Nonlinear Properties of Multifactor Financial Models

Kim Heng Chen*, Venkata K. Jandhyala†, & Stergios B. Fotopoulos‡

This paper provides a comprehensive analysis of the nonlinear properties of multifactor pricing models. Beginning with the generalized geometric Brownian motion, we develop a method whereby the log returns of a set of d-assets or portfolios admit a scale mixture model. This is followed by an analytical study on the conditional behavior of a subset of assets given another subset. Expressions for the first two conditional moments are provided under the scale mixture family. The regression equation when the scaling variable is constant (unity) corresponds with the renowned APT. Computable conditional moment expressions for the scaling variable are derived under both inverse gamma and gamma distributions. These moment equations are nonlinear in parameters, apart from containing the usual linear terms under the APT. We then apply the above nonlinear methodology to the log asset returns of four major companies in the U.S. stock market.

JEL classification: C31, C44, G12

Keywords: Generalized Geometric Brownian Motion; Heteroskedasticity; Scale Mixture of Normal Distributions.

INTRODUCTION

One of the major developments in modern financial economics is the Capital Asset Pricing Model (CAPM), which enabled economists to quantify the tradeoff between risk and expected return associated with holding a particular asset. Based on the foundation laid by Markowitz (1959), Sharpe (1964) and Lintner (1965) develop economy-wide implications of the CAPM. They postulate that investors of homogeneous expectations will hold mean-variance efficient portfolios and in the absence of market frictions, the market portfolio of its own accord is a mean-variance efficient portfolio. However, empirical evidence has indicated that the CAPM beta does not completely explain the cross section of expected asset returns: an interesting review of such literature can be found in Fama (1991).

* Assistant Professor, School of Business and Management, American University of Sharjah, P.O. Box 26666, Sharjah, United Arab Emirates.
* Corresponding Author. Tel: (971) - 6-5152365; fax: (971) - 6-5585065; e-mail: kachen@aus.edu
† Professor, Department of Mathematics and Department of Statistics, Washington State University, Pullman, WA99164-3113.
‡ Professor, Department of Management & Operations and Department of Statistics, Washington State University, Pullman, WA99164-4736.
As such, the Arbitrage Pricing Theory (APT) was introduced by Ross (1976) as an alternative to the CAPM and the former has since developed into a powerful tool for modeling the return-generating process of a population of assets or portfolios. Numerous studies on APT have emerged: Roll and Ross (1980), Huberman (1982), Jobson (1982), Chamberlain (1983), Chamberlain and Rothschild (1983), Dhrymes et al. (1984), Ingersoll (1984), Connor and Korajczyk (1988), Milne (1988), Bansal and Viswanathan (1993), Ferson and Korajczyk (1995) \textit{inter alia}. The APT begins with the assumption that the return-generating process can be governed by a linear relationship with a number of systematic factors. Under this and other assumptions, it establishes that the excess return for any arbitrary asset or portfolio may be captured by a linear relationship with the excess returns of a few portfolios with systematic risks. For empirical verification (estimation and testing) purposes, the APT is generally treated as a multifactor model, relating conditional expected returns of assets with diversified portfolios.

Jobson (1982) presents this conditional relationship between expected asset returns and expected portfolio returns in a multivariate regression setup. Relying on the homoskedasticity and normality assumptions, he develops the likelihood ratio test for APT. Financial analysts rather routinely rely upon the normality assumption in this framework for parameter estimation and the associated finite sample inferential properties. Otherwise, inferences are usually restricted to an asymptotic setting, where again, the normal distribution of the underlying parameter estimates is utilized via the central limit theory. However, such dependence upon normality-based theory may not be practical when high-frequency financial data is used in the data analyses. For example, the normality assumption appears to be adequate for characterizing aggregated returns such as monthly or yearly asset returns. But it does not work well with high-frequency financial data such as intraday asset returns. Empirical evidence has shown that these high-frequency financial data exhibit a leptokurtic distribution, that is, the distribution of the high-frequency financial data has “heavier” tails and “sharper” peak relative to the normal distribution, see e.g., Bollerslev (1986), Bollerslev et al. (1992), Feinstone (1987), Gerhard and Hautsch (2002). The violation of the normality assumption will result in biased or even erroneous parameter estimation, which in turn, produces inferences that are misleading at best.

In this paper, we take a rather different approach in the analysis of asset returns from that of Roll and Ross (1980) and Jobson (1982). Beginning with the generalized geometric Brownian motion, we develop a model for the cross-sectional log asset returns (natural logarithm of asset returns) where the volatility matrix is assumed to be stochastic. The stochastic model admits a general form of scale mixtures of the multivariate Gaussian family where the latter is a broad family of distributions that encompasses an array of heavy-tailed distributions and may be closely related to the stable family. This enables us to capture sharp departures from normality in log asset returns. With the derived scale mixture model for log asset returns as the basis, we investigate the conditional behavior of a set of log asset returns given another set. Thus, the basic goal overlaps with that of the APT.

The remainder of this paper is organized as follows. The next section reviews the econometric approach to APT by Roll and Ross (1980) and Jobson (1982). The section after that details the development and properties of the generalized multifactor financial model. This is followed by the section on data analysis, where we perform a detailed empirical analysis on the log returns data from four major U.S. companies, namely, Cisco Systems, Inc. (CISCO), Coca-Cola
Company (COKE), Dell Computer Corporation (DELL), and the Microsoft Corporation (MFST), conditional upon the log returns of the S&P500 index. The final section summarizes, concludes, and sketches directions for future research.

**REVIEW OF APT**

Roll and Ross (1980) derive the APT by postulating that the returns of a set of \( d \) assets are governed by a \( K \)-factor model of the form:

\[
R_i = \mu_i + \sum_{j=1}^{K} b_{ij} \delta_j + \epsilon_i, \quad i = 1, \ldots, d, \quad \text{...}(1)
\]

where \( R_i \) and \( \mu_i \) are the random return and the expected return associated with the \( i^{th} \) asset, \( \delta_j \) is the \( j^{th} \) systematic factor risk assumed to be random with mean zero and is common to all assets, \( b_{ij} \) quantifies the sensitivity of the \( i^{th} \) asset to the \( j^{th} \) systematic factor, and finally, \( \epsilon_i \) is the unsystematic risk component or the noise term. They argue that under equilibrium, there exist weights \( \lambda_1, \ldots, \lambda_K \left( K << d \right) \) such that the expected asset returns will satisfy

\[
\mu_i - \mu_0 = \sum_{j=1}^{K} \lambda_j b_{0j}, \quad i = 1, \ldots, d, \quad \text{...}(2)
\]

where \( \mu_0 \) is the rate of return of a riskfree asset. Upon identifying the factor risk premia \( \lambda_j \), they arrive at the APT of the following form:

\[
\mu_i - \mu_0 = \sum_{j=1}^{K} (\mu^j - \mu_0) b_{0j}, \quad i = 1, \ldots, d, \quad \text{...}(3)
\]

where \( \mu^j - \mu_0 \) denotes the excess returns on portfolios with only \( j^{th} \) systematic factor risk, \( j = 1, \ldots, K \). Thus, the central conclusion of the APT is that the mean premium returns lie in a \( K \)-dimensional subspace spanned by the factor loadings.

Inspired by the work of Roll and Ross (1980), Jobson (1982) formulated the APT into a multivariate regression framework. Specifically, he begins with the time-dependent version of Roll and Ross (1980) \( K \)-factor model by rewriting eq. (1) as

\[
R_t = \mu + B \delta_t + \epsilon_t, \quad t = 1, \ldots, n, \quad \text{...}(4)
\]

where \( \mu \) is the \( d \times 1 \) mean premium return vector, \( B \) is a \( d \times K \) matrix of factor loadings, \( \delta_t \) is a \( K \times 1 \) vector of systematic factors, and \( \epsilon_t \) is an \( d \times 1 \) vector of error terms with \( \delta_t \) being independent of \( \epsilon_t \). Assuming that there exists a subset of \( K \) independent assets or
portfolios with return vector $R_{2t}$, he partitions the model in eq. (4) into the following two equations:

$$R_{1t} = \mu_1 + B_1 \delta_t + \varepsilon_{1t}, \quad \ldots \quad (5)$$

$$R_{2t} = \mu_2 + B_2 \delta_t + \varepsilon_{2t}, \quad \ldots \quad (6)$$

where the partitions of $\mu$, $B$, and $\varepsilon_t$ conform to the partitioning of $R_j$ into $R_{1t}$ and $R_{2t}$. He then derives the APT, namely, the conditional model for $R_{1t}$ given $R_{2t}$, as

$$R_{1t} = \mu^* + B^* R_{2t} + \eta_{1t}, \quad \ldots \quad (7)$$

where $\mu^*$ is $(d - K) \times 1$, $B^*$ is $(d - K) \times K$, and $\eta_{1t}$ is $(d - K) \times 1$. Under the assumption that the premium return vector $R_j$ follows a multivariate normal distribution with mean $\mu$ and covariance $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, he then continues to investigate the inferential properties for the model in eq. (7). In short, Jobson (1982) states that, typically, in regression analysis, one assumes normality to hold or asymptotic theory is employed to justify this unequivocal commitment to normality as well as the linearity associated with the APT in eq. (7).

**ASSET PRICES AND RETURNS MODELING**

As mentioned in Section I, there is growing evidence in the finance literature that asset returns exhibit strong stochasticity in volatility, especially for high-frequency data. To better model the behavior of the asset price dynamics, researchers have developed various methodologies: Hull and White (1987) and Melino and Turnbull (1990) on processes with diffusion stochastic volatility, Rachev and SenGupta (1993) and Cambanis et al. (2000) on mixtures of normal and other distributions, McCulloch (1996) and Willinger et al. (1999) on Stable Paretian models, Madan and Milne (1991) and Barndoff-Nielsen (1998) on subordinated Levy processes in finance.

This leads us to question whether the linearity conditions, commonly imposed on the conditional asset returns in financial modeling continue to hold under strong departures from the deterministic nature of volatility. As shown below, when stochastic volatility is incorporated into financial returns, it will have important implications to various salient features such as linearity, homoskedasticity, as well as the normality of the usual APT-type financial models.

In developing the methodology, we adapt the conditional and scale mixture of normal distribution approaches of Jobson (1982) and Cambanis et al. (2000), respectively. In particular, we derive a cross-sectional model for the log returns of a set of assets or portfolios from the classical generalized Brownian formulation for asset prices. We incorporate stochasticity into the volatility matrix by formulating it as a scale mixture of normal distribution. This establishes
a general structural form for the log asset returns, which subsequently plays a fundamental role in understanding the behavior of conditional asset returns.

Generalized Geometric Brownian Motion and Model for Log Asset Returns

As presented in Karatzas and Shreve (1998), we consider a financial market \( \mathbb{M} \) comprising several \( (d) \) assets, \( d \in \mathbb{N} \). Let \( \{S_i(t), i = 1, \ldots, d, t \in [0, T]\} \) denote the prices of these financial instruments at time \( t \) with \( \{S_i(0), i = 1, \ldots, d\} \) being positive constants. Then, the prices evolve according to the generalized geometric Brownian motion:

\[
dS_i(t) = S_i(t) \left\{ \mu_i(t) dt + \sigma_i^{1/2}(t), dW(t) \right\}, \quad i = 1, \ldots, d, \quad (8)
\]

where \( \langle \cdot, \cdot \rangle \) denote the inner product, \( W = (W^{(1)}, \ldots, W^{(d)})' \) denotes a standard Brownian motion in \( \mathbb{R}^d \) defined on a complete probability space \( (\Omega, \mathcal{F}, \mathbb{P}), \{\mathcal{F}_t, t \in [0, T]\} \) denotes the \( \mathbb{P} \)-augmentation of the filtration \( \mathcal{F}^W_t = \sigma \left( W(s) : 0 \leq s \leq t \right) \) (s-algebra generated by \( W \)), \( \mu_i(t) \) and \( \sigma_i^{1/2}(t) \) denotes the appreciation rate and volatility of \( i^{th} \) asset at time \( t \), respectively. Note that both \( \mu_i(t) \) and \( \sigma_i^{1/2}(t) \), \( i = 1, \ldots, d \), are assumed to be progressively measurable with respect to \( \{\mathcal{F}_t\} \) (i.e., independent of the standard Brownian motion \( W \)) and are bounded uniformly in \( (t, w) \in [0, T] \times \Omega \).

The process of interest in this study is the log asset price process, i.e., \( Y_i = \log S_i(t), \quad i = 1, \ldots, d \). Since \( S_i(t) \) is an Ito process, \( Y_i \) is also an Ito process and is given by

\[
dY_i(t) = \left\{ \mu_i(t) - \frac{1}{2} \sigma_i^{1/2}(t), 1 \right\} dt + \sigma_i^{1/2}(t) dW(t), \quad i = 1, \ldots, d, \quad (9)
\]

or equivalently in vector form

\[
dY(t) = \left\{ \mu(t) - \frac{1}{2} \sigma(t), 1 \right\} dt + \sigma^{1/2}(t) dW(t), \quad \quad (10)
\]

where \( \mathbf{1} = (1, \ldots, 1)' \). By imposing the strong non-degeneracy condition

\[
y' \sigma(t) y \geq \varepsilon \|y\|^2, \quad \forall (t, y) \in [0, T] \times \mathbb{R}^d \text{ a.s. on the volatility matrix } \sigma \quad \text{where } \varepsilon > 0 ,
\]

and
assuming \( \int_0^t \left( \| \mathbf{u}(u) \|^2 + \| \sigma(u) \|^2 \right) du < \infty \ a.s., \) together with the Lipschitz condition

\[
\| \mathbf{u}(u) - \mathbf{u}(s) \|^2 + \| \sigma(u) - \sigma(s) \|^2 \leq c |u - s|, \quad u, s > 0, \quad \ldots (11)
\]

the following may be shown to hold as a solution of the differential equation in eq. (10):

\[
Y(t) = Y_0 + \int_0^t \left( \mathbf{u}(u) - \frac{1}{2} \sigma(u) \mathbf{1} \right) du + \int_0^t \left( \sigma^{1/2}(u), dW(t) \right), \quad t \in [0, T]. \quad \ldots (12)
\]

Next, we partition the time horizon \([0, T]\) into \(0 = t_{0,k} < t_{1,k} < \ldots < t_{j,k} \leq T\) such that

\[
\lim_{k \to \infty} \sup_{j} \left( t_{j,k} - t_{j-1,k} \right) = 0. \quad \text{In practice, the length of time between two successive partitioned periods may represent years, months, or even weeks. The latter can be further sub-partitioned into \(n\) intervals with length \(\delta\), i.e., \([t_{j,k} + (i-1)\delta, t_{j,k} + i\delta], i = 1, \ldots, n; j = 1, \ldots, k, \quad k \geq 1.\)
\]

The log return of the asset for the \(j\)th interval at \(j\)th period can then be expressed as

\[
R(t_{j,k}, \delta, i) = Y(t_{j,k} + i\delta) - Y(t_{j,k} + (i-1)\delta) = \left( \mathbf{u}(t_{j,k} + (i-1)\delta) - \frac{1}{2} \sigma(t_{j,k} + (i-1)\delta) \mathbf{1} \right) \delta \\
+ \left( \sigma^{1/2}(t_{j,k} + (i-1)\delta), W(t_{j,k} + i\delta) - W(t_{j,k} + (i-1)\delta) \right). \quad \ldots (13)
\]

Applying the Lipschitz condition in eq. (11) upon the appreciation rate vector and the volatility matrix, for sufficiently large \(k\), the above return vector may then be approximated as

\[
R(t_{j,k}, \delta, i) \approx \left( \mathbf{u}(t_{j,k}) - \frac{1}{2} \sigma(t_{j,k}) \mathbf{1} \right) \delta + \left( \sigma^{1/2}(t_{j,k}), \sigma^{1/2}Z \right) = R(t_{j,k}, \delta) \quad a.s. \quad \ldots (14)
\]

Eq. (14) suggests that the returns (or log returns) within any \(d\) interval are independent of the position at which the interval is considered. In the modeling literature, researchers usually let \(d\) represents the length of a day, hour, minute, or even second. Rydberg (1999) and Rydberg and Shephard (2001) consider an intraday analysis of the U.S. stock prices by letting \(d\) be in seconds. Note that the stochastic difference equation in eq. (14) can be interpreted as the Euler approximation of the log asset returns. The discretized log asset price process then converges almost surely to its continuous counterpart in eq. (10) as \(\delta \to 0\) and \(k \to \infty\). Refer to Kloeden and Platten (1995) for details.

When the appreciation rate vector \(\mathbf{u}\) and the volatility matrix \(\sigma\) are deterministic, the process \(\{ S_i(t), i = 1, \ldots, d, t \in [0, T]\}\) follows the classical lognormal model. Though elegant
from an analytic point of view, empirical evidence does not support this situation: high-frequency
data indicates that the distributions of asset returns are by and large leptokurtic, i.e., the
distributions of asset returns have relatively heavier tails than that of the Gaussian process. This
suggests that either the appreciation vector or the volatility matrix within the framework of the
geometric Brownian motion is non-deterministic. The modus operandi is to propose stochastic
versions of these parameter processes within the generalized geometric Brownian motion
framework. Retaining the deterministic nature of the appreciation vector \( \mathbf{\mu}(\cdot) \), we incorporate
stochasticity into the volatility matrix by letting \( \mathbf{\sigma}(\cdot) = A\mathbf{\Sigma}(\cdot) \) a.s., where \( A \) is a strictly positive
random variable and \( \mathbf{\Sigma}(\cdot) \) is a deterministic variance-covariance matrix.

The focus of this study is on the returns from a fixed partitioned interval. Thus, we can
drop the time component \( t_{\mu} \) from eq. (14) and without loss of generality, we let the sub-
partitioned size parameter \( d \) be 1. As a result, eq. (14) can be written as

\[
\mathbf{R} = \mathbf{\mu} - \frac{A}{2} \mathbf{\Sigma} + A^{1/2} \mathbf{G},
\]

...(15)

where \( \mathbf{G} \) follows \( d \)-multivariate normal distribution with variance-covariance matrix \( \mathbf{\Sigma} \). Note
that the scale random variable \( A \) is independent of \( \mathbf{G} \). The above model can be expressed as a
special case of the following more general form:

\[
\mathbf{R} = \mathbf{\mu} + A\mathbf{m} + A^{1/2} \mathbf{G},
\]

...(16)

where \( \mathbf{m} \) is a \( d \)-dimensional parameter vector. A special case of the model in eq. (16) is
considered in the monograph by Kotz et al. (2001), where they study some distributional
behavior of \( \mathbf{R} \) when the scaling variable \( A \) follows the exponential distribution with mean one.

In the studies of stock market returns, scale mixture type volatility processes as mentioned
above have been considered by several authors to capture the excess kurtosis. For instance,
Pratetz (1972), Blattberg and Goneses (1974), Madan and Seneta (1990), Kuchler et al. (1995),
Hurst and Platen (1997), and Barndorff-Nielsen (1997, 1998) inter alia have proposed gamma,
inverse gamma, lognormal, and inverse Gaussian distributions for the mixing variable while
modeling the volatility process of a single asset by scale mixtures.

**Conditional Properties of Log Asset Returns**

Conditional distributions of log asset returns and their conditional moments play important
roles in financial modeling and forecasting of returns. The CAPM and APT amply illustrate the
importance of such conditional modeling. While these are well studied for the case where the
volatility process is in equilibrium, such models and their properties are not fully explored and
well established for the case of stochastic volatility, such as those represented by scale mixtures
as shown in eq. (16). As such, recent works by Cambanis and Fotopoulos (1995), Fotopoulos
We begin by partitioning $R$, $\mu$, $m$, $G$, and $\Sigma$ into $R' = (R'_1, R'_2)$, $\mu' = (\mu'_1, \mu'_2)$, $m' = (m'_1, m'_2)$, $G' = (G'_1, G'_2)$, and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ where $R'_2$, $\mu'_2$, $m'_2$, and $G'_2$ are $K$-dimensional and $\Sigma_{22}$ is a $K \times K$ variance-covariance matrix, $K < d$, etc. Utilizing the methodology in Fang and Zhang (1990), i.e., by defining $\Sigma = B'B$ and $\Sigma_{22} = B'_2B_2$, where $B_2$ is a positive definite matrix, it follows that

$$\Sigma = B'B = \begin{bmatrix} I & \Sigma_{12}\Sigma_{22}^{-1} \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ \Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix} = F'F,$$

...(17)

where $F = \begin{bmatrix} B_{1,1,2} & 0 \\ 0 & B_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ \Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix} = \begin{bmatrix} B_{1,1,2} & 0 \\ B_2\Sigma_{22}^{-1}\Sigma_{21} & B_2 \end{bmatrix}$. Note that $\Sigma_{11,2} = B'_{1,1,2}B_{1,1,2}$ and with these results, eq. (16) may be equivalently expressed as

$$R = \mu + A'm + A'^{1/2}B'Z^d = \mu + A'm + A'^{1/2}F'Z,$$

which can further be expressed as

$$R = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + A \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} + A'^{1/2} \begin{pmatrix} B'_{1,1,2}Z_1 + \Sigma_{12}\Sigma_{22}^{-1}B'_2Z_2 \\ B'_2Z_2 \end{pmatrix},$$

...(18)

where $Z$ is a $d$-dimensional standard normal vector.

The above representation of the log asset returns turns out to be well suited for in-depth study of the conditional properties of $(R_1 | R_2)$. This class of scale mixture type random vectors, which encompasses the Gaussian systems as distinguished members, exhibits conditional behavior that is distinct from the normal theory. For example, regression involving jointly Gaussian random variables is always linear. However, this is not the case for the current model. Moreover, the conditional covariance of jointly normal random variables leads to degenerate, non-random quantities. Again, this is not the case for the model in eq. (16). Cambanis et al. (1981), Hardin (1982), Cambanis and Wu (1992), Samorodnitsky and Taqqu (1994), inter alia, consider situations in which the conditional regression is nonlinear and the conditional variance-covariance matrix is non-degenerate.

The theorem below provides expressions for the regression and skewastic (conditional variance of $R_1 | R_2$ a.s.) functions of $R_2$ along with necessary and sufficient conditions for their
existence. The proof can be obtained by simple conditioning arguments on eq. (16) and is thus omitted.

THEOREM 1. I. The regression equation of $R_2$ (conditional mean of $R_1$ given $R_2$) is given as

$$E\left[ R_1 | R_2 \right] = \mu_1 + m_{11,2} E\left[ A | R_2 \right] + \Sigma_{12,22}^{-1} (R_2 - \mu_2), \ a.s.$$

if and only if $E\left[ A | R_2 \right] < \infty \ a.s.$, where $m_{11,2} = \mu_1 - \Sigma_{12,22}^{-1} \Sigma_{11,2} \mu_2$.

II. The skedastic equation of $R_2$ (conditional variance of $R_1$ given $R_2$) is given as

$$V\left[ R_1 | R_2 \right] = \Sigma_{11,2} E\left[ A^2 | R_2 \right] + m_{11,2} m'_{11,2} V\left[ A | R_2 \right], \ a.s.$$

if and only if $E\left[ A^2 | R_2 \right] < \infty \ a.s.$, where $\Sigma_{11,2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$.

By setting $m_1 = -\frac{1}{2} (\Sigma_{11} 1_1 + \Sigma_{12} 1_2)$ and $m_2 = -\frac{1}{2} (\Sigma_{21} 1_1 + \Sigma_{22} 1_2)$ leads one to the model in eq. (15). Consequently, the corresponding regression and skedastic functions for this special case can be expressed as

$$E\left[ R_1 | R_2 \right] = \mu_1 - \frac{1}{2} \Sigma_{11,2} 1_1 E\left[ A | R_2 \right] + \Sigma_{12,22}^{-1} (R_2 - \mu_2), \ a.s. \quad \text{(19)}$$

$$V\left[ R_1 | R_2 \right] = \Sigma_{11,2} E\left[ A^2 | R_2 \right] + \frac{1}{2} \Sigma_{11,2} 1_1' \Sigma_{11,2} V\left[ A | R_2 \right] a.s. \quad \text{(20)}$$

Note that the regression equation in eq. (19) contains the conditional expectation of the mixing variable, $E\left[ A | R_2 \right]$, in addition to the usual linear term in $R_2$. We will show in the next subsection that for some specific distributions of $A$, $E\left[ A | R_2 \right]$ is nonlinear in both $R_2$ and the parameters, thus making the regression intrinsically nonlinear. Similarly, the expression for the skedastic function in eq. (20) also contains $E\left[ A | R_2 \right]$ besides the conditional variance term $V\left[ A | R_2 \right]$.

When the scaling variable $A$ is constant, i.e., $A = 1$, the regression equation in eq. (19) coincides with the expectation part of the APT as in eq. (7). Furthermore, the conditional variance-covariance matrix is degenerate.

When the scaling variable $A$ is random and $m = 0$, it follows from Theorem 1 that the regression equation continues to be linear. However, the corresponding variance-covariance
matrix is non-degenerate, thus, leading to heteroskedastic situations. Cambanis et al. (2000) characterize various nonlinear properties of $E[A|R_2]$ when $m = 0$ and $\mu = 0$. These properties enabled them to explain heteroskedasticity in the presence of the mixing random variable $A$. In particular, they consider the case where $A \sim S_{\alpha/2} \left( \cos \left( \frac{\pi}{4} \right), 1, 0 \right), 0 < \alpha < 2$, i.e., $A$ may have a stable distribution totally skewed to the right. The concentration of mass in the tail areas of the stable family of distributions has made it a suitable candidate for modeling financial data. They have noted that the conditional expectation of the mixing variable, $E[A|R_2]$, is finite even though its unconditional counterpart, $E[A]$, may be infinite.

When the scaling variable $A$ is random and $m \neq 0$, in addition to the linear component, the regression equation necessarily contains the nonlinear factor associated with the scaling random variable. Similarly, the variance-covariance matrix is non-degenerate, involving both $E[A|R_2]$ and $V[A|R_2]$.

Undoubtedly, the model in eq. (16) with $A$ being random and $m \neq 0$ captures the behavior of return premiums for the situation where the volatility is stochastic. Under this setup, the unsystematic risks are modeled by scale mixtures and the heavy-tailed nature of errors is accommodated. The model generalizes the APT to accommodate stochastic volatility and should be of significant interest to financial analysts. In what follows, we study the conditional properties of the model in eq. (16) for some special distributions of $A$.

Log Asset Returns Under Inverse Gamma and Gamma Mixture of Normal Distributions

The expressions for both the regression and skedastic functions in eqs. (19) and (20), respectively, involve the two conditioning moments, $E[A|R_2]$ and $V[A|R_2]$. Computation of these moments requires closed-form expressions for certain distributions of $A$. In this subsection, we shall first provide explicit analytical expressions for these conditional moments for cases where the mixing variable $A$ follows an inverse gamma distribution with $p$ degrees of freedom, i.e., $A \sim p/\chi_p^2$, and where the mixing variable $A$ follows a gamma distribution with shape parameters $p$ and 1, respectively, i.e., $A \sim \Gamma(p, 1)$.

Letting $A \sim p/\chi_p^2$ implies that the vector of log returns $R$ in eq. (16) follows a non-central multivariate $t$-type distribution. This is a natural non-Gaussian family that allows the modeling of both heavy-tailed (for small $p$) as well as lighted-tailed distributions (for moderate $p$). For instance, Hurst and Platten (1997) remark that the inverse gamma distribution for the mixing variable fits well for modeling log asset returns. On the other hand, if one lets $A \sim \Gamma(p, 1)$, then the vector of log returns $R$ in eq. (16) follows a hyperbolic-type multivariate distribution. Madan and Seneta (1990) obtain some distributional properties of the log asset returns over time under gamma scale mixtures.
Nonlinear Properties of Multifactor Financial Models

The following theorem provides closed-form expressions for the conditional moments of the scaling variable \( A \) when the latter follows either the inverse gamma or gamma distributions.

**Theorem 2.1.** Suppose that \( A \sim p \left( \chi^2_p \right) \) with probability density function given as

\[
f_A(a) = \frac{p^{p/2}}{2^{p/2} \Gamma(p/2)} a^{p-1} e^{-p/2a}, \quad a > 0.
\]

Then,

\[
E \left[ A^j \mid \mathbf{R}_2 = \mathbf{r}_2 \right] = \left( \frac{x}{y} \right)^j \frac{K_{(p+k-2j)/2}(xy)}{K_{(p+k)/2}(xy)}, \quad \mathbf{r}_2 \in \mathbb{R}^k; j \geq 1,
\]

where \( x^2 = p + \left( \mathbf{r}_2 - \mu_2 \right)^2 \), \( y^2 = \left\| \mathbf{m}_2 \right\|^2 \), \( K_\alpha(.) \) is the modified Bessel function of the third kind and the norm \( \left\| \mathbf{r} \right\|^2 = \left\langle \Sigma^{-1} \mathbf{r}, \mathbf{r} \right\rangle \).

**Proof:** See Appendix.

II. Suppose that \( A \sim \Gamma(p, 1) \) with probability density function given as

\[
f_A(a) = \frac{1}{\Gamma(p)} a^{p-1} e^{-a}, \quad a > 0.
\]

Then,

\[
E \left[ A^j \mid \mathbf{R}_2 = \mathbf{r}_2 \right] = \left( \frac{u}{v} \right)^j \frac{K_{p-j-1/2}(uv)}{K_{p-1/2}(uv)}, \quad \mathbf{r}_2 \in \mathbb{R}^k; j \geq 1,
\]

where \( u^2 = \left\| \mathbf{r}_2 - \mu_2 \right\|^2 \), \( v^2 = 2 + \left\| \mathbf{m}_2 \right\|^2 \), \( K_\alpha(.) \) is the modified Bessel function of the third kind and the norm \( \left\| \mathbf{r} \right\|^2 = \left\langle \Sigma^{-1} \mathbf{r}, \mathbf{r} \right\rangle \).

**Proof:** See Appendix.

It is well known that the Bessel function \( K_\lambda(z) \) is analytic on the \( z \)-plane except at branch points \( z = 0 \) and \( z = \infty \). Moreover, if \( -\pi < \arg z < \pi \), then \( K_\lambda(z) \) is real and positive whenever \( \lambda \) is real, and \( z \) is real and positive. From the above theorem, it is clear that the index \( \lambda \) is real and that the arguments in both Parts I and II of Theorem 2 are positive. Thus, \( K_\lambda(z) \) is always real and positive in both Parts I and II of Theorem 2.
Using Theorem 2, the two conditional moments for eqs. (19) and (20) may be expressed as

When \( A \sim p / \chi_2^2 \),

\[
E \left[ A \mid \mathbf{R}_2 = \mathbf{r}_2 \right] = \frac{x y}{\alpha_{(p+k)/2}} \left( xy \right), \quad \mathbf{r}_2 \in \mathbb{R}^k, \quad \ldots (21)
\]

\[
V \left[ A \mid \mathbf{R}_2 = \mathbf{r}_2 \right] = \left( \frac{x y}{\alpha_{(p+k)/2}} \left( xy \right) \alpha_{(p+k)/2} (xy) - \alpha_{(p+k)/2} (xy) \right), \quad \mathbf{r}_2 \in \mathbb{R}^k. \quad \ldots (22)
\]

When \( A \sim \Gamma (p,1) \),

\[
E \left[ A \mid \mathbf{R}_2 = \mathbf{r}_2 \right] = \frac{\mu}{V} \alpha_{p-k} (uv), \quad \mathbf{r}_2 \in \mathbb{R}^k, \quad \ldots (23)
\]

\[
V \left[ A \mid \mathbf{R}_2 = \mathbf{r}_2 \right] = \left( \frac{\mu}{V} \alpha_{p-k} (uv) \alpha_{p-k} (uv) - \alpha_{p-k} (uv) \right), \quad \mathbf{r}_2 \in \mathbb{R}^k, \quad \ldots (24)
\]

where \( \alpha_{\lambda} (z) = K_{\lambda-1} (z) / K_{\lambda} (z) \), \( z > 0, \lambda \in \mathbb{R} \).

DATA ANALYSIS

Data Description

The data used in this study: daily and intraday (15-minute) stock prices of the Cisco Systems, Inc. (CISCO), Coca-Cola Company (COKE), Dell Computer Corporation (DELL), Microsoft Corporation (MSFT), and S&P500 index (S&P500) for the period spanning from January 1998 to December 2000 are obtained from AnalyzerXL LLC. These stocks are chosen mainly because of their liquidity, particularly the technology stocks which were very popular in the late 90’s. For example, the 15-minute trading volume of the CISCO in 1998 averaged to about 2 million and reached the year high of 18 million. As for the S&P500, it is one of the commonly-used proxies for the market portfolio.

The log asset returns are then computed by taking the natural logarithms of the price ratios. Table 1 reports the summary statistics of the daily and intraday log asset returns. The large and positive (unconditional) excess kurtosis values of the intraday log asset returns indicate that their distributions have “heavier” tails and “sharper” peaks relative to that of the normal distribution; this concurs with the previous empirical findings.

Parameter Estimation

For the purpose of conciseness, we perform the comparative study of the scale mixture
<table>
<thead>
<tr>
<th></th>
<th>Daily (252 obs.)</th>
<th>Intraday (6530 obs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CISCO</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.00406</td>
<td>0.00367</td>
</tr>
<tr>
<td></td>
<td>-0.00044</td>
<td>0.00014</td>
</tr>
<tr>
<td></td>
<td>-0.00013</td>
<td>-0.00006</td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td>0.02886</td>
<td>0.02640</td>
</tr>
<tr>
<td></td>
<td>0.04267</td>
<td>0.00560</td>
</tr>
<tr>
<td></td>
<td>0.00542</td>
<td>0.00806</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-0.59128</td>
<td>0.05456</td>
</tr>
<tr>
<td></td>
<td>0.51734</td>
<td>-0.51730</td>
</tr>
<tr>
<td></td>
<td>0.65486</td>
<td>-0.14116</td>
</tr>
<tr>
<td><strong>Excess Kurtosis</strong></td>
<td>3.88971</td>
<td>3.1198</td>
</tr>
<tr>
<td></td>
<td>1.17345</td>
<td>27.6850</td>
</tr>
<tr>
<td></td>
<td>12.8634</td>
<td>13.0440</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>-0.13531</td>
<td>-0.06269</td>
</tr>
<tr>
<td></td>
<td>-0.12575</td>
<td>-0.10114</td>
</tr>
<tr>
<td></td>
<td>-0.06573</td>
<td>-0.11219</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>0.09924</td>
<td>0.07128</td>
</tr>
<tr>
<td></td>
<td>0.16667</td>
<td>0.05584</td>
</tr>
<tr>
<td></td>
<td>0.05604</td>
<td>0.06454</td>
</tr>
<tr>
<td><strong>COKE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.00024</td>
<td>-0.00032</td>
</tr>
<tr>
<td></td>
<td>0.00059</td>
<td>0.00000</td>
</tr>
<tr>
<td></td>
<td>-0.00002</td>
<td>0.00001</td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td>0.01948</td>
<td>0.01966</td>
</tr>
<tr>
<td></td>
<td>0.02692</td>
<td>0.00392</td>
</tr>
<tr>
<td></td>
<td>0.00393</td>
<td>0.00481</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-0.56106</td>
<td>0.51137</td>
</tr>
<tr>
<td></td>
<td>0.27313</td>
<td>-0.42679</td>
</tr>
<tr>
<td></td>
<td>-1.41467</td>
<td>0.41457</td>
</tr>
<tr>
<td><strong>Excess Kurtosis</strong></td>
<td>3.41665</td>
<td>2.19799</td>
</tr>
<tr>
<td></td>
<td>0.89365</td>
<td>22.9822</td>
</tr>
<tr>
<td></td>
<td>45.0368</td>
<td>13.2502</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>-0.10524</td>
<td>-0.06404</td>
</tr>
<tr>
<td></td>
<td>-0.09069</td>
<td>-0.05439</td>
</tr>
<tr>
<td></td>
<td>-0.08744</td>
<td>-0.04461</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>0.05088</td>
<td>0.08344</td>
</tr>
<tr>
<td></td>
<td>0.09654</td>
<td>0.05022</td>
</tr>
<tr>
<td></td>
<td>0.04058</td>
<td>0.06027</td>
</tr>
<tr>
<td><strong>DELL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.00559</td>
<td>0.00191</td>
</tr>
<tr>
<td></td>
<td>-0.00329</td>
<td>0.00019</td>
</tr>
<tr>
<td></td>
<td>0.00005</td>
<td>-0.00017</td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td>0.03521</td>
<td>0.03443</td>
</tr>
<tr>
<td></td>
<td>0.04367</td>
<td>0.00687</td>
</tr>
<tr>
<td></td>
<td>0.00682</td>
<td>0.00835</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-0.28358</td>
<td>-0.03126</td>
</tr>
<tr>
<td></td>
<td>0.01918</td>
<td>0.24399</td>
</tr>
<tr>
<td></td>
<td>-1.10832</td>
<td>-2.21543</td>
</tr>
<tr>
<td><strong>Excess Kurtosis</strong></td>
<td>1.53678</td>
<td>0.11266</td>
</tr>
<tr>
<td></td>
<td>1.98852</td>
<td>26.8646</td>
</tr>
<tr>
<td></td>
<td>31.0459</td>
<td>53.5474</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>-0.15790</td>
<td>-0.11779</td>
</tr>
<tr>
<td></td>
<td>-0.18943</td>
<td>-0.09580</td>
</tr>
<tr>
<td></td>
<td>-0.10380</td>
<td>-0.18056</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>0.10071</td>
<td>0.08943</td>
</tr>
<tr>
<td></td>
<td>0.17790</td>
<td>0.09422</td>
</tr>
<tr>
<td></td>
<td>0.07735</td>
<td>0.10606</td>
</tr>
<tr>
<td><strong>MSFT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.00332</td>
<td>0.00235</td>
</tr>
<tr>
<td></td>
<td>-0.00329</td>
<td>0.00012</td>
</tr>
<tr>
<td></td>
<td>0.00008</td>
<td>-0.00015</td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td>0.02383</td>
<td>0.02398</td>
</tr>
<tr>
<td></td>
<td>0.03544</td>
<td>0.00437</td>
</tr>
<tr>
<td></td>
<td>0.00462</td>
<td>0.00668</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-0.08075</td>
<td>0.40155</td>
</tr>
<tr>
<td></td>
<td>-0.00784</td>
<td>0.35719</td>
</tr>
<tr>
<td></td>
<td>0.57973</td>
<td>-2.20855</td>
</tr>
<tr>
<td><strong>Excess Kurtosis</strong></td>
<td>0.34590</td>
<td>0.69517</td>
</tr>
<tr>
<td></td>
<td>5.72548</td>
<td>18.1038</td>
</tr>
<tr>
<td></td>
<td>18.1379</td>
<td>101.290</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>-0.08848</td>
<td>-0.06494</td>
</tr>
<tr>
<td></td>
<td>-0.15598</td>
<td>-0.03870</td>
</tr>
<tr>
<td></td>
<td>-0.05760</td>
<td>-0.16476</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>0.06237</td>
<td>0.09880</td>
</tr>
<tr>
<td></td>
<td>0.19565</td>
<td>0.06614</td>
</tr>
<tr>
<td></td>
<td>0.06454</td>
<td>0.12474</td>
</tr>
<tr>
<td><strong>S&amp;P 500 Index</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>0.00109</td>
<td>0.00082</td>
</tr>
<tr>
<td></td>
<td>-0.00027</td>
<td>0.00003</td>
</tr>
<tr>
<td></td>
<td>0.00003</td>
<td>-0.00001</td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td>0.01277</td>
<td>0.01136</td>
</tr>
<tr>
<td></td>
<td>0.01416</td>
<td>0.00282</td>
</tr>
<tr>
<td></td>
<td>0.00239</td>
<td>0.00277</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>-0.52335</td>
<td>0.08816</td>
</tr>
<tr>
<td></td>
<td>0.07537</td>
<td>-0.28310</td>
</tr>
<tr>
<td></td>
<td>0.21440</td>
<td>-0.33209</td>
</tr>
<tr>
<td><strong>Excess Kurtosis</strong></td>
<td>4.51938</td>
<td>0.10190</td>
</tr>
<tr>
<td></td>
<td>1.37649</td>
<td>22.5907</td>
</tr>
<tr>
<td></td>
<td>9.21598</td>
<td>11.5762</td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
<td>-0.06799</td>
<td>-0.02779</td>
</tr>
<tr>
<td></td>
<td>-0.05947</td>
<td>-0.03917</td>
</tr>
<tr>
<td></td>
<td>-0.02449</td>
<td>-0.03592</td>
</tr>
<tr>
<td><strong>Maximum</strong></td>
<td>0.05058</td>
<td>0.03563</td>
</tr>
<tr>
<td></td>
<td>0.04709</td>
<td>0.03582</td>
</tr>
<tr>
<td></td>
<td>0.02051</td>
<td>0.02618</td>
</tr>
</tbody>
</table>
nonlinear model (scaling variable $A$ is random) with the usual linear APT (scaling variable $A$ is constant, i.e., $A = 1$) by conditioning the individual log asset returns of CISCO, COKE, DELL, and MSFT upon S&P500. In addition, we will only report the results when the scaling variable follows the inverse gamma distribution as it has been found that the inverse gamma is more appropriate for both daily and intraday log asset returns analyses.

Given a set of data on the log asset returns, computation of $E[R_1 | R_2]$ and $V[R_1 | R_2]$ in Theorem 1 requires that one is able to compute $E[A | R_2]$ and $V[A | R_2]$. Throughout the analysis, we let $m_1 = \frac{1}{2} (\sigma_{11} + \sigma_{12})$ and $m_2 = \frac{1}{2} (\sigma_{21} + \sigma_{22})$, as derived based on the geometric Brownian motion. Then, the relevant expressions for these conditional moments of the scaling variable are provided in eqs. (21) and (22) for the case where the scaling variable follows inverse gamma distribution. Computations of these expressions involve estimation of the unknown parameters $\sigma_{11}$, $\sigma_{12}$, and $\sigma_{22}$ as well as the computation of the modified Bessel function, $K_{\alpha}(\cdot)$. It may be noted that the mean of log asset returns $\mu' = (\mu_1, \mu_2)'$ may be assumed to be zero under the usual fair market scenario and hence, requires no estimation.

The method of moments is well suited for estimating the unknown covariance matrix $\Sigma$ in this case. The method involves the derivation of theoretical expressions for $\Sigma$ through eq. (18), the first two moments of the variable $A$ via the inverse gamma distribution and the computation of sample means and sample covariance matrix of $R_1$ and $R_2$. While maximum likelihood estimation would also be appropriate, we here choose to implement the method of moments in light of the highly nonlinear nature of the likelihood in the parameters $\Sigma$.

Utilizing eq. (18), it is not hard to obtain the following expressions:

$$\sigma_{11} = \frac{E[A]^2 V[R_1] - V[A]}{E[A]} \left( \mu_1 - E[R_1] \right)^2, \quad \ldots(25)$$

$$\sigma_{22} = \frac{E[A]^2 V[R_2] - V[A]}{E[A]} \left( \mu_2 - E[R_2] \right)^2, \quad \ldots(26)$$

$$\sigma_{12} = \frac{E[A] C[R_1, R_2] - V[A]}{E[A]} \left( \mu_1 - E[R_1] \right) \left( \mu_2 - E[R_2] \right), \quad \ldots(27)$$

where $C[R_1, R_2]$ denotes the covariance of $R_1$ and $R_2$. Furthermore, when $A \sim \rho / \chi_p^2$, the
first two moments are given by $E[A] = p/(p-2)$ and $V[A] = 2p^2/(p-2)^2(p-4)$.

The parameters $\sigma_{11}$, $\sigma_{12}$, and $\sigma_{22}$ may then be estimated by substituting the sample moments for $R_1$ and $R_2$ into eqs. (25)-(27). In solving these equations, one sets $\mathbf{u}' = (0, 0)'$.

The Bessel function $K_1(\cdot)$ may be computed through standard software package. One thus proceeds to compute the conditional moments in eqs. (21) and (22) and then $E[R_1|R_2]$ and $V[R_1|R_2]$ in Theorem 1. The latter for the linear model under normality are computed by simply letting $A = 1$ in the Theorem.

Noting that our model and inference are based on the mixture setup leading to a general form of the APT for scale mixture family of distributions, it is convenient to let our model be MAPT, the mixture-based APT. For each of the log asset returns data considered in our analysis, we then compare the MAPT with the APT for effectiveness in performance with respect to both regression and skedastic functions. The comparison is based upon the residuals of the respective models. Thus, let $\hat{\epsilon}_i$ and $\tilde{\epsilon}_i$ for $i = 1, \ldots, n$, be the respective residuals for MAPT and the APT. It is intuitively appealing to capture the mixture model effects (MME) upon the two conditional moments by computing

$$
MME_E = \frac{SSE(MAPT_E) - SSE(APT_E)}{SSE(APT_E)} \times 100\%, \quad \ldots(28)
$$

$$
MME_V = \frac{SSE(MAPT_V) - SSE(APT_V)}{SSE(APT_V)} \times 100\%, \quad \ldots(29)
$$

where $SSE(MAPT_E) = \sum_{i=1}^n \hat{\epsilon}_i^2$, $SSE(APT_E) = \sum_{i=1}^n \tilde{\epsilon}_i^2$, $SSE(MAPT_V) = \sum_{i=1}^n (\hat{\epsilon}_i^2 / \hat{V}[R_1|R_2])_{MAPT} - 1)'$ and $SSE(APT_V) = \sum_{i=1}^n (\tilde{\epsilon}_i^2 / \tilde{V}[R_1|R_2])_{APT} - 1)'$. Note that under the above definition, with respect to the expectation, the MAPT will be more effective than the APT whenever $MME_E$ is negative. Otherwise, the APT will be more efficient. Similar criterion applies to the variance via $MME_V$.

Even though we let the scaling variable follow the inverse gamma distribution, in reality, we still do not know the true value of $p$, the degrees of freedom parameter. Noting that the two conditional moments exist only when $p > 4$, we compute the above scale mixture effects in (28) and (29) for various values of $p$ beginning with $p = 6$. The computations for the intraday data on all the data sets against S&P500 resulted in smallest values for both $MME_E$ and $MME_V$ when $p = 6$. For purposes of comparison, we maintained the same $p = 6$ for the analysis of
the daily returns data as well. Consequently, we only present below the computations when the scaling variable follows inverse gamma distribution with 6 degrees of freedom. We do, however, present subsequently two graphs relating to the analysis of daily returns data from COKE and DELL that illustrate the behavior of $MME_p$ against $p$, the degrees of freedom. The following table provides a summary of the relevant computations including the mixture model effects upon the two conditional moments.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>( \text{SSE} (\text{MAP}) )</th>
<th>( \text{SSE} (\text{APT}) )</th>
<th>( \text{SSE} (\text{MAP}^p) )</th>
<th>( \text{SSE} (\text{APT}^p) )</th>
<th>( \text{MME}_p )</th>
<th>( \text{MME}^p )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CISCO</strong></td>
<td></td>
<td>( 0.11153 ) &amp; ( 0.08893 ) &amp; ( 0.23634 ) &amp; ( 0.12124 ) &amp; ( 0.11138 ) &amp; ( 0.22474 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 0.10990 ) &amp; ( 0.08739 ) &amp; ( 0.23595 ) &amp; ( 0.12115 ) &amp; ( 0.11371 ) &amp; ( 0.22469 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 1369.86 ) &amp; ( 846.70 ) &amp; ( 838.78 ) &amp; ( 97400 ) &amp; ( 116028 ) &amp; ( 46992 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 1318.42 ) &amp; ( 608.91 ) &amp; ( 801.34 ) &amp; ( 274306 ) &amp; ( 116924 ) &amp; ( 85929 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 1.49 ) &amp; ( 1.77 ) &amp; ( 0.17 ) &amp; ( 0.08 ) &amp; ( 0.06 ) &amp; ( 0.02 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 3.90 ) &amp; ( 39.51 ) &amp; ( 4.61 ) &amp; ( -64.49 ) &amp; ( -0.77 ) &amp; ( -45.31 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>COKE</strong></td>
<td></td>
<td>( 0.05792 ) &amp; ( 0.07805 ) &amp; ( 0.18070 ) &amp; ( 0.05983 ) &amp; ( 0.08325 ) &amp; ( 0.14413 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 0.05786 ) &amp; ( 0.08698 ) &amp; ( 0.18040 ) &amp; ( 0.05983 ) &amp; ( 0.08326 ) &amp; ( 0.14414 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 1237.81 ) &amp; ( 1224.72 ) &amp; ( 742.12 ) &amp; ( 71798 ) &amp; ( 83897 ) &amp; ( 86866 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 992.09 ) &amp; ( 1064.62 ) &amp; ( 690.35 ) &amp; ( 138057 ) &amp; ( 368970 ) &amp; ( 109912 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 0.11 ) &amp; ( 0.09 ) &amp; ( 0.17 ) &amp; ( 0.01 ) &amp; ( -0.01 ) &amp; ( -0.01 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 24.77 ) &amp; ( 15.04 ) &amp; ( 7.50 ) &amp; ( -47.99 ) &amp; ( -77.26 ) &amp; ( -20.97 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DELL</strong></td>
<td></td>
<td>( 0.19148 ) &amp; ( 0.22048 ) &amp; ( 0.34110 ) &amp; ( 0.18408 ) &amp; ( 0.22106 ) &amp; ( 0.33951 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 0.18734 ) &amp; ( 0.22044 ) &amp; ( 0.33992 ) &amp; ( 0.18390 ) &amp; ( 0.22108 ) &amp; ( 0.33945 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 1104.44 ) &amp; ( 839.58 ) &amp; ( 851.16 ) &amp; ( 102359 ) &amp; ( 136667 ) &amp; ( 145873 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 713.12 ) &amp; ( 699.83 ) &amp; ( 884.14 ) &amp; ( 205534 ) &amp; ( 370057 ) &amp; ( 437400 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 2.21 ) &amp; ( 0.02 ) &amp; ( 0.35 ) &amp; ( 0.09 ) &amp; ( -0.01 ) &amp; ( 0.02 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 54.87 ) &amp; ( 19.97 ) &amp; ( -3.73 ) &amp; ( -50.20 ) &amp; ( -63.07 ) &amp; ( -66.65 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MSFT</strong></td>
<td></td>
<td>( 0.07336 ) &amp; ( 0.08685 ) &amp; ( 0.22460 ) &amp; ( 0.08233 ) &amp; ( 0.09121 ) &amp; ( 0.21278 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 0.07231 ) &amp; ( 0.08641 ) &amp; ( 0.22292 ) &amp; ( 0.08228 ) &amp; ( 0.09120 ) &amp; ( 0.21264 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 771.66 ) &amp; ( 1262.05 ) &amp; ( 2448.35 ) &amp; ( 96418 ) &amp; ( 151145 ) &amp; ( 430029 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 483.27 ) &amp; ( 1025.31 ) &amp; ( 2111.51 ) &amp; ( 161518 ) &amp; ( 210279 ) &amp; ( 910136 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 1.45 ) &amp; ( 0.51 ) &amp; ( 0.75 ) &amp; ( 0.06 ) &amp; ( 0.02 ) &amp; ( 0.06 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( 59.68 ) &amp; ( 23.09 ) &amp; ( 15.95 ) &amp; ( -40.30 ) &amp; ( -28.12 ) &amp; ( -52.75 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In what follows, we present selected graphs representing daily and intraday data analyses for CISCO in year 1998. The graphs represent the corresponding model fits, prediction curves based upon twice and thrice the conditional standard deviation and the normal probability plots for the residuals.
Figure 1a. Fits for both APT and MAPT models. The APT linear fit and the MAPT scale mixture fit for the daily log returns data for CISCO, year 1998.

Figure 1b, 1c. Residual plots and bands. Figure 1b plots residuals for the MAPT model together with two and three standard deviation bands for CISCO, year 1998. Figure 1c represents the same for the APT model.

Figure 1d, 1e. Normal probability plots. Figure 1d represents the normal probability plot for residuals from the MAPT fit for CISCO, year 1998. Figure 1e represents the same for APT fit.
Figure 2a.  Fits for both APT and MAPT models. The APT linear fit and the MAPT scale mixture fit for the intra-day log returns data for CISCO, year 1998.

Figure 2b, 2c.  Residual plots and bands. Figure 2b plots residuals for the MAPT model together with two and three standard deviation bands for CISCO, year 1998. Figure 2c represents the same for the APT model.

Figure 2d, 2e. Normal probability plots. Figure 2d represents the normal probability plot for residuals from the MAPT fit for CISCO, year 1998. Figure 2e represents the same for APT fit.
Discussion

In comparing the two models, MAPT and APT, we first note from Table 2 that the $MME_{y}$ percentages range between a minimum of -0.01% (DELL, intraday, in year 1999) to a maximum of 2.21% (DELL, daily, in year 1998) with median value of 0.08%. Thus, it is quite evident that the two models are close to each other for all fits considered in the analysis. The linear APT and the nonlinear MAPT appear to behave in the same manner with regard to the expectation fits. This implies that the contribution of the nonlinear term in the MAPT is hardly significant, that is, on the whole, the regression is captured equally well by both the nonlinear MAPT and its linear counterpart in daily as well as intraday data analyses.

Now, one might ask whether the same is true for the conditional variance. Again, from Table 2, one notes that the $MME_{y}$ percentages behave rather differently for the daily and the intraday analyses. In particular, all the $MME_{y}$ percentages for the daily returns are positive except for DELL in year 2000, whereas the $MME_{y}$ percentages for the intraday returns are all negative. The median value of the $MME_{y}$ percentages for the daily returns works out to 17.96%. On the other hand, the median for the $MME_{y}$ percentages for the intraday analyses is -49.10%. Thus, with regard to conditional variance, the APT seems to outperform the MAPT for the daily returns, whereas the MAPT is a far more superior fit for the analyses of the intraday returns.

We believe the difference between the MAPT and APT in capturing the variability can be traced back to the distributional properties of the log asset returns. As such, normal probability plots of the residuals from the MAPT and APT fits in the daily and intraday analyses are constructed and carefully studied. To conserve space, we only include the normal probability plots of the residuals from the MAPT and APT fits in the daily and intraday analyses of CISCO for the year 1998. It is important to note that the normal probability plots presented in Figures 1d and 1e for the daily returns are representative of what we have seen for most of the other daily return analyses. Similarly, the plots presented in Figures 2d and 2e also overwhelmingly represent the other plots for the intraday return analyses.

Figures 1d and 1e clearly demonstrate that the underlying daily log asset returns follow the normal distribution whereas Figures 2d and 2e show beyond doubt that the intraday log asset returns depart sharply from the normal distribution. Moreover, one can conclude from Figures 2d and 2e that the intraday log asset returns satisfy symmetry even as they demonstrate a heavy-tailed nature in their distribution. Note that the sharp departure from normality and strong indication of symmetry support the choice of the inverse gamma distribution for the scaling variable leading to a heavy-tailed $t$-type of distribution for the conditional intraday returns. It is this symmetry and heavy-tailed nature of the intraday return distribution as opposed to the normal distribution behavior of the daily log asset returns that together explain the differences in MAPT and APT.

Specifically, the scale mixture family with a heavy-tailed nature, particularly with small degrees of freedom ($p = 6$), is a clear misspecification of the underlying distribution for the daily returns. Hence, it is only appropriate that the linear APT fitted under the assumption of normality should outperform the scale mixture model in this case. Recall that the median $MME_{y}$
percentage for the daily returns is positive, 17.96%. The symmetric nature of the normal distribution ensures that the misspecified MAPT compares evenly with the APT in terms of the expectation fit.

On the other hand, the sharp departure from normality and strong indication of symmetry of the intraday returns support the choice of the inverse gamma distribution for the scaling variable leading to a heavy-tailed $t$-type distribution for these returns. Thus, one is able to simultaneously rationalize the observations regarding the intraday return analysis. Namely, the symmetric nature of the intraday data, on average, compares evenly with both the APT fit as well as the MAPT fit, whereas the heavy-tailed nature of the data is captured more sharply by the MAPT only (median $MME_\nu = -49.10\%$).

It is easy to see the reason behind the normality of the daily returns data as well as the non-normality of the intraday data. The intraday data being high-frequency data has little scope for aggregation effect, whereas the daily data can be assumed to have substantial aggregation effect.

**Choice of Degrees of Freedom, $p$**

Finally, we wish to discuss the behavior of $MME_\nu$ as a function of $p$, the degrees of freedom. Note that the expectation fit $MME_\nu$ is practically unaffected by the choice of $p$. We illustrate the behavior of $MME_\nu$ as a function of $p$ through the analysis of daily returns for COKE in year 1999 and also the analysis of daily returns for DELL in year 1998. There is some contrasting evidence in both of these analyses. In the case of the analysis for COKE, the normal probability plot in Figure 3a shows clear departure from normality. On the other hand, the normal probability in Figure 4a shows the normal distribution to be equivocal in this case. Accordingly, one anticipates that the MAPT should fit the COKE data better perhaps with somewhat higher degrees of freedom rather than the very low value of $p = 6$. In contrast, the APT should fit the DELL data well and the value of $p$ in reality should be extremely large. Figures 3b and 4b plot the behavior of $MME_\nu$ against $p$ for COKE and DELL, respectively. As expected, $MME_\nu$ in Figure 3b drops sharply for increasing $p$ and in fact becomes negative subsequent to $p = 17$, thus indicating that the MAPT with some $p$ value below 17 would be the best fit. Due to some computational overruns associated with the Bessel functions, we are unable to go beyond $p = 25$ as provided in the plot. Otherwise, one can capture the best fitting $p$ in this case.

In the case of the analysis of DELL, the $MME_\nu$ in Figure 4b always stays positive for increasing $p$ even as it monotonically decreases in $p$. Therefore, the best fitting $p$ will perhaps be achieved only at a very high value. This is consistent with the fact that the underlying distribution is normal.

**CONCLUDING REMARKS AND DIRECTIONS FOR FUTURE RESEARCH**

The main objective of this study is to understand the nonlinear aspects of the multifactor pricing models. Empirical evidence has disproved the omnipresence of the conventional normality assumption on the financial asset returns, which prompts the need to consider situations under...
Figure 3a. Normal probability plot. The plot is based on residuals from the linear APT model fit to log returns data for AT&T, year 2000.

Figure 3b. \( MME, \) against \( p, \) the degrees of freedom. The mixture model effect for the conditional variance against the APT model is computed for daily log returns data for AT&T, year 2000, at various degrees of freedom parameter \( p \) under the inverse gamma distribution for the scaling variable.
Figure 4a. **Normal probability plot.** The plot is based on residuals from the linear APT model fit to log returns data for DELL, year 1998.

Figure 4b. **$MME$, against $p$, the degrees of freedom.** The mixture model effect for the conditional variance against the APT model is computed for daily log returns data for DELL, year 1998, at various degrees of freedom parameter $p$ under the inverse gamma distribution for the scaling variable.
strong departures from normality. We develop a general framework for understanding the properties of the multifactor pricing models when the log asset returns are modeled by members of the scale mixture family. The derived analytical expressions contain nonlinear terms such as the conditional expectation and variance-covariance of the scaling variable. The presence of these terms may explain the nonlinear and heteroskedastic features exhibited in high-frequency financial data.

Empirical analyses are performed on data sets involving both daily and intraday log returns of the stocks from the Cisco Systems, Inc. (CISCO), Coca-Cola Company (COKE), Dell Computer Corporation (DELL), and the Microsoft Corporation (MSFT) conditional upon the log returns of the S&P500 index using the methodologies discussed in Section III. The results reveal that the nonlinear model dominates over its linear counterpart in modeling high-frequency financial data with symmetrical and heavy-tailed features by a median mixture model effect (MME) of -49.10%.

In conclusion, we recommend the use of the nonlinear model (MAPT) over the linear model (APT) in the analysis of (non-normal) intraday data since the former does a better job in explaining the conditional variance. However, if there is no dispute on the normality assumption, especially in the analysis of daily data, the linear model (APT) is preferred because of its simplicity.

Some open questions and directions for future research have been identified and thus proposed. First, this study only considers the formulation of nonlinear models in the cross-sectional framework. Reformulation of the techniques discussed here to incorporate time-series effects is deemed feasible, though difficult. This will definitely improve the practicability and predictability of the proposed model. Second, instead of restricting the volatility process to a scale mixture of normal distribution, it could be generalized even more by being a random matrix. Finally, as seen in Section IV, low degrees of freedom of the inverse gamma mixing variable produce better results and are thus preferred. However, the estimation of the optimal value for the degrees of freedom is outside the scope of this study. A possible way of determining this value may be attributed to the use of kernel estimation in nonparametric statistics.
Proof of Theorem 2

From Cambanis et al. (2000), it can be seen that

\[
E[A^j | R_z = r_z] = \int_0^\infty a^{j-\frac{1}{2}} \exp \left( -\frac{1}{2a} \| r_z - \mu_z - m_z a \|^2 \right) f_A(a) \, da
\]

\[
= \int_0^\infty a^{j-\frac{1}{2}} \exp \left( -\frac{1}{2a} \| r_z - \mu_z - m_z a \|^2 \right) f_A(a) \, da,
\]

where \( f_A(a) \) denotes the probability density function of the scaling random variable \( A \). Thus, it suffices to derive the expression for the integral in the numerator (say \( NUM \)) for any \( j \geq 0 \). We shall provide the necessary details for Part I of Theorem 2, where \( A \sim p / \chi_p^2 \). The details for Part II, where \( A \sim \Gamma( p,1 ) \), can be obtained similarly and are thus omitted.

\[
NUM = \frac{p^{p/2}}{2^{p/2} \Gamma(p/2)} \int_0^\infty a^{2j+(p-k)/2} \exp \left( -\frac{a}{2} \| m_z \|^2 - \frac{1}{2a} \| r_z - \mu_z \|^2 \right) \left( p + \| r_z - \mu_z \|^2 \right) \, da
\]

\[
= \frac{p^{p/2} \Gamma(2j+(p-k)/2)}{2^{p/2} \Gamma(p/2)} \int_0^\infty a^{2j+(p-k)/2} \exp \left( -\frac{a}{2} - \frac{1}{2a} (xy)^2 \right) \, da
\]

where \( x^2 = p + \| r_z - \mu_z \|^2 \) and \( y^2 = \| m_z \|^2 \). Using eq. (8.432.6) from Gradshteyn and Ryzik (1980), the above expression can be simplified as

\[
NUM = \frac{p^{p/2} \Gamma(2j+(p-k)/2)}{2^{p/2} \Gamma(p/2)} \left( \frac{x}{y} \right)^{2j+(p-k)/2} K_{\frac{2j+(p-k)/2}{2}}(xy).
\]

The expressions for the two conditional moments in Part I of Theorem 2 then follow easily by substituting eq. (A2) with \( j = 1 \) and 2, respectively, in the numerator of eq. (A1), and (A2) with \( j = 0 \) in the denominator of eq. (A1).

REFERENCES


Nonlinear Properties of Multifactor Financial Models


