Jau-Lian Jeng

EMPIRICAL ASSET PRICING MODELS

Data, Empirical Verification, and Model Search

Empirical Asset Pricing Models

Jau-Lian Jeng

Empirical Asset Pricing Models

Data, Empirical Verification, and Model Search

palgrave
macmillan

Jau-Lian Jeng
School of Business and Management
Azusa Pacific University
Stevenson Ranch, CA, USA

ISBN 978-3-319-74191-8 ISBN 978-3-319-74192-5 (eBook)
https://doi.org/10.1007/978-3-319-74192-5

Library of Congress Control Number: 2017964504

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book discusses several issues concerning the construction of empirical asset pricing models, including: (1) the setting of essential properties in asset pricing models of stock returns, (2) the statistical inferences that can be applied to verify the necessary properties of empirical asset pricing models, and (3) the model search approach where any model can be considered as only a tentative approximation for asset returns given their time-changing nature.

The main aim of the book is to verify that statistical inferences and time series analysis for asset returns should not be confined to the verification of certain structures or variables based simply on statistical significance alone. These statistical verifications can only be meaningful if the intent or hypothesis for the model is related to the properties developed in the theoretical setting of asset pricing models where systematic components of asset returns are considered.

Blaming the existing models for their deficiency or lack of forecasting superiority is not necessarily a solid way to refute the theories. In fact, unless we have some solid understanding of the ultimate mechanism of stock returns, it is premature to claim the depletion of current existing theories based only on predictability or forecasting. A rigorous justification must originate from more profound alternatives that may belittle the currently existing theoretical framework. Profitability (through forecasting, for instance) can't even be a unique determinant for the validity of empirical models on asset returns.

Speculative profits (through forecasting) may result from technical analysis where no theoretical background of financial economics (or anything else) is discussed at all. Superiority in forecasting with certain proposed models or mechanisms may prevail with short-term horizons among different data sets. Yet, it is not surprising to find that this advantage quickly resolves over time which entails the needs to update and modify the presumed models continuously. Thanks to their properties, this is precisely why financial markets are sufficiently interesting to attract enormous resources in exploring the quintessence of their evolving mechanism. What is really essential for empiricists is how to accommodate this possibly time-changing nature of stock returns, and to strive for the pricing kernels with meaningful interpretation of them.

Part I of this book covers the essential properties of theoretical asset pricing models, especially when linear (factor-pricing) models are of interest. Since the focus of the book is on empirical asset pricing models, only discrete-time models are discussed. From the theoretical issues, the conventional specification tests are also discussed with their possible implications for the models of interest. This leads to the discussion of model searching with various model selection criteria where emphases are mainly of reduction of dimensionality and predictability.

Given the pitfalls of these model selection criteria, Part II provides an alternative methodology where various justifications of the cross-sectional properties of stock returns is emphasized and additional model searching is devised with the specification tests provided. Hence the aim of this book is to reconsider the necessary cautions involved in the analyses of empirical asset pricing models and to provide some alternatives. The book may be used as a technical reference for researchers, graduate students, and professionals who are interested in exploring the possible alternatives that may provide more tractable methods for empirical asset pricing models for various applications in the future.

Stevenson Ranch, CA, USA Jau-Lian Jeng

Acknowledgments

This book is dedicated to my family and my parents for their support with encouragement and patience regarding my stubborn and unrelenting pursuit of academic goals—even when the environment for the pursuit didn't seem appealing or yielding. I would like to thank, in particular, the editorial assistance of Sarah Lawrence and Allison Neuburger at Palgrave Macmillan. Many thanks are also offered to Dr. Jack Hou for his encouragement, comments, and reviews. Above all, I thank especially the one who said "…I am!" for giving me the inspiration that has endured over the decades of my exile and to recognize my limits.

Introduction

Ever since the pioneering work of the capital asset pricing model, theoretical and empirical discussion on the pricing kernel of asset returns has been huge in the financial economics literature. Although many alternative methodologies and theories have been devised, the difficulty in empirical application of asset pricing models still remains unresolved in many areas such as model instability over different time horizons, variable selection on proxies for factors, and (possibly) applicable robust statistical inferences. It is likely that we will discover that an empirical asset pricing model, once selected, can only apply to a certain time period before the model validity quickly disappears when an extended time horizon or data set is considered.

Unfortunately, this phenomenon seems to prevail in many data sets (domestic or foreign) that are applied. The disappointing results in turn lead to the pervasive discontent with the theoretical foundation of asset pricing models. Emphasis on (time series) predictability becomes the norm for model validity for empirical asset pricing models. With the keen demand for validating empirical asset pricing models, statistical verification (with predictability and specification tests) when certain proxies or variables of interest are used becomes the mainstream for financial time series modeling on asset returns.

Essentially, emphases in finding the common features or characters of asset returns (in an attempt to reduce the dimensionality, for instance) through statistical significance should be dealt with using additional caution since these features, once identified, may only prevail tentatively (or contempornaeously) over the selected time horizon.

Part I surveys (a) the quintessential issues of asset pricing models as the pricing kernels for asset returns and (b) the conventional specification tests that consider the possible reduction of dimensionality with statistical significance, which leads to (c) the importance of model searching for the normal (or expected) returns where model selection criteria are applied. Although various specification tests or model selection criteria have been developed for empirical asset pricing models, few of them emphasize the prerequisite that these included variables (in empirical asset pricing models) should satisfy the systematic properties of pricing kernels such as non-diversifiability so that the separation between normal (or expected) returns and abnormal returns or idiosyncratic risks can be well stated.

In essence, empirical asset pricing models must fulfill a set of more restrictive conditions whereas statistical significance in explanatory power (such as *p*-value) on certain (pre-)selected variables can only be considered as exploratory. After all, as the purpose of empirical asset pricing models is to identify the intrinsic structure that governs the (possibly time-changing) core or pricing kernel of asset returns, statistical inference of the significance of certain variables or structures is not entirely sufficient.

Developments on the conventional studies in testing empirical asset pricing models focus mainly on asymptotic arguments of time series data. However, for the validity of any empirical asset pricing model, the focus should be on whether the set of selected variables or proxies by which one attempts to explain the pricing kernel of asset returns constitutes the cross-sectional (asymptotic) commonality among the asset returns or not. It appears, if experience in empirical finance is applied, that identification of some statistically significant explanatory variables for asset returns is not too difficult to provide.

The difficulty, however, is whether these identified variables or proxies truthfully reveal the essential (cross-sectional) commonality of asset returns or not. What is misleading in many empirical findings is that the essence of asset pricing models as pricing kernels was sacrificed when statistical verification of the significance and predictability of explanatory variables in the presumed models is advocated through time series data.

Notice that this empirical verification (of predictability) is mostly (if not all) based on known or collected time series data. As a matter of fact in empirical finance, even if the verification is carried out through out-of-sample time series data, these data are usually known in advance. In other words, the models are fitted with a given training sample of presumed time horizon. And then, time series forecastability is verified with the left-over data in the data set which the modeler has already obtained. The major dilemma lies in the trade-offs as to whether the model specification on empirical asset pricing models is to find something that may help to describe the (short-run) dynamics of asset returns or to identify the quintessence of pricing kernels when short-run predictability could be sacrificed.

Although these trade-offs are not immediately clear-cut, given the notorious time-changing nature of financial markets, it is unlikely that there exists an omnipotent model that encompasses all others across all time horizons. To the best that can be shown, the winning model (through statistical verification or otherwise) only represents a tentative explanation or approximation for the underlying pricing kernel of asset returns. Time changes everything.

Hence, even with the contemporaneous model that encompasses all other competitive alternatives, the empirical result only shows the current notion for the underlying determinant of asset returns. What is more critical, however, is whether the tentative model obtained helps us understand more about the pricing kernel of the asset returns or not. And perhaps more essentially, it helps us to modify diligently the model(s) for different time horizons or data sets.

In Chap. 1 of Part I, the discussions focus on the conventional linear models for asset returns. Given the enormous volume of literature on asset pricing models, this book only surveys and develops the discussions on parametric model building and variable selection. The recent developments on semi-parametric (factor) modeling for asset pricing are also briefly discussed.

Starting from the capital asset pricing model (CAPM), the methods for reduction of dimensionality are covered where factor-pricing models are typical examples. It is not too difficult to find that the empiricist in applied finance may criticize these models as somewhat useless in the usage of profit-taking transactions. Nonetheless, from the perspectives of the financial economist, this is precisely the result of a properly working market mechanism where the advantage in any attempt at speculative opportunity should quickly resolve to zero. Does this mean that these theories are all useless in empirical application? We can only be sure if we have some better theories to explain the mechanism of capital markets and the ultimate determinants for pricing kernels of stock returns.

Although many alternative approaches such as the nonlinearity and behavioral assumptions are developed, the question to ask is “Are these alternative approaches good enough to substitute for the original models we have?" or "Are they competitive enough to provide better insights for the pricing mechanism of stock returns?" Up to the current date, these known alternatives (or models), although rigorous and promising, remain as supplementaries, but they are inadequate as substitutes for existing theories on the pricing kernels of stock returns.

For empirical asset pricing models, the basic criteria for model building are: (1) the procedures for identifying a (or a group of) proper model(s) should be easy to implement in statistical inferences (or with other analytical tools); (2) these candidate models must have well-established theoretical foundations to support the findings; and (3) they provide further directions to cope with the developing status of information and model searching.

Chapter 2 in Part I, for instance, will discuss the methodologies that are currently applied in empirical asset pricing models on asset returns. The chapter includes up-to-date coverage on theoretical setting and model specification tests developed for empirical asset pricing models. However, it is not difficult (in empirical application) to find that these identified, presumed to be economic, variables may not necessarily provide better specification and forecasts than the application of simple time series modeling of asset returns. Chapter 3 in Part I surveys the model selection criteria in determining the number of factors of asset returns. Chapter 4 in Part II discussed alternative methods for detecting hidden systematic factors without assuming that there exists a correct factor structure. Chapter 5 considers model search in empirical asset pricing models.

As such, the search for empirical asset pricing models cannot be succinctly accomplished with the in-sample statistical inferences over some limited time horizons or data sets. Various model specification tests have been developed toward robust methods in (dynamic) asset pricing models. However, it seems that most analyses emphasize the asymptotic properties from time series perspectives. One possible reason for this is that the shadow of forecastability still plays an essential role in the robustness of empirical asset pricing models. Nevertheless, what is essential in such models is the strength of (cross-sectional) coherence or association for these identified economic variables/factors that possibly describes the intrinsic mechanism or pricing kernel of asset returns.

Given the evolving nature of these pricing kernels, forecastability of presumed models over an out-of-sample time horizon is usually limited. Instead, tractability is the goal for empirical asset pricing models: that model specification should emphasize the capability and properties of the underlying intrinsic mechanism of asset returns (or so-called pricing kernels) to administer and accommodate the model search when various available information is applied. The interest of study should be on what method (or methods) is (are) to apply in the search for empirical asset pricing models which is often perceived as evolving through time where many data sets have been applied to trace them. Hence, a model search for empirical asset pricing models should focus on the fundamental properties that any pricing kernel (based on any available information) should prevail in addition to the statistical significance of certain (economic) variables identified or their forecastability.

Contents

PART I

Asset Pricing Models: Discussions and Statistical Inferences

CHAPTER 1

Asset Pricing Models: Specification, Data and Theoretical Foundation

The author surveys and discusses linear asset pricing models with the intent to identify some sets of variables or factors with reduced dimensionality to approximate the core or pricing kernel of asset returns. A theoretical foundation may start with discussion on factor pricing models where asset returns are projected onto some lower-dimensional sets of factors that possibly explain the major variations of asset returns. The aim is to identify major determinants for the fluctuations of asset returns where these determinants satisfy some systematic properties that ensures their indispensable roles.

Controversies begin with questions of measurability of factors and their justification. Classical issues such as the measurability of market portfolio in the capital asset pricing model (CAPM) and selection of market indices, for instance, all incur the problems of measurability and representation for the verification of a theoretical framework. Developments and extensions of arbitrage pricing theory (APT) and multi-factor asset pricing models do not make the hope of attaining robust asset pricing models any brighter.

Statistical inferences do not always mediate the severity of problems mentioned if caution regarding their limitations is not taken into account. Given that all measurable factors presumed for asset pricing models may contain some measurement errors, it is unlikely that empirical asset pricing models will resolve the difficulty of completeness in model justification. At their best extent, empirical asset pricing models can only mimic the systematic patterns or properties of asset returns that provide the tractability

J.-L. Jeng, *Empirical Asset Pricing Models*,
https://doi.org/10.1007/978-3-319-74192-5_1

and direction where the pursuit of economic explanations on asset returns may be feasible.

For instance, the earlier study by Fama and MacBeth (1973) with two-path regressions of asset returns (although these may be considered biased or inconsistent in some statistical properties) is an example of where justification of presumed factor(s) in asset pricing models should undertake further cautious verification in their systematic properties rather than statistical inferences based on significance levels (such as p-values) for any proposed/identified regularity of the data. Even then, work on empirical asset pricing models only provide a direction where further contemporaneous elaborations or searches are needed for model developments on asset returns. Statistical properties are indeed very important for the justification of empirical results. However, the quintessence of empirical asset pricing models in asset returns is to identify the intrinsic mechanism and its role that determines the coherence of these returns in the capital market.

Statistical inference does indeed help. However, statistical verification of empirical asset pricing models should offer the direction or tractability (and plausibility) for the searching of models. In fact, the tractability of model searching should involve optimal usage of available information which provides the common essentiality that may prevail (in all asset returns, for instance) and allow the evolving nature of the models through applications of various data sets.

In particular, statistical analyses must be accompanied with theoretical properties or reasonings developed under economic/financial theories. Statistical inferences and econometrics provide rigorous verification through the extended study of rigorous layouts on the time-series and cross-sectional properties of data steams. However, those works are never exhaustive. To provide some helpful insights on empirical asset pricing models, innovative thoughts that incorporate new theoretical frameworks for analytical issues and explanation are needed. Otherwise, empirical asset pricing models may simply fall into being criticized as measurement without theory as stated in Koopmans (1947).

Although there are also limits for theories (such as that they can't be treated purely as insights that will provide accurate guidance in decision making or anything else) those developments may offer conceivable hypotheses for empirical work that results in confirmation or rebuttal. Whether statistical inferences from empirical data are confirmatory or not, the introduction of economic theory improves our understanding of the alternatives to underlying schemes of interest. Likewise, innovation of thought may result from such refutations. This is discussed in Wolpin (2013) for the risk of inferences without theory, although Rust (2014) mentions that there are also limits to using theory, pronouncing that it is essential to incorporate (economic) theory into test and empirical work to improve the theories instead of purely relying on randomized experiments and asking the data to speak for themselves.

1.1 Theories, Asset-Pricing Models, and Reduction of Dimensionality

Given that there is a vast amount of literature covering the discussions on the asset pricing model, this section will only survey some basic theoretical developments that are well-known in the field. If loosely defined, these models can be denoted as rational pricing models that attempt to provide basic pricing regularity for security returns. These theoretical developments offer further possibilities to extend the asset pricing models for modification for up-dated information. Even though theories don't necessarily predict the stock returns better than a crystal ball in any short period of time, the rigors of these theoretical works lay down the foundation for extended studies that may accommodate more closely the evolving nature of the capital market.

The aim (for the theoretical foundation), however, is not simply for the search for further extensions to try to cope with the results of empirical findings. It is not surprising that one may discover these empirical findings sometimes even contradict each other when various data, time horizons, test statistics, and/or sampling schemes are applied. Empirical findings (when using statistical inferences or otherwise) are only tools or devices for attempting to identify possible features or characteristics of security returns, for instance. However, these empirical findings are not necessarily so universal as to describe the ultimate or intrinsic regularities of the systems of interest. They are only indications which may depict certain features (of the system of interest) that require attention.

Hence, if the results (at any stage of empirical asset pricing models) are identified, it is necessary that one reflects on the theoretical foundation for the justification of rationality in modeling. And that reflection is not to forge theories to match the data. Although scientific analyses do require the steps from hypotheses/theoretical developments to refutation with data verification, the modification of theoretical work afterward (whether the refutation confirms or rejects the initial hypotheses) should strive for the provision of improvement in understanding (of the system of interest) and not for alteration of theoretical work simply so as to coordinate it with the empirical results.

Many theoretical models and empirical works have provided various insights in the finance and economics literature over recent decades. In spite of the huge volume of articles and research, the intent (of contribution in asset pricing) seems oftentimes to encourage the adaptation to fashionable or contemporary trends of thought. However, reviewing the past literature indicates that the inspiration of epoch-breaking research doesn't just follow such fashions. These advancements, either empirical or theoretical, rarely keep trace of those trends or fashion with further empirical examplification or alternative data sets. Instead, the contribution is of various perspectives and inspiration that incur different schools of thought. Reviewing the past literature for CAPM (or APT) shows that these contributions are not simply dedications to tradition or technicality. Instead, they involve ingenuity and path-breaking thoughts.

1.1.1 Market Model and Capital Asset Pricing Model (CAPM)

Typical asset pricing models start from discussions of one-factor model such as CAPM. The model requires the parameter "beta" to describe the association between market/systematic risk and the rates of return for stocks/portfolios. Much empirical verification and evidence is shown to identify the linear trade-offs of these two. The conventional CAPM begins with a simplified analysis such that, for all assets (or portfolios), the expected rates of return can be expressed as a simple linear model such that for $i = 1, \ldots, n$,

$$E[R_i] = R_f + \beta_i[E(R_m) - R_f],$$

or, in terms of excess returns,

$$E[r_i] = E[R_i] - R_f = \beta_i E[R_m - R_f],$$

where R_f stands for the risk-free rate and $\beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\sigma_m^2}$ as the systematic risk, R_m stands for the rate of return of the market portfolio. In brief, the asset's risk premium depends on the systematic risk and the market premium $E(R_m) - R_f$. In applying the CAPM onto the stock return data, conventional studies consider the time series regression model as

$$r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}, \quad i = 1, \dots, n, \ t = 1, \dots, T,$$

where $r_{it} = R_{it} - R_{ft}$ is the excess return of asset i at time t, and $r_{mt} = R_{mt} - R_{ft}$. The time series regression will give the estimates for the "betas" of the excess returns of included assets. Accordingly, if the theory holds true, the intercept in the time series regression should be close to zero. Ideally, if the market portfolio is correctly identified then the theory should result in the second-pass regression, such that for $i = 1, \dots, n$,

$$\overline{r}_i = \gamma_o + \gamma_1 \hat{\beta}_i + \eta_i,$$

where $\overline{r}_i$ is the (time-series) average excess return for asset i, and $\hat{\beta}_i$ is the estimate of beta for asset i from the first-pass regression. Under the model CAPM, the coefficient γ_o should be equal to zero, and γ_1 is the coefficient for the market premium.

However, since the market indices may not precisely represent the market portfolio and the cross-sectional dependence, and since heteroskedasticity may make the conclusion of second-pass regression misleading, Fama and MacBeth (1973) has developed the "grouping" portfolios as applying the estimates of "betas" from the time series regressions so as to consider the second-pass regression such that

$$\overline{r}_p = \mu + \gamma \overline{\beta_p} + \upsilon_p,$$

where $\overline{r}_p$ is the average excess return for portfolio p, and $\overline{\beta_p}$ is the average betas of the assets included in the portfolio p. The reason for regrouping the assets' excess returns and betas into portfolios (according to predetermined characteristics) is to reduce the impact from the errors-in-variables problem since the market indices chosen for the market portfolio in time series regressions may contain measurement errors. In particular, in Fama and MacBeth (1973), a scheme of "rolling" estimates for the betas is implemented over different subperiods of the time series data which possibly reduces the impacts from time-varying coefficients in "betas" when time series regressions are applied.

Regardless of the theoretical appeals that the market-wise risk should be compensated by a suitable risk premium for the risky assets, the empirical findings show that a single market risk premium seems insufficient to explain the risk premiums of asset returns, especially when Chen et al. (1986) show that some other economic variables in addition to market index returns may contribute explanations for asset returns, it is conceivable to see that the development of theoretical modeling toward multi-factor extension will follow logically.

1.1.2 Linear Factor Pricing Models and Arbitrage Pricing Theory

Earlier works on a linear factor pricing model can be found in Ross (1976) and Chamberlain and Rothschild (1983). Grinblatt and Titman (1985) extend the approximate factor structure of Chamberlain (1983) to show that it can be transformed into the exact factor structure in Ross (1976) and vice versa. Following from the works of Grinblatt and Titman (1985) and others, Reisman (1988) applies the Hahn-Banach theorem for the well-defined return space with continuous functionals to establish a similar pricing model. All these theoretical results are based on the existence of a well-defined (approximate) factor structure of return processes.

Intuitively, in a one-factor case, it is easy to see that the return processes can be projected to the factor as

$$r_i = E[r_i] + b_i f + e_i,$$

where r_i is the excess return for asset i, e_i is the idiosyncratic risk, and f is the systematic factor, $E[e_i] = E[f] = \mathrm{Cov}[e_i, f] = 0$. Under no arbitrage condition, the expected premium for asset i should be expressed as

$$E[r_i] \approx \gamma\, b_i.$$

Reisman (1992) extends the analysis to consider the “beta” when defined on the reference variable(s), which establishes a similar result. That is, given the beta $\tilde{b}_i$ (with respect to the reference variable g), it is feasible (if the factor structure is correct) to write

$$\tilde{b}_i \approx \mathrm{Cov}(f, g) b_i.$$

Hence, as long as $\mathrm{Cov}(f, g) \neq 0$, it can be shown that $\gamma\, b_i \approx \lambda \tilde{b}_i$, where

$$\lambda = \frac{\gamma}{\mathrm{Cov}(f, g)}.$$

Rewrite the above equations as

$$E[r_i] - \lambda \tilde{b}_i = (E[r_i] - \gamma\, b_i) + \left(\gamma\, b_i - \lambda \tilde{b}_i\right)$$

which is equivalent to stating that $E[r_i] \approx \lambda \tilde{b}_i$. That is, there exists a pricing functional for expected premiums when using the reference variables.

Following Chamberlain and Rothschild (1983), Reisman (1992), the excess returns for all assets are projected onto the set of factors (or reference variables) $\{f_1, \ldots, f_k\}$such that

$$r_i = \alpha_i + \sum_{j=1}^{k} \beta_{ij} f_j + \epsilon_i,$$

where the $\{\epsilon_i\}_{i=1,\ldots,n}$ represent the idiosyncratic risk of the return processes. The factor structure (with k factors) will hold if the eigenvalues of the covariance matrix of $\{\epsilon_i\}_{i=1,\ldots,n}$ are all bounded from above. Given the factor structure (and under the continuity assumption), it is feasible for the expected returns (or risk premiums) to be expressed as

$$E[r_i] \approx \sum_{j=1}^{k} \beta_{ij} \phi(f_j),$$

where $\phi(f_j)$ is the risk premium of factor j, $j = 1, 2, \ldots, k$, where the pricing errors are square-summable even when the number of assets n increases.

From empirical perspectives, the arbitrage pricing theory extends the conventional models (such as CAPM) into multi-factor space to explain the possible risk premiums of assets. The merit of the model is that the possible explanatory variables (such as reference variables in Reisman 1992) can be extended to higher dimensions, particularly when the number of assets grows large. The difficulty, however, is that there is no indication as to what the "true factors" are and that these factors are usually unobservable. Although proxies or reference variables can be applied to express the risk premiums, there is no justification (or selection rule) as to which set of proxies or reference variables can be considered as optimal. Besides, given that the "true factors" are not observable, little is known of the correlations or associations of the "true factors" and proxies (or reference variables). In fact, as indicated in Lewellen et al. (2010), the cross-sectional regressions in using the proxies for factors may overstate the fittedness of models even when the factors and proxies are only slightly related.

Another question concerns how many factors are needed for these asset returns. Although statistical inferences such as factor analysis seems feasible for identifying the factors, it is usually limited to finite (cross-sectional) dimensions. Expansions in both the cross-sectional and time-series dimensions will require additional analyses (such as Bai and Ng (2002), Bai (2003), and many others) where conventional factor analysis cannot apply. An additional difficulty is that the factors and factor loadings are not identifiable. Any nonsingular linear transform on factors and their loadings will suffice as the same model for describing the return processes. Likewise, empirical applications on these factor-oriented models require extra caution when interpreting estimated results.

1.2 Predictability or Tractability?

Stock return predictability is almost always an issue that attracts much research, using different schemes, data, time periods, and methods. Although there are various studies that cover all possible issues in this field, controversies are bountiful. Different empirical results and claims are covered in much financial literature. The main question still concerns what and how the information of asset returns can be considered when derived from this kind of evidence. For simplicity, the discussions on the forecastability and predictability are used interchangeably. The difference between these two concepts is outlined in Clements and Hendry (1999), namely that forecasts may require additional information on data generating mechanisms and processes while predictability is related to feasible (and possibly lagged) information sets.

When a newer asset pricing model or device is developed, it is usually (and pervasively) assumed that the model will most likely improve the forecastability of asset returns in empirical applications. In addition, the new finding or concept may easily turn into fashionable research when pronounced in the finance literature or elsewhere. More explicitly, various research articles published attempt to find any possible mechanism (including time series models and economic/business attributes such as dividend yields, inflation rates, or earning forecasts) to assess stock return predictability. These earlier studies may have conveyed the notion that predictability is an essential tool for checking the validity of model specification, particularly on asset returns or empirical asset pricing models. However, the proving or disproving of stock return predictability should be to assist the understanding of the underlying systems and to provide better guidance in formatting or modeling the data. It's not merely a demonstration of empirical findings. There are a few issues to be discussed here:

1. Is predictability a necessary result of a good asset pricing model?
2. Is the empirical evidence of finite sample predictability (from the models or other variables) really a reliable source for verification so that stock returns can be traced down accordingly? Are these findings only the confirmations of scholarly trends, or simply the echoes of common knowledge?
3. What role should theoretical asset pricing models play?

For the first issue, according to Clements and Hendry (1999), the unpredictability (for a stochastic process ν_t) is defined as

$$D_{\nu_t}(\nu_t|I_{t-1}) = D_{\nu_t}(\nu_t),$$

where $D_{\nu_t}(\nu_t|I_{t-1})$, and $D_{\nu_t}(\nu_t)$ stand for the conditional and unconditional distribution of ν_t, respectively. That is to say, the (un)predictability is defined upon the available information set I_{t-1}. Choices over different information may cause the empirical results to differ from each other.

Hence, based on the definition, when $\mathscr{I}_{t-1} \subset I_{t-1}$, it is possible to have $D_{\nu_t}(\nu_t|\mathscr{I}_{t-1}) = D_{\nu_t}(\nu_t)$, while $D_{\nu_t}(\nu_t|I_{t-1}) \neq D_{\nu_t}(\nu_t)$—even though the correct information set I_{t-1} may not be known or feasible. In other words, (un)predictability depends on the content of the relevant information sets applied.

Unless justification can be shown on the inclusion relation of the various content of (possibly overlapping) information sets in empirical practices, proofs or disproofs of the predictability of asset returns or asset pricing models (in using different information sets) fall short of being insights into understanding the pricing kernel of asset returns. One of the possible reasons is that these verifications are usually based on different information sets over time horizons, data sources, and constructions of relevant variables. In addition, the notion of (un)predictability can also be considered as time-varying such that for time index $H \geq 1$,

$$D_{\nu_t}(\nu_t|I_{t-1}) = D_{\nu_t}(\nu_t), \quad t = 1, \ldots, T,$$
$$D_{\nu_t}(\nu_t|I_{t-1}) \neq D_{\nu_t}(\nu_t), \quad t = T+1, \ldots, T+H,$$

or,

$$D_{\nu_t}(\nu_t|I_{t-1}) \neq D_{\nu_t}(\nu_t), \quad t = 1, \ldots, T,$$
$$D_{\nu_t}(\nu_t|I_{t-1}) = D(\nu_t), \quad t = T+1, \ldots, T+H.$$

More explicitly, empirical results for predictability of asset returns over different time horizons may vary. Therefore, empirical verifications of (un)predictability for asset returns or asset pricing models across different time horizons may not be identical. As a matter of fact, if the underlying data generating mechanism of asset returns is time-varying, different empirical results of (un)predictability are likely to appear. Likewise, developments in assessing this time-varying nature of asset returns versus asset pricing models need to consider both the detection of the possible time frames for the parameters of interest and the tracking methods (or modelings) for these parameters, too.

For the second issue on empirical verification of asset pricing models, the so-called predictability in many empirical studies is actually discovered from the already known observations collected, in contrast to pure simulations. That is, the empirical results (of predictability) are usually formed from the statistics of so-called in-sample and out-of-sample observations of historical data. Notwithstanding that dividing the historical data into in-sample and out-of-sample observations is subjective and arbitrary, it is rarely feasible that one would simply apply the proposed methods or hypothesized models on the real-time observations from the day-to-day transactions and wait for the results to show. In other words, the entire verification of predictability when using historical data is more likely "in-sample" by the researchers' selection through prior knowledge or otherwise. Random simulations may provide more robust results. However, different schemes of replication may also affect the results of verification. More explicitly, even with high-frequency trading and technology where real-time observations are available in a short period of time, this so-called predictability may vary and/or vanish over different time periods.

Various methods have been designated to the search, including the high-end technology of financial econometrics. However, since any investor will almost always endeavor to pursue speculative profits once the technology is discovered, it is inconceivable that the findings will be published and become enunciable. Hence, it is perhaps not too surprising to say that if one successfully develops a predictive mechanism that belittles others in the market, one should keep it completely confidential.

In particular, under the pavilion of contemporaneous financial econometrics, predictability is essential in either the model verifications or the empirical applications. Unfortunately, even though predictability is only the verification of the model's validity, the provision of some empirical studies may lead to misunderstanding and the attempt to search for the possibility of speculative profits with better forecasts. More strikingly, overemphasis on the predictions and forecasts may lead the financial modeling toward tracking asset returns with devices or mechanisms of short-term validity where no plausible explanation (in financial economics or otherwise) may be feasible. In fact, this kind of emphasis and motivation (for new tracking methods) may simply destroy the validity of verifying model specification (with predictability) since the devised mechanism is only for "tracing and chasing the prey". More specifically, it is not merely due to the possibility of a time-varying pricing mechanism that the theoretical models may not perform better in predictions or forecasts. As a matter of fact, this lack of soundness in predictability simply shows that all models are only approximations for the data generating mechanism. Namely, theoretical models are not developed solely for predictability. Developing models or hypotheses in asset pricing models is to improve the search for understanding and approaches toward some better direction for decision making, if not more.

Thirdly, notice that the ultimate objective of asset pricing models is to identify effectively a tractable explanation for the pricing mechanism (and perhaps on its changing nature) based on the accessible information. Emphasis should be on the continuous effort and work (of finance professionals or academics) of searching for the determinants of the pricing mechanism of stock returns with rigor and explanation—provided that the underlying system of asset returns is almost always evolving through different time periods, regimes, systems, or economies. The goal, however, is to search for any tractable mechanism from which important economic/business decisions can be made. And that (out-of-sample) predictability of the model is only one dimension of the possible statistical properties that may validate the plausibility of the proposed or devised mechanism.

Unfortunately, given the empirical evidence that shows the time series predictability in various aspects, it seems that the fashion leads to finding some plausible alternative theories or else to accommodate the outcomes from empirical studies. In other words, the model search becomes finding theoretical explanations for whatever the empirical statistics may show. This approach could be misleading and counter-productive since one may be attempting to find some plausible excuses for the empirical findings when the past theories (or models) do not hold up completely. The focus for empirical asset pricing models should be on finding the guidelines and theoretical reasonings that may assist the search for the governing structure of asset returns. One should not rely on the empirical findings from various data series, and then develop some plausible ways to accommodate the systems observed empirically.

Although it is not legitimate to take whichever side as favoring predictability or against predictability, it is necessary to stress that the main purpose for empirical asset pricing models is not merely to achieve predictability through any meticulous methodology devised. Discovering the evidence of stock return predictability from various sources is useful indeed. Justification of (say) the profitability of certain devised strategies in considering the long-term predictability of security returns may help us to understand the essence of investments. Yet, verification of these findings can only provide some snapshots within the evolving path of financial markets and the functioning of their participants. How to include these time-changing qualities and to run the empirical asset pricing models to acquire increased plausibility on security returns is far more important.

In short, the provision of a rigorous and/or theoretical analytical apparatus for explaining or forecasting security returns is not to provide any forecast-dominating model or scheme. (In essence, there may be no such ubiquitous model for security returns.) Rather the provision is to offer some guidelines and expectations for pursuing the search for descriptions and rationales that may provide traits in understanding data generating mechanisms. Shmueli (2010) for instance discusses this issue in greater detail.

Explicitly Shmueli (2010) states that statistical models can be used in several essential categories: explanatory modeling, predictive modeling, and descriptive modeling. In essence, the predictability and explanatory power of a model can be considered as two-dimensional requirements for any scientific approach to model building.

As stated in Shmueli (2010), "...the predictive qualities of a model should be reported alongside its explanatory power so that it can be fairly evaluated in terms of its capabilities and compared to the other models. Similarly, a predictive model might not require causal explanation in order to be scientifically useful; however, reporting its relation to causal theory is important for purpose of theory building."

Shmueli (2010) provides a simple example that shows that when a misspecified model is chosen and compared to the so-called true model of the data source, it can be seen that, although the misspecified model suffers from a larger bias in estimating coefficients, the predictive errors of the misspecified model can be less than that of the correct model when some conditions are provided. In that way, if one is emphasizing only predictability, it is easy to choose the misspecified model instead of the true one.

For instance, if a correctly specified statistical model is given as

$$\underline{y} = x_1\beta_1 + x_2\beta_2 + \underline{\varepsilon}, \tag{1.2.1}$$

where unbiased estimation for the model is provided on the correctly specified model such that

$$\mathrm{Var}(\hat{y}) = \sigma^2 \underline{x}'(X'X)^{-1}\underline{x}, \tag{1.2.2}$$

and $\underline{x} = (\underline{x}_1', \underline{x}_2')'$ is the vector of $x_1,\ x_2$, and X is the design matrix of $\underline{x}$. The prediction error can be expressed as

$$\begin{aligned} &E(\underline{y} - \underline{\hat{y}})'(\underline{y} - \underline{\hat{y}}) \\ &\quad = \sigma^2(1 + \underline{x}'(X'X)^{-1}\underline{x}). \end{aligned} \tag{1.2.3}$$

For an incorrectly specified model such as

$$\underline{y} = x_1\gamma_1 + \underline{\nu}, \tag{1.2.4}$$