanwhile, market inputs, such as advertising, price cuts, promotions, R&D, and competitor investments, may affect market outputs, such as sales. Managers can indeed actively manage their marketing budgets and decisions. They search for market opportunities in which positive effects of a one-time investment can persist at least partly (i.e., have a hysteresis effect) and hope that the negative effect of a price war or product defect is short lived and has no hysteresis. Accurately understanding market input–output relationships is critical for managers to make the right investment decisions and take real-time actions to capture fleeting market opportunities.

For simplicity, we use the advertising-sales relationship as an example of the general marketing and business input–output dynamics. To clearly demonstrate basic concepts, first-order models are used whenever possible. Our first objective is to discriminate the different market dynamics, as shown in Figure 9.

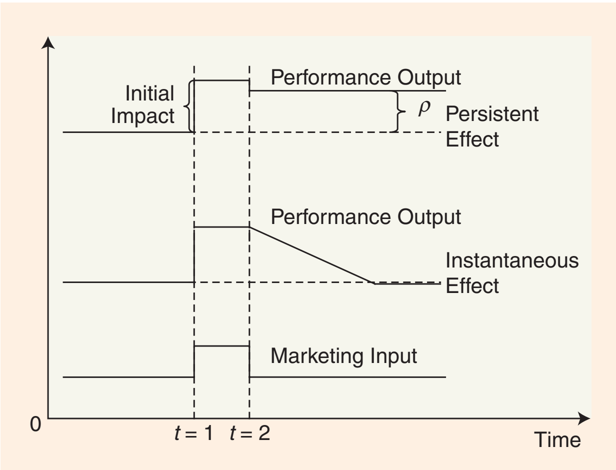

FIGURE 9. The different market input–output dynamics.

To identify market output characteristics, conventionally the following first-order time-series model is used to test whether the sales S_t is evolving (a unit-root test with null hypothesis $H_0 : \phi = 1$):

$$S_t = \mu + \phi S_{t-1} + e_t.$$

When the sales S_t is evolving, managers need to know the role of the advertising investment A_t in creating and supporting sales evolution to budget advertising investment. The following market input–output model is established given known A_t:

$$S_t = c + \alpha S_{t-1} + \beta A_t + e_t.$$

Note that the sales S_t and advertising A_t are monetary value time series. To examine whether the sales evolution is intrinsic (i.e., a favorable market feature) or supported by advertising (i.e., a less favorable market feature with which continuous advertising needs to be budgeted), an intrinsic market evolution (IME) test is proposed [10] to test null hypothesis $H_0 : \alpha = 1$. If H_0 is rejected, the market is intrinsically stationary; i.e., a short-term advertising investment has a short-term instantaneous effect on sales. Sales evolution needs to be supported by the persistent advertising spending. In SP language, the system function has no poles on or outside the unit circle. The unit root (pole) of the output $S(Z)$ is indeed generated by the unit root of input $A(Z)$. In reality, when advertising of a product increases, sales increase, but when advertising is withdrawn, sales revert to the original level. Marketing managers need to evaluate the investment return to maintain the most profitable sales level.

Conversely, if H_0 is not rejected with the IME test, the market is inherently evolving. A short-term advertising investment will have a long-term persistent effect on sales. In SP language, the system function itself has poles on or outside the unit circle. The unit root (pole) of the output $S(Z)$ is indeed generated by the intrinsic market dynamic. Such a phenomenon is not common in SP systems, and researchers may question why it is important to identify intrinsically evolving markets. In the economic and business world, an intrinsically evolving market indicates a more favorable business environment in which a short-term advertising campaign drives sales up; when the advertising is withdrawn, sales maintain their level or even continue to grow. Such a phenomenon could be caused by the intrinsically superior product characteristics that retain loyal customers and attract new customers by word of mouth, or a growing emerging market yearning for such products. Short-term advertising only acts as ignition. When an intrinsically evolving market is identified, managers can increase their marketing campaign budgets to capture such opportunities in time [12].

In addition to identifying the market nature, SP models build budgeting strategy. For example, a simple percentage budgeting model is

$$A(t) = f(S(t-1)) = \gamma_b S(t-1).$$

That is, the marketing budget is based on sales of the previous period. By this budgeting model, the optimal budgeting sequence can be quantified according to various constraints and costs to support the sales evolution in an intrinsically stationary market. Figure 10 shows an example of various advertising budget effects with different γ_b given $\alpha = 0.8$. More detailed methodology, analysis, and applications are available in [10]–[12].

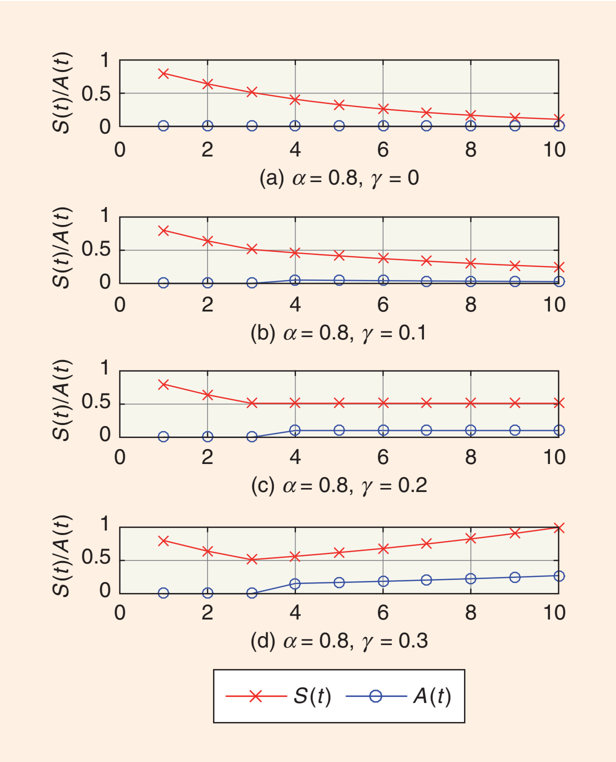

FIGURE 10. The different sales effects of percentage advertising budget practices with different γ_b given $\alpha = 0.8$, i.e., in an intrinsically stationary market. (a) The sales die out without maintaining advertising. (b) The sales still die out at a slower rate (along with the advertising) with inadequate advertising percentage budget. (c) The continuous advertising budget as a percentage of sales induces and supports the sales evolution. (d) The continuous advertising budget as a larger percentage of sales leads to increasing sales (and subsequent increasing advertising).

Real-world marketing dynamics can be more sophisticated. With possible lag effects of market inputs and outputs and multiple variables, multivariate [12] and higher-order models [11] can be built based on these concepts. With more complexity in data and model analysis, SP model-based econometric analysis can certainly play an instrumental role in quantitative marketing and business research.

Time-varying risk models based on Kalman filtering and Gaussian processes

Time-varying systematic risk analysis based on Kalman filtering

In financial markets, the systematic risk is represented by the market risk beta in the CAPM or the multifactor betas in the FF three-factor model, including market beta, size beta, and value beta. In practice, the ordinary least squares (OLS) time-series regression is used to estimate beta by assuming constant beta over the observation period. Prior research has shown evidence of time-varying beta [55]. The OLS time-series regression does not account for any significant economic event that could affect systematic risks of an asset.

The piecewise mean-reverting (PMR) model is based on the observations of the FF three-factor systematic risk behaviors in empirical tests. The betas tend to jump with relevant significant events and revert to their means with different rates depending on the type of events. The reverting rate indicates how quickly the systematic risk of a stock recovers from a sudden change. For the CAPM model, the PMR model consists of a system equation that represents the PMR dynamic of the hidden time-varying variable beta and an observation equation of the stock return given by the CAPM. The system equation is

$$\beta_t = (1-\phi_t)\bar{\beta} + \phi_t \beta_{t-1} + z_t u_t + \xi_t,$$

where $\bar{\beta}$ represents the average β over time and ϕ_t is the mean-reverting rate in the range of [0,1]. The larger is ϕ_t, the slower is the mean reverting. The jump process is based on the Bernoulli random variable z_t, and a zero-mean normally distributed variable, $u_t \sim N(0, \sigma_u^2)$, represents the amount of jump, i.e., z_t takes the value of 1 with a given probability p and the value of 0 with probability $1-p$; $\xi_t \sim N(0, \sigma_\xi^2)$ represents a random perturbation of β. The observation equation is simply the CAPM, i.e., the asset return time series

$$R_t - r_f = \alpha_t + \beta_t (R_{M,t} - r_f) + \varepsilon_t.$$

This model assumes that significant economic events can lead to the abnormal changes in beta. A modified Kalman filter can be used to estimate and track the PMR beta [56], [57]. In addition, the methodology can be extended to multifactor models, such as the FF three-factor model [57].

As previously discussed, model validation is always an open issue. To achieve validation, researchers can compare the model with alternative models and try to obtain the real-world data to examine whether betas change when major events occur. However, such case studies are confirmatory rather than conclusive.

Gaussian process regression stochastic volatility model

Volatility modeling is one of the most active research areas of financial time series. With the recent development of Bayesian nonparametric modeling in SP and machine-learning communities, flexible tools and modeling methods, such as the Gaussian process (GP) [58]–[60] and copula process [61], can be applied to model financial data volatility.

By combining the GP state-space modeling framework with the stochastic modeling concept [62], a GP regression stochastic volatility (GPRSV) model can be built to solve the problem of modeling and predicting time-varying variance of financial time-series data. In GPRSV models, a GP prior is placed over the state transition function, and the state transition function is a random function sample from the GP. It is therefore not limited to a fixed linear AR form, as in GARCH-type models and pure stochastic volatility (SV) models. For a zero-mean normally distributed asset return time series, Y_t, with time-varying variance σ_t^2, a GPRSV model is represented by the following set of equations:

$$\begin{aligned} a_t &= Y_t - \mu = \sigma_t \varepsilon_t, \\ v_t &= \log(\sigma_t^2) = f(v_{t-1}) + \tau \eta_t, \\ f &\sim \mathcal{GP}(m(x), k(x, x')), \\ \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} &\sim N(0, \Sigma), \\ \Sigma &= \begin{bmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{bmatrix}, \end{aligned}$$

where μ is the mean of Y_t, a_t is the innovation of the return series, v_t is the logarithm of variance σ_t^2 at time t, and ε_t and η_t are independent and identically distributed zero-mean standard Gaussian distributed white noises. A well-known asymmetric effect is called *financial leverage*, representing a negative correlation between today's return and tomorrow's volatility [63], [64]. This asymmetric leverage effect is captured by the correlation ρ between ϵ_t and η_t. The unknown parameters τ and ρ are to be estimated. The hidden state transition function f is assumed to follow a GP, defined by the mean function $m(x)$ and covariance function $k(x, x')$. The parameters in $m(x)$ and $k(x, x')$ are called *hyperparameters*. For example, if the mean function is defined as $m(x) = cx$, then c is a hyperparameter. The mean function $m(x)$ can encode prior knowledge of system dynamics. Figure 11 shows the graphical model representation of a GPRSV model.

Particle-filter-based Markov chain Monte Carlo learning methods can be used to efficiently estimate the GPRSV model and make volatility inferences. The GPRSV model demonstrates

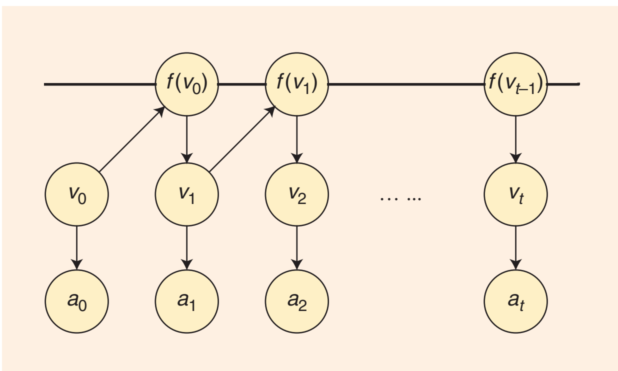

FIGURE 11. A graphical model representation of a GPRSV model of time-varying volatility/ a_t: the observation variable at time t, v_t: the hidden variable (logarithm of volatility), and f_t: the transition function sampled from a GP. The thick horizontal line represents fully connected nodes.

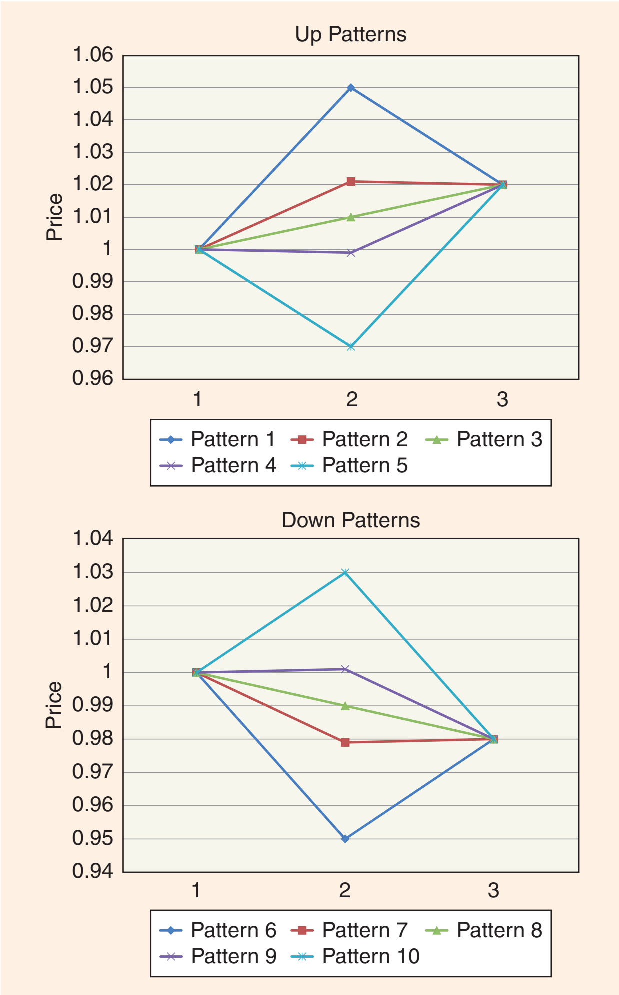

FIGURE 12. The ten preset three-point patterns: five with up direction with different paths and five with down direction with different paths.

Table 1. The statistical analysis of historical stock returns conditional on ten three-point patterns.

				KS Test		t-Test	
Pattern	Matches	Mean	Std	*H*	*p*	*H*	*p*
All	2,855,470	0.000	0.996	0	1	0	1.000
1	189,760	−0.016*	0.969	1	0	1	0.000
2	333,443	0.003	0.937	1	0	0	0.067
3	445,424	0.004*	0.916	1	0	1	0.021
4	362,097	0.010*	0.937	1	0	1	0.000
5	170,181	0.007*	0.938	1	0	1	0.002
6	168,005	−0.001	0.975	1	0	0	0.598
7	337,782	0.011*	1.051	1	0	1	0.000
8	435,628	0.012*	1.106	1	0	1	0.000
9	306,364	−0.032*	1.076	1	0	1	0.000
10	163,672	−0.034*	1.012	1	0	1	0.000

Null hypothesis H_0: the mean return is not significantly different from zero. The null hypothesis is rejected in eight of ten patterns (significant values are marked by an asterisk under $\alpha = 0.05$).

superior volatility prediction performance with both simulated and empirical financial data compared with alternative GARCH and SV models. The previous examples show that SP can contribute to investigating time-varying characteristics of finance and economic systems.

Big data analysis of financial data based on SP

Returning to stock returns, we present some empirical statistical facts to illustrate the potential information that can be sifted from financial big data.

Does the path matter in momentum patterns?

Stock price momentum is a well-known phenomenon in which the stock return continues its direction in the short run. The momentum factor is included in the Carhart four-factor model [25]. From an SP perspective, the momentum is a simple two-point pattern. A natural question is, would multiple point patterns (paths) predict the stock return behavior?

The data set is CRSP monthly price data from 1965 to 2012. The CRSP data set is the gold standard in historical stock return research. To compare stock returns of different stocks, all stock returns are normalized using the methodology in [6]. Ten preset three-point price patterns are constructed. Specifically, ten groups of three-point price patterns in the data set are created according to their correlation with each of the ten preset patterns, as shown in Figure 12. The correlation similarity threshold is 0.95. The subsequent one-month stock returns conditional on each of the ten three-point patterns are analyzed. The null hypothesis H_0 is that the mean conditional return of the subsequent month is zero, i.e., the same as unconditional returns. Both a nonparametric Kolmogorov–Smirnov (KS) test and parametric t-test are conducted. Table 1 summarizes the results.

There are more than 2.8 million three-point patterns. The means of all stock returns are normalized to zero. As can be seen, eight of ten conditional returns have statistically significant means with different directions. The p-values are reliable given the large number of samples. This statistical fact shows that in historical stock return data, 1) the three-point patterns contain information about future returns and 2) the path does matter in addition to the two-point momentum pattern.

That said, the results do not necessarily mean that one can profit from these patterns because potential constraints, such as liquidity and transaction costs, exist to prevent trading profit, and the results are just a summary of historical data. Rather, these results along with the results presented in [6] show that fine structures exist in stock returns and markets (more than traditional economic research indicates) and that SP can provide powerful analytical tools.

Can the past inform about the future?

All theories and models are based on historical data with the assumption that the past can represent the future. In SP, researchers do not often ask the question of whether the past can inform about the future because they are confident about natural physical laws. However, in financial markets, people have always been skeptical about models and statistical results summarized from historical data, especially so after the 2008 financial crisis. Some people have even gone so far as to say that the past cannot inform about the future, and therefore models and statistics coming from historical data are not useful or trustworthy.

In the following, we show some interesting empirical results as a small step forward to answer this question. The stock daily returns of Russell 3000 component stocks from January 1995 to September 2014 are examined. The Russell 3000 is used because those stocks are relatively liquid and the price data are a good reflection of real market transactions. Note that the component stocks of the Russell 3000 are changing over time. The test period is set to about four years from January 2011 to September 2014. Starting from the first trading day of 2011, the return distribution of the historical data of the past 16 years is used to predict the return distribution of the next day. We want to examine overall how accurate it is to use the historical stock return distribution as a representation of the future distribution. The following procedure is used to examine the differences between historical distributions and future distributions.

All theories and models are based on historical data with the assumption that the past can represent the future.

At time t, we use all historical daily returns of the preceding 16 years to create a distribution f. We then define K equal probability quantile bins $Q_k, k = 1, \ldots, K$ with bin (stock return) boundaries $q_1, \ldots q_{K-1}$. Apparently, $P(R_\tau \in Q_k) = (1/K), \forall \tau \le t$. If the future distribution is the same as the past distribution, we can expect that at time t, the returns of the 3000 stocks will fall into each bin uniformly, i.e., $P(R_t \in Q_k) = (1/K)$. The chi-square test can be conducted to test this hypothesis. We aggregate all bin counts and plot two scenarios in Figure 13 with 50 bins, i.e., $K = 50$. Figure 13(a) shows the performance of the past unconditional distribution. Figure 13(b) shows the performance of the past distribution conditional on the preceding ten-day patterns.

The bar charts should be flat if the past distribution and the future distribution are the same. Given the large number of samples, by chi-square tests, we can statistically reject the null hypothesis that the future distribution is the same as the past distribution for the testing periods. Note that both distributions underestimate tail risks. The statistical test results justify people's concern that the past data do not represent the future. However, as Figure 13 shows the past distribution does contain some information about the future, e.g., the mean values. In addition, the past distributions conditional on the preceding ten-day price pattern contain more information about the future than the unconditional past distributions.

These findings from the data are encouraging for SP because they indicate that sophisticated structures containing information in data need to be identified and understood. Meanwhile, these findings also pose challenges because hypotheses and systems in these economic data are different from those in typical SP applications. Related SP problems and solutions for such big data and modeling applications need to be carefully formulated.

Other related works in SP

We have mainly focused on the literature in economics and business research and provided the SP understanding on the literature. Note that there have been three special issues on SP for finance [1]–[3]: one in *IEEE Signal Processing Magazine* in 2011, and the other two in *IEEE Journal of Selected Topics in Signal Processing* in 2012 and 2016. Readers can find more SP examples there, especially on portfolio and risk analysis, HFT, and algorithmic trading. When going through SP technologies, readers can focus more on the economic problem formulation and evaluations of solutions with the concepts discussed in this tutorial. Also, an overview of business analytics related to SP can be found in [65] and [66], which provide perspectives on system modeling of a business.

Conclusions and thoughts

In this tutorial, we present an introduction to some fundamental economic theories that govern financial markets and people's economic decisions, including expected utility, RP, portfolio theory and asset-pricing models, EMH, prospect theory, and behavioral economics. We also introduce basic econometric tools and theories from an SP perspective.

We emphasize that when analyzing data and building economic models, researchers should keep in mind existing economic theories and hypotheses. The burden of proof is high when findings contradict or are inconsistent with existing theories. For example, any price anomaly found in financial markets needs to be carefully examined because it may contradict existing models and the EMH and could be a chance result. It is always difficult to predict prices because these are usually determined by market equilibrium no matter how

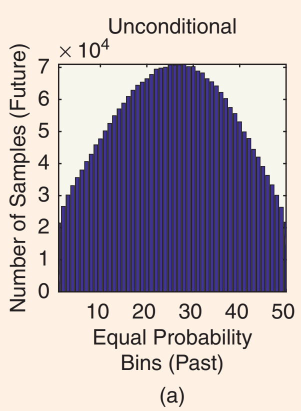

FIGURE 13. The bar charts should be flat if the past distribution and the future distribution are the same.

unreasonably they may seem. Excess economic benefits are associated with risks.

Time-series models, system models, and statistical models are certainly useful tools to analyze economic data. SP researchers can try to formulate the economic problems into filtering, array SP, impulse responses, finite-state machine, Markov models, and state-space models. Whenever a lagged transform seems to be useful, researchers can consider the *Z* transform, Laplace transforms, and spectral analysis, which may provide additional insights into economic systems and problem solutions.

Caution also needs to be exercised when formulating input–output relationships and objective functions for an economic system. In many cases, humans are only observers of the system and have little power to change the input, e.g., consumer price indices, or the output, e.g., sales or GDP. Meanwhile, humans are oftentimes also part of an economic system, and any human actions or predictions may change the system or system equilibrium. For example, a buying order for a stock may inevitably change the price of the stock. There are times when humans are indeed actively responsible for the input of the system, such as an investment budget decision on R&D or marketing for a company. The different nature of inputs and outputs shapes not only problems and models but also the methodology to the solution and interpretation of results.

In terms of mathematical formulations and modeling tools, many similarities exist between SP and econometrics. However, the weaknesses of SP methodology come from what is often taken for granted in SP, such as the knowledge of physical systems, ground-truth data, controlled experiments, or the confidence in simulated data. Mean-squared errors or variances are often good statistics in SP performance evaluation but are not sufficient in economic and business studies. Indeed, when making investment suggestions and decisions, i.e., putting one's money to bet on a model, researchers need to understand more of the probability and statistical significance of their predictions, estimation, data samples, model validity, etc. They need to understand causal relationships and examine all possible alternative models and hypotheses about the data because an accurate mechanism of a social or economic system is always unknown.

In the big data era, there is a wealth of data analysis and processing work for which SP tools are useful, e.g., denoising and data cleaning, feature extraction, pattern detection, tracking, and abnormality detection. Note that traditional economics are more about equilibrium, and SP can play a big role in understanding the process and transition patterns to reach equilibrium from data analysis.

Indeed, numerical data are at the core of SP. SP is promising in exploring economic big data to learn information better and faster and in finding sophisticated subtle structures within data. Researchers need to be careful not to be biased by the data at hand.

As SP research and applications are expanding, SP researchers can broaden their vision and take advantage of the methodologies and advancements in econometrics, economics, and statistical methods for SP problems. For example, game theory has been an active research tool in communication systems and resource allocation. More cross-pollination between SP and economics will produce fruitful results in research as well as practical applications affecting people's daily lives.

Acknowledgments

We thank EidoSearch Inc. (www.eidosearch.com) for providing the empirical test results in the section "Big Data Analysis of Financial Data Based on SP."

Authors

Xiao-Ping (Steven) Zhang (xzhang@ryerson.ca) received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in electronic engineering and the M.B.A. degree in finance, economics, and entrepreneurship from the University of Chicago, Illinois. He is a professor of electrical and computer engineering and is cross appointed to the Finance Department at the Ted Rogers School of Management at Ryerson University, Toronto, Canada. His research interests include signal processing, electronic systems, machine learning, big data, finance, and marketing. He is the cofounder and chief executive officer for EidoSearch, an Ontario-based company offering a content-based search and analysis engine for financial big data.

Fang Wang (fwang@wlu.ca) received the M.B.A. degree in finance from the University of Texas, San Antonio, and the Ph.D. degree in management information systems from Michael G. DeGroote School of Business, McMaster University, Hamilton, Ontario, Canada. She is an associate professor of marketing at the Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, Ontario, Canada. Her research interests include market response models, marketing data mining/analysis, long-term marketing productivity, information systems management, big data, and social media applications in marketing and finance.

References

[1] *IEEE Signal Process. Mag. (Special Issue on Signal Processing for Financial Applications)*, vol. 28, no. 5, Sept. 2011.

[2] *IEEE J. Sel. Topics Signal Process. (Special Issue on Methods in Finance and Electronic Trading)*, vol. 6, no. 4, Aug. 2012.

[3] *IEEE J. Sel. Topics Signal Process. (Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading)*, vol. 10, no. 6, Sept. 2016.

[4] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, May 1970.

[5] J. H. Cochrane, *Asset Pricing*, rev. ed. Princeton, NJ: Princeton Univ. Press, 2005.

[6] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *J. Finance*, vol. 55, no. 4, pp. 1705–1765, Aug. 2000.

[7] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, Mar. 1979.

[8] R. J. Shiller, "Do stock prices move too much to be justified by subsequent changes in dividends?" *Am. Econ. Rev.*, vol. 73, no. 1, pp. 236–237, Mar. 1983.

[9] E. F. Fama, "Market efficiency, long-term returns, and behavioral finance," *J. Finance Econ.*, vol. 49, no. 3, pp. 283–306, Sept. 1998.

[10] F. Wang and X. P. Zhang, "Reasons for market evolution and budgeting implications," *J. Marketing*, vol. 72, no. 5, pp. 15–30, Sept. 2008.

[11] F. Wang, X. P. S. Zhang, and M. Ouyang, "Does advertising create sustained firm value? The capitalization of brand intangible," *J. Acad. Marketing Sci.*, vol. 37, no. 2, pp. 130–143, 2009.

[12] D. M. Hanssens, F. Wang, and X. P. Zhang, "Performance growth and opportunistic marketing spending," *Int. J. Research Marketing*, vol. 33, no. 4, pp. 711–724, Dec. 2016.

[13] E. Budish, P. Cramton, and J. Shim, "The high-frequency trading arms race: Frequent batch auctions as a market design response," *Quarterly J. Economics*, vol. 130, no. 4, pp. 1547–1621, Nov. 2015.

[14] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.

[15] H. Levy and H. M. Markowtiz, "Approximating expected utility by a function of mean and variance," *Am. Econ. Rev.*, vol. 69, no. 3, pp. 308–317, 1979.

[16] Y. Kroll, H. Levy, and H. Markowitz, "Mean-variance versus direct utility maximization," *J. Finance*, vol. 39, no. 1, pp. 47–61, 1984.

[17] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[18] W. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, 1964.

[19] J. Lintner, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Rev. Economics Statistics*, vol. 47, no. 1, pp. 13–37, 1965.

[20] F. Black, "Capital market equilibrium with restricted borrowing," *J. Business*, vol. 45, no. 3, pp. 444–455, 1972.

[21] E. F. Fama and K. R. French, "The capital asset pricing model: Theory and evidence," *J. Economic Perspectives*, vol. 18, no. 3, pp. 25–46, Sept. 2004.

[22] J. V. Jacob Viner, "Cost curves and supply curves," *Zeitschrift fr Nationalkonomie/J. Economics*, vol. 3, no. 1, pp. 23–46, 1931.

[23] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *J. Financial Economics*, vol. 33, no. 1, pp. 3–56, Mar. 1993.

[24] E. F. Fama and K. R. French, "The cross-section of expected stock returns," *J. Finance*, vol. 47, no. 2, pp. 427–465, June 1992.

[25] M. M. Carhart, "On persistence in mutual fund performance," *J. Finance*, vol. 52, no. 1, pp. 57–82, 1997.

[26] E. F. Fama and K. R. French, "A five-factor asset pricing model," *J. Financial Economics*, vol. 116, no. 1, pp. 1–22, 2015.

[27] E. F. Fama, "Multifactor portfolio efficiency and multifactor asset pricing," *J. Financial Quantitative Analysis*, vol. 31, no. 4, pp. 441–465, 1996.

[28] R. C. Merton, "An intertemporal capital asset pricing model," *Econometrica*, vol. 41, no. 5, pp. 867–887, 1973.

[29] H. L. Van Trees, *Detection, Estimation, and Modulation Theory,* part 4, *Optimum Array Processing*. Hoboken, NJ: Wiley, 2002.

[30] G. Ganesan, "A subspace approach to portfolio analysis," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 49–60, Sept. 2011.

[31] M. U. Torun, A. N. Akansu, and M. Avellaneda, "Portfolio risk in multiple frequencies," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 61–71, Sept. 2011.

[32] A. N. Akansu and M. U. Torun, "Toeplitz approximation to empirical correlation matrix of asset returns: A signal processing perspective," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 319–326, Aug. 2012.

[33] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 4, pp. 337–350, Aug. 2012.

[34] D. M. Malioutov, A. A. Corum, and M. Çetin, "Covariance matrix estimation for interest-rate risk modeling via smooth and monotone regularization," *IEEE J. Sel. Topics Signal Process., vol. 10*, no. 6, pp. 1006–1014, Sept. 2016.

[35] M. Senneret, Y. Malevergne, P. Abry, G. Perrin, and L. Jaffrs, "Covariance versus precision matrix estimation for efficient asset allocation," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 6, pp. 982–993, Sept. 2016.

[36] E. Jay, P. Duvaut, S. Darolles, and A. Chrétien, "Multifactor models," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 37–48, 2011.

[37] B. Malkiel, *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*, 10th ed. New York: Norton, 2011.

[38] M. S. Fischer Black, "The pricing of options and corporate liabilities," *J. Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.

[39] J. Y. Campbell and R. J. Shiller, "The dividend-price ratio and expectations of future dividends and discount factors," *Rev. Financial Studies*, vol. 1, no. 3, pp. 195–228, 1988.

[40] J. H. Cochrane, "Presidential address: Discount rates," *J. Finance*, vol. 66, no. 4, pp. 1047–1108, 2011.

[41] R. J. Shiller, *Irrational Exuberance*, 2nd ed. Princeton, NJ: Princeton University Press, 2005.

[42] J. H. Cochrane. Eugene Fama: Efficient markets, risk premiums, and the Nobel Prize. presented at the University of Chicago Syposium, Nov. 4, 2013. [Online]. Available: https://faculty.chicagobooth.edu/john.cochrane/research/papers/Fama_panel_nov_2013.pdf

[43] X. P. Zhang and D. Kedmey, "Techware: Financial data and analytic resources," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 138–141, Sept. 2011.

[44] R. Tsay, *Analysis of Financial Time Series*. Hoboken, NJ: Wiley, 2010.

[45] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Statist. Assn.*, vol. 74, no. 366, pp. 427–431, 1979.

[46] C. Granger and P. Newbold, "Spurious regressions in econometrics," *J. Econometrics*, vol. 2, no. 2, pp. 111– 120, 1974.

[47] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.

[48] C. Granger, "Testing for causality: A personal viewpoint," *J. Economic Dynamics Control*, vol. 2, pp. 329–352, 1980.

[49] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969. [Online]. Available: http://www.jstor.org/stable/1912791

[50] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, pp. 307–327, 1986.

[51] R. Engle, "New frontiers for arch models," *J. Applied Econometrics*, vol. 17, no. 5, pp. 425–446, 2002.

[52] N. Wiener, "The theory of prediction," in *Modern Mathematics for the Engineer*, E. F. Beckenbach, Ed. New York: McGraw-Hill, 1994.

[53] D. Creal, "A survey of sequential Monte Carlo methods for economics and finance," *Econometric Revs.*, vol. 31, no. 3, pp. 245–296, 2012.

[54] D. Hanssens, L. Parsons, and R. Schultz, *Market Response Models: Econometric and Time Series Analysis* (International Series in Quantitative Marketing). New York: Springer-Verlag, 2012.

[55] S. Sunder, "Stationarity of market risk: Random coefficients tests for individual stocks," *J. Finance*, vol. 35, no. 4, pp. 883–896, 1980.

[56] T. Rajbhandary, X. P. Zhang, and F. Wang, "Piecewise constant modeling and Kalman filter tracking of systematic market risk," in *Proc. 2013 IEEE Global Conf. Signal Information Processing (GlobalSIP),* p. 1144.

[57] L. Vo, X.P. Zhang, and F. Wang, "Multifactor systematic risk analysis based on piecewise mean reverting model," in *Proc. 2013 IEEE Global Conf. Signal Information Processing (GlobalSIP),* p. 1142.

[58] J. Han, X. P. Zhang, and F. Wang, "Gaussian process regression stochastic volatility model for financial time series," *IEEE J. Sel. Topics Signal Process.* (*Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading*), vol. 10, no. 6, Sept. 2016.

[59] R. Frigola, Y. Chen, and C. Rasmussen, "Variational Gaussian process state-space models," in *Proc. Advances Neural Information Processing Systems*, 2014, pp. 3680–3688.

[60] Y. Wu, J. M. Hernández-Lobato, and Z. Ghahramani, "Gaussian process volatility model," in *Proc. Advances Neural Information Processing Systems*, 2014, pp. 1044–1052.

[61] A. Wilson and Z. Ghahramani, "Copula processes," in *Proc. Advances Neural Information Processing Systems*, 2010, pp. 2460–2468.

[62] S. J. Taylor, "Modeling stochastic volatility: A review and comparative study," *Mathematical Finance*, vol. 4, no. 2, pp. 183–204, 1994.

[63] A. A. Christie, "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *J. Financial Economics*, vol. 10, no. 4, pp. 407–432, 1982.

[64] G. Wu, "The determinants of asymmetric volatility," *Rev. Financial Studies*, vol. 14, no. 3, pp. 837–859, 2001.

[65] K. R. Varshney and A. Mojsilovi, "Business analytics based on financial time series," *IEEE Signal Process. Mag.*, vol. 28, no. 5, pp. 83–93, Sept. 2011.

[66] K. R. Varshney. (2012). Introduction to business analytics. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing tutorial. [Online]. Available: https://informationashvins.files.wordpress.com/2012/04/varshney_icassp2012.pdf

[67] J. H. Cochrane. (2008, Nov. 12). Is now the time to buy stocks? *Wall Street Journal*. [Online]. Available: https://www.wsj.com/articles/SB122645226692719401