Phoebe Koundouri, Nikolaos Kourogenis, Nikitas Pittis, Panagiotis Samartzis

Factor models of stock returns: GARCH errors versus time-varying betas

Article (Accepted version) (Refereed)

Original citation:
Koundouri, Phoebe, Kourogenis, Nikolaos, Pittis, Nikitas and Samartzis, Panagiotis (2016) *Factor models of stock returns: GARCH errors versus time-varying betas.* Journal of Forecasting. ISSN 0277-6693

DOI: 10.1002/for.2387

Factor Models of Stock Returns: GARCH Errors versus Time - Varying Betas

Phoebe Koundouri*† Nikolaos Kourogenis‡ Nikitas Pittis‡

Panagiotis Samartzis‡

Abstract

This paper investigates the implications of time-varying betas in factor models for stock returns. It is shown that a single-factor model (SFMT) with autoregressive betas and homoscedastic errors (SFMT-AR) is capable of reproducing the most important stylized facts of stock returns. An empirical study on the major US stock market sectors shows that SFMT-AR outperforms, in terms of in-sample and out-of-sample performance, SFMT with constant betas and conditionally heteroscedastic (GARCH) errors, as well as two multivariate GARCH-type models.

Keywords: autoregressive beta, stock returns, single factor model, conditional heteroscedasticity, in-sample performance, out-of-sample performance.

JEL Classification: C22, G10, G11, G12

---

*Department of International and European Economic Studies, Athens University of Economics and Business and Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, UK.

† *Correspondence to*: Phoebe Koundouri, Department of International and European Economic Studies, Athens University of Economics, 76, Patission Street, 104 34 Athens, Greece. Email: pkoundouri@aueb.gr. Tel: +302108203455.

‡Department of Banking and Financial Management, University of Piraeus.

1 Introduction

The analysis of the statistical properties of stock returns has been a research area of great interest since the beginning of 1950s. One of the most useful and intuitive statistical models for stock returns is the single factor model (SFM) which, together with its multivariate generalization (the multiple factor model), form the basis for many asset pricing models, such as the Arbitrage Pricing Model (APT), put forward by Ross (1976) or the Intertemporal Capital Asset Pricing Model (ICAPM), introduced by Merton (1973). The SFM attempts to capture the intuitive idea that asset returns are driven by unanticipated changes (surprises) of a common underlying factor. More specifically, in the context of SFM, the return, r_i, of a security (or a portfolio) i, $i = 1, 2, ..., n$, is linearly related to an exogenous (zero-mean) variable M through the linear regression $r_i = a_i + \beta_i M + u_i$. The error term, u_i, in this model has zero mean, finite variance and satisfies the condition $E(u_i \mid M) = 0$, $\forall i = 1, 2, ..., n$. Furthermore, the theoretical assumption that the correlation between r_i and r_j, $i \neq j$ stems solely from the common "causal" factor M entails the assumption that $Cov(u_i, u_j) = 0$, for every i and j, $i \neq j$. The slope coefficient, β_i, is interpreted as a measure of the systematic risk of the stock i, and is usually referred to as the "beta coefficient", or simply the "beta" of the stock i.

SFM is a single period model. In the estimation of this model using time series data, it is usually assumed that the aforementioned linear relationship between r_i, $i = 1, 2, ..., n$ and M is time-invariant. Under this (often implicit) assumption, the stochastic process $\{r_{i,t}\}$, $i = 1, 2, ..., n$ is probabilistically caused by the stochastic process $\{M_t\}$ through the temporal relationship $r_{i,t} = a_i + \beta_i M_t + u_{i,t}$, hereafter referred to as SFMT, with $\{u_{i,t}\}$ being an *iid* process with zero mean and finite variance, $\sigma^2_{\nu_i}$. As a consequence, all the statistical properties of $\{r_{i,t}\}$, $i = 1, 2, ..., n$, are determined solely by those of $\{M_t\}$ and $\{u_{i,t}\}$. This means that SFMT is a well-specified statistical model and hence, empirically adequate. Empirical adequacy of SFMT means that the parameters a_i, β_i and $\sigma^2_{\nu_i}$ are time-invariant, and the error term $u_{i,t}$ is an *iid* process.

Has SFMT been found to be empirically adequate? The answer is negative. There are at least two sources for the empirical failure of SFMT. The first one lies in the fact that the error process $\{u_{i,t}\}$ has been found to exhibit temporal dependence, which is usually identified as conditional heteroscedasticity (CH). The second source of empirical inadequacy of SFMT comes from studies suggesting that the regression coefficient β_i is not constant over time. Important studies offering evidence for a time-varying beta, $\beta_{i,t}$, include Blume (1971, 1975), Fabozzi and Francis (1978), Fisher and Kamin (1985), Sunder (1980), Ohlson and Rosenberg (1982), Bos and Newbold (1984), Collins, Ledolter and Rayburn (1987), Bos and Fetherston (1992, 1995) and Faff, Lee and Fry (1992).

The response of the empirical literature to the aforementioned empirical failures of SFMT has taken various forms among which the following two are the most prominent. The first response consists in replacing the assumption of independence of the error sequence with the assumption that $\{u_{i,t}\}$ exhibits non-linear dependence, which usually takes the form of a GARCH-type model. Note that the resulting model, hereafter referred to as SFMT-GARCH, retains the (rather strong) assumption of a time-invariant beta. The second response focuses on the problem of beta instability, thus specifying models with stochastic parameters. For example, Shanken (1990) models the time varying beta as a linear function of observable state variables. Alternatively, the time varying beta is often treated as a stochastic (hidden) process. To this end, Fabozzi and Francis (1978) assumed that $\beta_{i,t}$ is an i.i.d process with finite variance, while Fisher and Kamin (1985), Sunder (1980), Bos and Newbold (1984) and Jostova and Philipov (2005) allowed for persistence in the variation of beta by assuming that $\beta_{i,t}$ follows a first-order autoregressive (AR(1)) process (including the case of a random walk). Ohlson and Rosenberg (1982) and Collins, Ledolter and Rayburn (1987) proposed a hybrid of these two models by assuming that $\beta_{i,t}$ is the sum of a random and an AR(1) processes[1]. Overall, these studies suggest the emergence of another variant of SFMT, namely the one in which the slope coefficient is modeled as an autoregressive process, whilst the error term $u_{i,t}$ retains its independence property. The resulting model in which $\beta_{i,t}$ is assumed to follow an AR(1) process, will be hereafter referred to as SFMT-AR.

Both SFMT-GARCH and SFMT-AR may be thought of as emerging from imposing alternative sets of restrictions on the vector stochastic process $\{Z_{i,t}\}$, $Z_{i,t} = [r_{i,t}, M_t]'$. Under this point of view, the question of which of the two models is empirically adequate is translated into the question of which of the two sets of restrictions is supported by the data, which has both empirical and theoretical interest. Indeed, moving from SFMT-GARCH to SFMT-AR may be *theoretically* interpreted as shifting interest from imposing conditions on the temporal behavior of the non-systematic risk to modeling explicitly the dynamics of the (theoretically more interesting) systematic risk. In other words, in spite of the fact that SFMT-GARCH and SFMT-AR may be thought of as alternative parameterizations of the same process, these two models offer quite different theoretical explanations of the observed regularities. In the context of SFMT-AR and SFMT-GARCH, the stylized facts of stock returns are explained (at least partly) by the persistent variation of the systematic risk or that of the idiosyncratic risk, respectively.[2]

The preceding discussion leads, quite naturally, to the following question: Is there any SFMT-type model that combines the main features of both SFMT-GARCH and SFMT-AR? In an attempt to produce such a model, one may assume that $\{Z_{i,t}\}$ follows a bivariate GARCH process. In such a case, the model that arises by con-

[1] More recently, Andersen et al. (2005) offered convincing evidence for the autoregressive nature of betas. Building on their previous work on the relationship between realized volatility and conditional covariance matrix (Andersen et al., 2003), they constructed quarterly and monthly realized betas for 25 stocks of the Dow Jones Industrial Average index using high-frequency returns. These realized beta series exhibit positive serial correlation, which is adequately captured by stationary, low-order autoregressive models (see also Jostova and Philipov, 2005, for additional evidence on the autoregressive nature of beta).

[2] The motivation for a comparative study of SFMT-AR and SFMT-GARCH is enhanced by the fact that this remark remains valid when SFMT-AR and SFMT-GARCH are augmented by additional risk factors.

ditioning on M_t, hereafter referred to as SFMT-B-GARCH, exhibits a time varying beta (under the usual covariance/variance interpretation) and a conditionally heteroscedastic error. It is important to note, however, that although SFMT-B-GARCH is perfectly eligible as a statistical model, it nonetheless lacks the "theoretical flavor" of SFMT-AR and SFMT-GARCH. This is because, SFMT-B-GARCH treats $r_{i,t}$ and M_t as causally symmetrical, instead of explicitly assuming that M_t is the sole causal factor of $r_{i,t}$. Put differently, the presence of M_t on the right-hand side of the SFMT equation should not merely be the result of "conditioning on M_t," but it should reflect the theoretical role of M_t as the common cause of all $r_{i,t}$'s, $i = 1, 2, ..., n$. However, since quite often, the shortage of theoretical elegance is more than compensated by forecasting performance, we include SFMT-B-GARCH in our set of competing SFMT-type models.

What is the empirical performance of SFMT-GARCH, SFMT-AR and SFMT-B-GARCH? Since all these models exhibit mean conditional independence properties (since M_t represents unanticipated changes of the risk factor), their comparison should focus on how well each of these models approximates the second-order effects of $\{Z_{i,t}\}$. To this end, we distinguish between *in-sample* and *out-of-sample* performance. In-sample performance of a given model is satisfactory, if each and every probabilistic assumption that defines this model is supported by the available data. On the other hand, the out-of-sample performance of any of the aforementioned models is determined by the ability of the model to predict the covariance matrix $\Sigma_{t|t-1}$ of $\mathbf{r}_t$, $\mathbf{r}_t = [r_{1,t}, r_{2,t}, ..., r_{n,t}]'$, accurately, based on the information available up to $t-1$. Since $\Sigma_{t|t-1}$ is unobservable, the question of the out-of-sample performance of the models under study may reduce to that of which of these models results in the most efficient diversification of the underlying n assets. More specifically, if these n assets are used at each t to construct optimal portfolios in the Markowitz sense, then which of the three competing models under consideration, namely SFMT-GARCH, SFMT-AR and SFMT-B-GARCH, comes closer to delivering the Markowitz ideal portfolio? Put differently, which of these models achieve the most efficient management of portfolio risk? Moreover, is the best of these models good enough? In other words, does any of the aforementioned models produce diversification gains that are superior to those achieved by the naive $(1/N)$ rule? This last question becomes particularly interesting in the light of the strong evidence, offered by De Miquel, Garlappi and Uppal (2007), against the ability of several standard methods for estimating $\Sigma_{t|t-1}$ to beat the $(1/N)$ rule in terms of portfolio efficiency. Note however, that the aforementioned results refer to an observation frequency, namely monthly, in which most of the second-order effects have been washed out via temporal aggregation. This leaves an important question unanswered: Does any of the aforementioned parametric models for CH - when applied to higher than monthly frequencies - produce any diversification gains over the $(1/N)$ rule?

The remainder of this paper is organized as follows: Section 2 defines the SFMT-AR model, analyzes its theoretical properties and studies the problem of estimating its parameters in some detail. More specifically, the first part of this section demonstrates that SFMT-AR implies that the generating process $\{r_{i,t}\}$ exhibits the theoretical properties of conditional heteroscedasticity and leptokurtosis. A rather interesting result, emerging from this analysis is that SFMT-AR produces CH even in the case in which the factor process $\{M_t\}$ is independent. This result, already introduced above, implies that the empirical regularities of stock returns may be caused not by the probabilistic properties of the underlying risk factor, but rather by the persistent time variation of the systematic risk. The second part of Section 2 discusses estimation issues concerning SFMT-AR and presents the results of a small Monte Carlo study, which show that the proposed estimator exhibits satisfactory finite-sample properties. Section 3 estimates SFMT-GARCH, SFMT-AR and SFMT-B-GARCH using weekly US stock returns data and compares their in-sample and out-of-sample forecasting performance. To account for the possibility that M_t is a poor proxy of the market portfolio, we also consider an additional model, hereafter referred to as SFMT-MGARCH, in which the errors of the ten factor models are jointly modelled as a multivariate GARCH process. The results from this section suggest that none of the four heteroscedastic factor models under consideration is fully adequate in terms of the adopted in-sample criteria. However, all these models offer significant portfolio efficiency gains over the $(1/N)$ rule. Moreover, with the exception of SFMT-B-GARCH, these models dominate, in terms of all the usual out-of-sample criteria adopted in the literature, both the homoscedastic SFMT model and the method of estimating $\Sigma_{t|t-1}$ via the sample moments. Among the four heteroscedastic factor models under consideration, SFMT-AR seems to achieve the best out-of-sample performance, closely followed by SFMT-GARCH. Interestingly, the performance of SFMT-B-GARCH, that is the model supposed to combine the virtues of SFMT-AR and SFMT-GARCH, is remarkably poor. Section 5 concludes the paper.

2 The Single Factor Model with Autoregressive Beta (SFMT-AR)

First, a note on notation. Throughout the paper, we will use normal letters for numbers or random variables, bold non-capital letters for vectors and bold capital letters for matrices. Let us consider a market with n assets (stocks) and let $r_{i,t}$ be the one-period continuously compounded return on an individual stock, defined as $r_{i,t} = p_{i,t} - p_{i,t-1}$, where $p_{i,t}$ is the natural logarithm of the price of the particular stock. Following the discussion of the previous section, we assume that $r_{i,t}$ is related to an observable factor, M_t via the following relationship:

$$r_{i,t} = \alpha_i + (\beta_i + \beta_{i,t})M_t + u_{i,t}, \; i = 1, 2, ..., n \tag{1}$$

where α_i and β_i are real numbers, and $u_{i,t}$, $\beta_{i,t}$, are zero-mean sequences of random variables whose exact properties will be defined below. Equation (1) can be written in vector form as follows:

$$\mathbf{r}_t = \boldsymbol{\alpha} + (\boldsymbol{\beta} + \boldsymbol{\beta}_t)M_t + \mathbf{u}_t, \tag{2}$$

where $\mathbf{r}_t' = [r_{1,t}, r_{2,t}, \ldots, r_{n,t}]$, $\boldsymbol{\alpha}' = [\alpha_1, \alpha_2, \ldots, \alpha_n]$, $\boldsymbol{\beta}' = [\beta_1, \beta_2, \ldots, \beta_n]$ and $\mathbf{u}_t' = [u_{1,t}, u_{2,t}, \ldots u_{n,t}]$.

Assumption M: $\beta_{i,t}$ follows a zero-mean AR(1) process,

$$\beta_{i,t} = \varphi_i \beta_{i,t-1} + \varepsilon_{i,t}, \;\; |\varphi_i| < 1, \, 1 \leq i \leq n \tag{3}$$

and

$$\begin{bmatrix} \mathbf{u}_t \\ M_t \\ \boldsymbol{\varepsilon}_t \end{bmatrix} \sim NIID \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_u & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_m^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_\varepsilon \end{bmatrix} \right) \tag{4}$$

where $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \ldots, \varepsilon_{n,t}]'$, $\boldsymbol{\Sigma}_u = diag\left(\sigma_{u_1}^2, \ldots, \sigma_{u_n}^2\right)$, and $\boldsymbol{\Sigma}_\varepsilon = diag\left(\sigma_{\varepsilon_1}^2, \ldots, \sigma_{\varepsilon_n}^2\right)$.

Remark: The assumption that M_t is independent may appear to be overly restrictive and inconsistent with the empirical properties of the variables that are usually called to play the role of M_t. However, if CH is deduced from a model in which M_t is independent, it is quite natural to assume that this result will continue to hold in the case that M_t exhibits properties similar to those that SFMT-AR attempts to explain. In other words, SFMT-AR with independent M_t constitutes the least favorable case for deriving CH.

From assumption **M** we have that, $\boldsymbol{\Sigma}_\beta := Var(\boldsymbol{\beta}_t) = diag\left(\sigma_{\beta_1}^2, \sigma_{\beta_2}^2, \ldots, \sigma_{\beta_n}^2\right)$, where,

$$\sigma_{\beta_i}^2 = Var\left(\beta_{i,t}\right) = \frac{\sigma_{\varepsilon_i}^2}{1 - \varphi_i^2} \, .$$

Equation (3) can be also written in vector form as

$$\boldsymbol{\beta}_t = \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \boldsymbol{\varepsilon}_t \ , \tag{5}$$

where $\boldsymbol{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_n)$.

Remark:

In the case of constant beta, i.e. $r_{i,t} = \alpha_i + \beta_i M_t + u_{i,t}$, $i = 1, 2, ..., n$, the assumption that $[\mathbf{u}_t', M_t]'$ is NIID with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_c$ defined as

$$\boldsymbol{\Sigma}_c = \begin{bmatrix} \boldsymbol{\Sigma}_u & 0 \\ 0 & \sigma_m^2 \end{bmatrix} ,$$

implies that $\mathbf{r}_t$ is *niid* with $E(\mathbf{r}_t) = \boldsymbol{\alpha}$ and $Var(\mathbf{r}_t) = \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}_u$. On the contrary, as will be shown below, the assumption that $[\mathbf{u}_t', M_t, \boldsymbol{\varepsilon}_t']'$ is *niid*, that is assumption (4), together with the assumption of autoregressive betas, that is assumption (3), imply that $\mathbf{r}_t$ is a non-Gaussian stationary process, exhibiting non-linear temporal dependence.

2.1 Theoretical Properties of SFMT-AR

Let us now analyze the probabilistic properties of the process $\mathbf{r}_t$, implied by SFMT-AR. Let $\mathcal{F}_{t-1} = \sigma(\mathbf{r}_1, ..., \mathbf{r}_{t-1}, M_1, ..., M_{t-1})$ to be the information up to time $t-1$, where $\sigma(\mathbf{r}_1, ..., \mathbf{r}_{t-1}, M_1, ..., M_{t-1})$ denotes the smallest sigma-algebra generated by the collection $\{\mathbf{r}_1, ..., \mathbf{r}_{t-1}, M_1, ..., M_{t-1}\}$.

(I) Conditional Heteroscedasticity

From assumption $\mathbf{M}$ we obtain:

$$\begin{aligned} Var(\mathbf{r}_t) &= E\left[((\boldsymbol{\beta} + \boldsymbol{\beta}_t)M_t + \mathbf{u}_t)((\boldsymbol{\beta} + \boldsymbol{\beta}_t)M_t + \mathbf{u}_t)'\right] \\ &= \sigma_m^2 E\left[(\boldsymbol{\beta} + \boldsymbol{\beta}_t)(\boldsymbol{\beta} + \boldsymbol{\beta}_t)'\right] + \Sigma_u = \sigma_m^2 \left(\Sigma_\beta + \boldsymbol{\beta}\boldsymbol{\beta}'\right) + \Sigma_u \end{aligned} \tag{6}$$

and

$$
\begin{aligned}
Var(\mathbf{r}_t \quad | \quad & \mathcal{F}_{t-1}) = \\
&= E\left[((\boldsymbol{\beta}+\boldsymbol{\beta}_t)M_t+\mathbf{u}_t)\left((\boldsymbol{\beta}+\boldsymbol{\beta}_t)M_t+\mathbf{u}_t\right)' \mid \mathcal{F}_{t-1}\right] = \\
&= \sigma_m^2 E\left[(\boldsymbol{\beta}+\boldsymbol{\beta}_t)(\boldsymbol{\beta}+\boldsymbol{\beta}_t)' \mid \mathcal{F}_{t-1}\right] + \Sigma_u = \\
&= \Sigma_u + \sigma_m^2\left(\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)' + \Sigma_\varepsilon\right) = \\
&= \sigma_m^2\Sigma_\varepsilon + \Sigma_u + \sigma_m^2\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)\left(\boldsymbol{\beta}+\boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}\right)' \qquad (7)
\end{aligned}
$$

Under the diagonality of Σ_u the returns $r_{i,t}$, $i = 1, 2, \ldots, n$, are related only through M_t, in the sense that the idiosyncratic terms $u_{i,t}$ and $u_{j,t}$ do not contribute in $Cov\left(r_{i,t}, r_{j,t}\right)$ and $Cov\left(r_{i,t}, r_{j,t} \mid \mathcal{F}_{t-1}\right)$. Equation (7) demonstrates that SFM-AR implies that $\mathbf{r}_t$ is a conditionally heteroscedastic process.

Remarks:

(i) Equation (6), together with the martingale-property of $\{\mathbf{r}_t\}$ discussed below, imply that $\{\mathbf{r}_t\}$ is a second-order stationary process.

(ii) Under assumption **M,** the constant beta SFM arises as a special case in which $\boldsymbol{\varepsilon}_t \equiv 0$ for every i and t and $\boldsymbol{\Phi} \equiv \mathbf{0}$. In such a case, $Var\left(\mathbf{r}_t \mid \mathcal{F}_{t-1}\right) = Var\left(\mathbf{r}_t\right) = \sigma_m^2\boldsymbol{\beta}\boldsymbol{\beta}' + \Sigma_u$, which is time invariant. Conditional homoscedasticity arises also in the case of non-persistent random betas. Indeed, when the autoregressive parameters, φ_i, of the stochastic betas are zero (see, for example, Fabozzi and Francis, 1977), we have $Var\left(\mathbf{r}_t \mid \mathcal{F}_{t-1}\right) = Var\left(\mathbf{r}_t\right) = \sigma_m^2\Sigma_\varepsilon + \Sigma_u + \sigma_m^2\boldsymbol{\beta}\boldsymbol{\beta}'$, which means that r_t is conditionally homoscedastic. On the other hand, in the general case in which $\varphi_i \neq 0$, equation (7) implies that $Cov\left(r_{i,t}, r_{j,t} \mid \mathcal{F}_{t-1}\right)$ is time varying. In other words, the presence of conditional heteroscedasticity cannot be accounted for solely by assuming that $\left\{\beta_{i,t}\right\}$ is a random sequence**.** Indeed, it is the persistence of $\beta_{i,t}$ that gives rise to conditional heteroscedasticity.

(iii) As already noted, assumption **M** implies independence for the factor sequence $\{M_t\}$. This means that SFMT-AR is capable of producing CH solely in terms of the autoregressive nature of betas. Put it differently, individual stock returns are likely to exhibit volatility clustering, even if the single factor affecting them has much simpler dynamic properties.

(II) Leptokurtosis

We now show that SFMT-AR implies that the unconditional distribution of stock returns is a mixture of normal distributions and derive the kurtosis coefficient, which implies a positive excess kurtosis. First note that from the independence between $\mathbf{u}_t$, M_t and $\boldsymbol{\varepsilon}_t$, postulated in assumption **M**, conditional on the realization of $\boldsymbol{\beta}_t$ and all the information that is generated up to time $t-1$, $\mathcal{F}_{t-1}$, we have that $E\left[\mathbf{r}_t \mid \boldsymbol{\beta}_t, \mathcal{F}_{t-1}\right] = \boldsymbol{\alpha}$ and $Var\left[\mathbf{r}_t \mid \boldsymbol{\beta}_t, \mathcal{F}_{t-1}\right] = \Sigma_u + \sigma_m^2(\boldsymbol{\beta}+\boldsymbol{\beta}_t)(\boldsymbol{\beta}+\boldsymbol{\beta}_t)'$. On the other hand, since $[\mathbf{u}_t', M_t, \boldsymbol{\varepsilon}_t']'$ is multivariate normal, we directly conclude the following proposition:

Proposition 1: *The unconditional distribution of $\mathbf{r}_t$ is a mixture of normal distributions and is described by:*

$$\mathbf{r}_t \sim MN\left(\boldsymbol{\alpha}, \Sigma_u + \sigma_m^2(\boldsymbol{\beta}+\boldsymbol{\beta}_t)(\boldsymbol{\beta}+\boldsymbol{\beta}_t)'\right) \ , \tag{8}$$

where MN stands for the mixed normal distribution.

The analytic expression of the kurtosis coefficient of r_t is given in Theorem 1:

Theorem 1: *Under Assumption M, the kurtosis coefficient of the unconditional distribution of $r_{i,t}$ is given by:*

$$Kurt(r_{i,t}) = \frac{E\left[\left(r_{i,t} - E[r_{i,t}]\right)^4\right]}{Var^2(r_{i,t})} = 3 + \frac{12\beta_i^2\sigma_{\beta_i}^2\sigma_m^4}{Var^2(r_{i,t})}. \tag{9}$$

Proof: see Appendix A.

Remarks:

(i) The above theorem proves that, in general, stock returns are leptokurtic, except for the case that the excess kurtosis,

$$\frac{12\beta_i^2\sigma_{\beta_i}^2\sigma_m^4}{Var^2(r_{i,t})}$$

is equal to zero, i.e. when $\sigma_{\beta_i} = 0$ or, when $\sigma_{\beta_i} \neq 0$ and $\beta_i = 0$ (the case $\sigma_m = 0$ is ruled out a-priori since it implies a degenerate process for M_t).

(ii) Equation (9) shows that the degree of persistence of $\beta_{i,t}$ as measured by φ_i is not the only factor that affects the degree of leptokurtosis of the distribution of $r_{i,t}$. In other words, leptokurtosis may be present even if $\varphi_i = 0$, provided that $\beta_{i,t}$ is a stochastic sequence, that is, $\sigma_{\beta_i} \neq 0$.

(iii) In the context of the linear SFMT with constant beta, the leptokurtosis of $r_{i,t}$ could be accounted for by either the leptokurtosis of M_t or that of $u_{i,t}$ or both. In the context of SFMT-AR, leptokurtosis arises even under the assumption that M_t and $u_{i,t}$ (as well as $\beta_{i,t}$) are Gaussian processes.

2.2 Estimation Issues

We first use a Kalman filter approach to derive the Gaussian log-likelihood function of SFMT-AR. The parameters of this model may be estimated using the maximum likelihood method. Note that assumption **M** implies that conditional on M_t and $\mathcal{F}_{t-1}$, we have

$$\left(\begin{bmatrix} \boldsymbol{\varepsilon}_t \\ \mathbf{u}_t \end{bmatrix} \middle| M_t, \mathcal{F}_{t-1} \right) \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_\varepsilon & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_u \end{bmatrix} \right), \tag{10}$$

where $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_\varepsilon$ are diagonal matrices defined in section 2.

Next, let us define

$$\begin{aligned} \boldsymbol{\beta}_{t/t-1} &= E[\boldsymbol{\beta}_t \mid \mathcal{F}_{t-1}] \\ P_{t/t-1} &= E[(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t/t-1})'(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t/t-1}) \mid \mathcal{F}_{t-1}] \end{aligned}$$

to be the conditional mean and the conditional covariance matrix of $\boldsymbol{\beta}_t$, respectively. Then we have,

$$
\begin{aligned}
E[\mathbf{r}_t \mid M_t, \mathcal{F}_{t-1}] &= \boldsymbol{\alpha} + (\boldsymbol{\beta} + \boldsymbol{\beta}_{t/t-1})M_t, \\
Var[\mathbf{r}_t \mid M_t, \mathcal{F}_{t-1}] &= M_t^2 \mathbf{P}_{t/t-1} + \boldsymbol{\Sigma}_u, \\
Cov[\mathbf{r}_t, \boldsymbol{\beta}_t \mid M_t, \mathcal{F}_{t-1}] &= M_t \mathbf{P}_{t/t-1}.
\end{aligned}
$$

By virtue of (10), it follows that

$$
\left(\begin{bmatrix} \boldsymbol{\beta}_t \\ \mathbf{r}_t \end{bmatrix} \middle| M_t, \mathcal{F}_{t-1} \right) \sim N \left(\begin{bmatrix} \boldsymbol{\beta}_{t/t-1} \\ \boldsymbol{\alpha} + (\boldsymbol{\beta} + \boldsymbol{\beta}_{t/t-1})M_t \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{t/t-1} & M_t \mathbf{P}_{t/t-1} \\ M_t \mathbf{P}_{t/t-1} & M_t^2 \mathbf{P}_{t/t-1} + \boldsymbol{\Sigma}_u \end{bmatrix} \right).
$$

The above result allows us to derive the updating equations:

$$
\begin{aligned}
\boldsymbol{\beta}_{t/t} &= E[\boldsymbol{\beta}_t \mid \mathcal{F}_t] = \boldsymbol{\beta}_{t/t-1} + M_t \mathbf{P}_{t/t-1} \mathbf{F}_{t/t-1}^{-1} \mathbf{v}_{t/t-1}, \\
\mathbf{P}_{t/t} &= Var[\boldsymbol{\beta}_t \mid \mathcal{F}_t] = \mathbf{P}_{t/t-1}(\mathbf{I} - M_t^2 \mathbf{P}_{t/t-1} \mathbf{F}_{t/t-1}^{-1}),
\end{aligned}
$$

where $\mathbf{v}_{t/t-1} = \mathbf{r}_t - E[\mathbf{r}_t \mid M_t, \mathcal{F}_{t-1}] = \mathbf{r}_t - \boldsymbol{\alpha} - (\boldsymbol{\beta} + \boldsymbol{\beta}_{t/t-1})M_t$, and $\mathbf{F}_{t/t-1} = Var[\mathbf{r}_t \mid M_t, \mathcal{F}_{t-1}] = E[\mathbf{v}_{t/t-1}\mathbf{v}_{t/t-1}' \mid M_t, \mathcal{F}_{t-1}) = M_t^2 \mathbf{P}_{t/t-1} + \boldsymbol{\Sigma}_u$.

Finally, the prediction equations are given by:

$$
\begin{aligned}
\boldsymbol{\beta}_{t/t-1} &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1/t-1}, \\
\mathbf{P}_{t/t-1} &= \boldsymbol{\Phi} P_{t-1/t-1} \boldsymbol{\Phi} + \boldsymbol{\Sigma}_\varepsilon.
\end{aligned}
$$

Note that the eigenvalues (i.e. the diagonal elements) of the matrix $\boldsymbol{\Phi}$ are assumed to lie inside the unit circle, implying that $\boldsymbol{\beta}_t$ is covariance-stationary and thus, we may set the starting value for the recursion, $\boldsymbol{\beta}_{1/0} = 0$ and its associated MSE $vec(\mathbf{P}_{1/0}) = (I - (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi}))^{-1} vec(\boldsymbol{\Sigma}_\varepsilon)$, where $\otimes$ is the Kronecker product and vec is the linear transformation of a $n \times n$ matrix into a column of size $n^2 \times 1$ under which the columns of the matrix are stacked on top of one another. With these initial values for the recursion and a set of values for the hyper-parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_u$, we obtain the sequences $\{\boldsymbol{\beta}_{t/t-1}\}_{t=1}^T$ and $\{\mathbf{P}_{t/t-1}\}_{t=1}^T$.

Given the results above, the sample log-likelihood is given by:

$$\begin{aligned}\sum_{t=1}^{T} \log f(\mathbf{r}_t \quad | \quad M_t, \mathcal{F}_{t-1}) &= \\ &= -\frac{TN}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log \mid \mathbf{F}_{t/t-1} \mid -\frac{1}{2}\sum_{t=1}^{T}\mathbf{v}_{t/t-1}^{\prime}\mathbf{F}_{t/t-1}^{\prime}\mathbf{v}_{t/t-1}\end{aligned} \tag{11}$$

Note that if $\boldsymbol{\Sigma}_\varepsilon = \mathbf{0}$ and $\boldsymbol{\Phi} \neq \mathbf{0}$, then we end up with a zero-mean AR(1) model whose coefficients vary deterministically. In this case the log-likelihood function does not provide an estimator for $\boldsymbol{\Phi}$, since it attains the same maximum for any $\boldsymbol{\Phi}$ whose eigenvalues are less than one in absolute value. In other words, this particular parameter configuration causes identification failure for $\boldsymbol{\Phi}$. Pagan (1980) provides sufficient conditions for the maximum likelihood estimates of the parameters of general state space models to be consistent and asymptotically normal. In the case of the SFMT-AR model, these conditions amount to: (i) model identification (this excludes the case $\Sigma_\varepsilon = \mathbf{0}, \boldsymbol{\Phi} \neq \mathbf{0}$), (ii) stationarity of the state process, that is $|\phi_i| < 1$, $i = 1, 2, ..., n$, (iii) second-order stationarity of $\{r_{i,t}\}$, $i = 1, 2, ..., n$ (see Remark (i) in section 2.1) and (iv) the model parameters taking values inside the permissible parameter space. To maximize (11), we employ the Levenberg–Marquardt algorithm, put forward by Levenberg (1944), which has been shown to be more robust than the Gauss–Newton algorithm.

For the initial estimates of the hyper-parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_u$, we use OLS estimators. More specifically, we estimate the regression:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} M_t + \mathbf{u}_t, \tag{12}$$

from which we obtain the initial value of the hyper-parameter $\boldsymbol{\Sigma}_u^0$. Then, we apply rolling OLS to (12) to obtain rolling estimates, $\left[\boldsymbol{\alpha}_1^0, ...\boldsymbol{\alpha}_{n-k+1}^0\right]$ and $\left[\boldsymbol{\beta}_1^0, ..., \boldsymbol{\beta}_{n-k+1}^0\right]$ of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively and set $\boldsymbol{\alpha}^0 = \sum_{i=0}^{n-k+1} \frac{\boldsymbol{\alpha}_i^0}{n-k+1}$ and $\boldsymbol{\beta}^0 = \sum_{i=0}^{n-k+1} \frac{\boldsymbol{\beta}_i^0}{n-k+1}$, where k is the estimation window. Finally, $\boldsymbol{\Phi}^0$ and $\boldsymbol{\Sigma}_{\varepsilon}^0$ are obtained from the AR(1) regression $\boldsymbol{\beta}_t^0 = \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}^0 + \boldsymbol{\varepsilon}_t$, $t = 1, ..., n-k+1$.[3]

In order to examine the finite-sample performance of the proposed ML estimator under alternative sets of the SFMT-AR parameters, we conduct a small Monte Carlo study. In all the simulations that follow, the number of replications is equal to 5000 and the sample size, T, is set equal to 250, 500, and 1000. Although many alternative parameter sets were examined, we report the results from the following four representative cases, for $T = 1000$:[4]

$$\begin{aligned}
1.\ \left(a, \beta, \phi, \sigma_u^2, \sigma_\varepsilon^2\right) &= (0.0005, 1.20, 0.30, 0.00015, 0.200), \\
2.\ \left(a, \beta, \phi, \sigma_u^2, \sigma_\varepsilon^2\right) &= (0.0005, 1.10, 0.90, 0.00015, 0.025), \\
3.\ \left(a, \beta, \phi, \sigma_u^2, \sigma_\varepsilon^2\right) &= (0.0005, 1.00, 0.99, 0.00015, 0.0035), \\
4.\ \left(a, \beta, \phi, \sigma_u^2, \sigma_\varepsilon^2\right) &= (0.0005, 0.90, 0.10, 0.00035, 0.250).
\end{aligned}$$

The cases above, are representative of the corresponding ML estimates obtained in the empirical applications of the next section. The first and last parameter settings correspond to the case where the process $\{\beta_t\}$ has small persistence and is driven mainly by the noise component, whereas the second and third parameter settings correspond to the case in which the process $\{\beta_t\}$ is very close to being non-stationary. The results for the four cases, are reported in Tables 1 and 2. Table 1 contains the average bias, standard deviation, kurtosis and skewness coefficients of the corresponding ML estimators. The empirical sizes of the corresponding t-statistic for the null hypothesis $H_0 : \widehat{\mu} = \mu$, $\mu \in \{a, \beta, \phi, \sigma_u^2, \sigma_\varepsilon^2\}$, at the 5% significance level, are also presented. In addition, table 2 includes the empirical sizes of the well-known BDS

[3] The betas have been demeaned.

[4] This sample size was chosen as representative of the actual sample size for the empirical results that will be discussed in the next section.

test proposed by Brock, Dechert, Scheinkman and LeBaron (1996), for testing the hypothesis that the standardized residuals are *iid*. To calculate the BDS, we must specify the values of two parameters, the embedding dimension, m, and the distance parameter, ε/s, where s denotes the sample standard deviation. Brock, Hsieh and LeBaron (1991), suggest that ε should take values in the interval $[0.5, 1.5]$, and that m should be in line with the number of observations. Given the selected sample size $T = 1000$, and the fact that ε affects the power of the test, the reported results correspond to $m = 2, 3, 4, 5$ and $\varepsilon/s = 1$.

The results may be summarized as follows:

(i) The ML estimators of all the parameters in SFMT-AR work sufficiently well for all the parameter configurations under study, including those in which the autoregressive coefficient for $\{\beta_t\}$ is near unity (case 3). The average biases and standard deviations decrease as the sample size increases and the biases are sufficiently small even for $T = 250$. For example, for $T = 500$, the bias of $\widehat{\phi}$ is equal to -0.041, -0.062, -0.055 and -0.03 for cases 1, 2, 3 and 4, respectively. When the sample size increases to $T = 1000$, the corresponding bias decreases, in absolute terms, to -0.015, -0.015, -0.008 and -0.005, respectively.

(ii) The t-statistics corresponding to the parameters a, σ_u^2 and σ_ε^2 are properly sized, for all the four cases under consideration, even for sample sizes as small as $T = 250$. The t-statistics for $\widehat{\beta} = \beta$, appear to be over-sized, even for $T = 1000$, in the cases of strongly persistent betas, namely cases 2 and 3. Size distortions in both directions are reported for the t-statistics of $\widehat{\phi} = \phi$ for all the four cases under consideration except for case 2. This means that testing the hypothesis $\widehat{\phi} = \phi$ is, in general problematic unless $\{\beta_t\}$ exhibits strong (but not extremely strong) persistence. This is attributed to the small rate of convergence in the case of very high and very low persistence. For example, in case 3, when the number of observations increases to $T = 4000$, the size distortions become much smaller.[5]

[5] The empirical sizes for testing the hypotheses $\widehat{\phi} = \phi$ and $\widehat{\beta} = \beta$ become 3.88 and 8.06, respectively.

(iii) The empirical sizes of the BDS tests (reported in table 2), are in general close to their nominal values, especially for $m = 5$ and $T = 1000$, for all the cases under consideration.

3 Empirical Results

The empirical analysis of this paper is based on the S&P500 sector data (see, for example, DeMiguel, Garlappi and Uppal (2007), and Anderson, Brooks and Katsaris (2010) among others for similar datasets). We follow the Global Industry Classification Standard (GICS), designed and maintained by Standard & Poor's (S&P) and Morgan Stanley Capital International (MSCI), which consists of the following 10 sectors: 1: Consumer Discretionary, 2: Consumer Staples, 3: Energy, 4: Financials, 5: Healthcare, 6: Industrials, 7: Information Technology, 8: Materials, 9: Telecommunications and 10: Utilities. The dataset consists of weekly returns on the value-weighted indices of the aforementioned sectors, the returns on the S&P 500 Index (used as an approximation of the single factor, M_t) and the return of the 90-day T-bill, which is used as a proxy for the risk-free rate[6]. All the series are obtained from Bloomberg (except for the risk-free rate which was obtained from Ken French's Data Library website) and cover the period 22/9/1989 - 28/12/2012.

3.1 In-sample Comparisons

Using this dataset, we estimate the SFMT, SFMT-GARCH, SFMT-B- GARCH and SFMT-AR models. As already mentioned, SFMT is the simple homoscedastic model $\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} M_t + \boldsymbol{\nu}_t$, in which $\{\boldsymbol{\nu}_t\}$ is assumed to be a *niid* process with zero mean and finite variance-covariance matrix, $\Sigma_v = diag\{\sigma^2_{v_1}, ..., \sigma^2_{v_n}\}$. SFMT-GARCH is defined as follows:

[6]Following the tradition of the empirical literature (see for example, Fama and French 1996, Ng, Engle and Rothschild 1992) we employ excess rather than simple returns in the empirical analysis of this section. To avoid additional notational burden, we shall refrain from changing the relevant notation, which means that from now on $r_{i,t}$ will denote the excess return on asset i.

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} M_t + \mathbf{z}_t, \tag{13}$$

where $\mathbf{z}_t$ is a vector whose elements are $z_{i,t} = \sqrt{h_{i,t}}\varepsilon_{i,t}$, where $h_{i,t} = c_i + \gamma_i \varepsilon_{i,t-1}^2 + \delta_i h_{i,t-1}, i = 1, ..., n$ and $\{\varepsilon_{i,t}\}$ is *NIID(0,1)*. Note that since M_t is assumed to be the only source of correlation among the elements of $\mathbf{r}_t$, the conditional covariance matrix, $\Sigma_{t,t-1}^z$, of $\mathbf{z}_t$ is diagonal. Then, $\mathbf{\Sigma}_{t|t-1,(SFMT-GARCH)} = \mathbf{\Sigma}_{t,t-1}^z + \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}'$.

SFMT-B-GARCH is defined as follows:

$$\left[\begin{smallmatrix} r_{i,t} \\ M_t \end{smallmatrix}\right] = \left[\begin{smallmatrix} \mu_i \\ \mu_m \end{smallmatrix}\right] + \left[\begin{smallmatrix} u_{i,t} \\ u_{m,t} \end{smallmatrix}\right] = \boldsymbol{\mu}^i + \mathbf{u}_t^i, i = 1, ..., 10$$

where

$$\mathbf{u}_t^i = \mathbf{z}_t \mathbf{H}_{i,t}^{1/2}$$

and $\mathbf{z}_t$ is a 2-dimensional IID process with zero mean and the identity covariance matrix. We employ the constant correlation model[7] of Bollerslev (1990) to parameterize $\mathbf{H}_{i,t}^{1/2}$, and therefore,

$$\mathbf{H}_{i,t} = \begin{bmatrix} c_{ii} + \gamma_{ii}\varepsilon_{i,t-1}^2 + \delta_{ii} h_{ii,t-1} & \rho_{im}\sqrt{h_{ii,t}}\sqrt{h_{mm,t}} \\ \rho_{im}\sqrt{h_{ii,t}}\sqrt{h_{mm,t}} & c_{mm} + \gamma_{mm}\varepsilon_{m,t-1}^2 + \delta_{mm} h_{mm,t-1} \end{bmatrix}.$$

Note that, under SFMT-B-GARCH, in contrast to our approach up to now, we also model explicitly the conditional variance of M_t. Under the assumptions thus far, we may write $r_{i,t} = a_{i,t} + b_{i,t} M_t + u_{i,t}$, where $a_{i,t} = \mu_i - b_{i,t}\mu_m, b_{i,t} = \frac{h_{im,t}}{h_{mm,t}}$ and $u_{i,t}$ is a zero-mean process with variance equal to and $\tilde{h}_{ii,t} = h_{ii,t} - \frac{h_{im,t}^2}{h_{mm,t}}$. As a consequence, the conditional covariance matrix of the 10 sectors is given by:

$$\mathbf{\Sigma}_{t|t-1,(SFMT-B-GARCH)} = \mathbf{b}_t \mathbf{b}_t' h_{mm,t} + \Sigma_{t,t-1}^u,$$

where $\mathbf{b}_t' = [b_{1,t}, b_{2,t}, \ldots, b_{10,t}]'$ and $\Sigma_{t,t-1}^u = diag\left(\tilde{h}_{11,t}, \ldots, \tilde{h}_{1010,t}\right)$. Note that the

[7] Other methodologies, such as the diagonal BEKK or VECH, produce similar results.

time-varying betas can be re-written as:

$$b_{i,t} = \frac{h_{im,t}}{h_{mm,t}} = \rho_{im} \frac{\sqrt{h_{ii,t}}}{\sqrt{h_{mm,t}}}, i = 1, ..., 10.$$

The estimation results for SFMT, SFMT-GARCH, SFMT-B-GARCH and SFMT-AR are reported in Tables 3, 4, 5 and 6, respectively. These tables include also the results from the application of the BDS test on the standardized residuals of the aforementioned models.[8] An additional standard test for the presence of second-order effects in the standardized residuals is also reported.

Additional diagnostic tests, aiming at assessing the degree of time variation in the beta coefficient, are reported in Figures 1 and 2 (Appendix B). These Figures contain rolling estimates of the beta coefficient for all the ten sectors under consideration and for both models which assume a time-invariant beta, namely, for SFMT and SFMT-GARCH[9]. The overall results may be summarized as follows:

(i) In the context of the constant-beta homoscedastic SFMT model, the OLS estimates of beta from the ten sectors under consideration are quite disperse, ranging from 0.58 for the Utilities sector to 1.35 for the Financials one. However, there is strong evidence that this model is seriously misspecified. For all the ten sectors, the aforementioned test for higher-order temporal dependence rejects the hypothesis that the standardized residuals form an independent sequence. Furthermore, the rolling OLS estimates of beta, reported in Figure 1, leave no doubt that the constant beta assumption does not enjoy empirical support. Indeed, in some cases the time variation of betas is impressive. For example, in the case of Consumer Staples sector, the estimates of beta range from -0.08 to 1.26 for the estimation periods 9-March 2001 and 29-October 1993, respectively.

(ii) As far as SFMT-GARCH is concerned, the ML estimates of its parameters are broadly consistent with the ones reported in the empirical literature, that is, the

[8]The reported results correspond to the case where $\varepsilon/s = 1$ and $m = 3$. Results for the cases $m = 2, 4$ and 5 (not reported) provide similar conclusions.

[9]The figures contain rolling betas for window size of 50 observations (1 year).

sum of the GARCH coefficients $\widehat{\gamma}_i + \widehat{\delta}_i$ is close to unity with $\widehat{\delta}_i$ being much larger than $\widehat{\gamma}_i$. As expected, SFMT-GARCH performs far better than SFMT in terms of in-sample performance criteria. Specifically, the standardized residuals of SFMT-GARCH appear to be independent for all the sectors under consideration, with the possible exceptions of Financials and Industrials. However, additional misspecification testing reveals that SFMT-GARCH is not empirically adequate. Specifically, the rolling ML estimates of beta in the context of SFMT-GARCH, reported in Figure 3, suggest that substantial (or even massive) parameter instability is still present. This in turn implies that SFMT-GARCH does not capture adequately the exact form of CH exhibited by the returns generating process. In other words, the time variation of betas may be interpreted as evidence of important discrepancy between the type of second-order effects that truly characterize $\{\mathbf{r}_t\}$ and those implied by SFMT-GARCH.

(iii) The SFMT-B-GARCH also performs far better than SFMT in terms of in-sample performance criteria. However, the standardized residuals of SFMT-B-GARCH as opposed to those of SFMT-GARCH, do not appear to be temporally independent in general. On the other hand, SFMT-B-GARCH captures, to some extent, the time variation of betas. These results show that the two GARCH models produce different sets of empirical results with neither of them being clearly superior to the other.

(iv) Turning to the SFMT-AR model, the first thing to observe is the emergence of two distinct patterns of persistence for the beta process. Specifically, there is one group of sectors (HP) consisting of Consumer Staples, Energy, Healthcare, Information Technology and Materials for which the beta is highly persistent. The rest five sectors form another subset (LP) for which the beta persistence is low (Consumer Discretionary, Financials, Industrials) or even zero (Telecommunications, Utilities). As a result, HP exhibits strong second-order effects as opposed to LP in which dynamic heteroscedasticity is weak, if present at all. This varying degree of second-order effects within the set of returns series under consideration implied by SFMT-AR is in sharp contrast with the uniformity of volatility persistence impinged upon the aforementioned series by SFMT-GARCH. As far as empirical adequacy is concerned, SFMT-AR does not succeed in delivering independent standardized residuals in any of the ten series under study. This means that SFMT-AR does not fully capture the second-order effects of $\{\mathbf{r}_t\}$. Another interesting question would be to compare the time-variation of betas produced by SFMT-AR to that of SFMT-B-GARCH. Table 7 reports the correlations between the conditional betas from the two approaches, for each sector. These correlations suggest that betas differ between the two approaches and sometimes, this difference is substantial (see, for example, the negative correlation in the case of the Telecommunication sector betas). The overall assessment of the results on SFMT-GARCH, SFMT-B-GARCH and SFMT-AR seem to suggest that neither of these models provide an adequate characterization of the second-order dynamics of the returns generating process. As a result, the relevant question becomes that of which of these models comes closer to approximating the true CH exhibited by $\{\mathbf{r}_t\}$. This question may also be stated in the form: which of these models fares better in forecasting the next period's covariance matrix of returns? This question is addressed in the next sub-section.

(v) The estimated SFMT-GARCH conditional variance process differs radically from the SFMT-AR one, even in the cases where SFMT-AR delivers highly persistent processes. Table 8 reports the correlation coefficients between the two conditional variance processes, for the ten sectors under consideration. These coefficients are, in general, close to zero or even negative. More specifically, the estimates of the correlation coefficient range from -0.561 to 0.613 for Information Technology and Financials, respectively. It is interesting to note the strong negative correlation between the SFMT-GARCH and SFMT-AR conditional variance processes for the Information Technology sector, which is characterized by the most persistent beta process among all the sectors under consideration. These results imply that in spite of the fact that both SFMT-GARCH and SFMT-AR entail second-order, persistent effects, the exact types of CH implied by these models are quite different.

3.2 Out-of-Sample Comparisons

To take into account the possibility that, either M_t is not a good proxy of the market portfolio, or that this is not the only factor accounting for the observed correlations between $r_{i,t}$ and $r_{j,t}$, we also consider an additional model, hereafter referred to as SFMT-MGARCH, in which the errors of the ten factor models are jointly modelled as a multivariate GARCH process. More specifically, define $\mathbf{r}_t^{'}$ to be the (10×1) time-series vector $[r_{1,t}, r_{2,t}, \ldots, r_{10,t}]^{'}$. Consider a system of 10 conditional mean equations $\mathbf{r}_t = \mathbf{a} + \mathbf{b}M_t + \mathbf{u}_t$, where $\mathbf{u}_t = \mathbf{z}_t\mathbf{H}_t^{1/2}$ and $\mathbf{z}_t$ is a 10-dimensional IID process with zero mean and the identity covariance matrix. Again, we employ the constant correlation model to parametrize $\mathbf{H}_t$, and therefore,

$$h_{ij,t} = \begin{cases} c_{ii} + \gamma_{ii}\varepsilon_{i,t-1}^2 + \delta_{ii}h_{ii,t-1}, i = j \\ \rho_{ij}\sqrt{h_{ii,t}}\sqrt{h_{jj,t}}, i \neq j \end{cases}, i, j = 1, ..., 10.$$

Then, $\mathbf{\Sigma}_{t|t-1,(SFMT-MGARCH)} = \mathbf{bb}'\sigma_m^2 + \mathbf{H}_t$.

The out-of-sample comparisons are carried out as follows: First, we select an initial sample, referred to as the estimation sample, for which all the competing models, namely SFMT, SFMT-GARCH, SFMT-B-GARCH, SFMT-MGARCH, SFMT-AR, are estimated. Although various alternative estimation samples were tried and produced similar results, our reported results refer to the period 22/9/1989 - 30/12/2005. Second, using the estimated parameters, we produce one-step ahead forecasts of the conditional covariance matrix, $\mathbf{\Sigma}_{t|t-1}$, for each of the aforementioned models for the period 6/1/2006 - 28/12/2012, thus obtaining 365 one-week ahead forecasts. To remind the reader, the conditional covariance matrices implied by SFMT, SFMT-GARCH, SFMT-B-GARCH, SFMT-MGARCH and SFMT-AR are given by:

$$\begin{aligned}
\boldsymbol{\Sigma}_{t|t-1,(SFMT)} &= \boldsymbol{\Sigma}_v + \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}', \\
\boldsymbol{\Sigma}_{t|t-1,(SFMT-GARCH)} &= \boldsymbol{\Sigma}_{t,t-1}^z + \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}', \\
\boldsymbol{\Sigma}_{t|t-1,(SFMT-BGARCH)} &= \Sigma_{t,t-1}^u + h_{mm,t}\mathbf{b}_t\mathbf{b}_t', \\
\boldsymbol{\Sigma}_{t|t-1,(SFMT-MGARCH)} &= \mathbf{H}_t + \sigma_m^2 \mathbf{b}\mathbf{b}', \\
\boldsymbol{\Sigma}_{t|t-1,(SFMT-AR)} &= \sigma_m^2 \boldsymbol{\Sigma}_\varepsilon + \boldsymbol{\Sigma}_u + \sigma_m^2 \left(\boldsymbol{\beta} + \boldsymbol{\beta}_{t/t-1}\right)\left(\boldsymbol{\beta} + \boldsymbol{\beta}_{t/t-1}\right)'.
\end{aligned}$$

Using these matrices we calculate the global minimum variance portfolios together with the corresponding realized portfolio returns. The reason for selecting the global minimum variance portfolio is to minimize the estimation errors relating to the estimation of the expected returns. This procedure results in 365 out-of-sample realized portfolio returns for each model. For comparison purposes, apart from the SFMT, SFMT-GARCH, SFMT-B-GARCH, SFMT-MGARCH and SFMT-AR portfolios, we also calculate the portfolio returns that correspond to the case in which in $\widehat{\boldsymbol{\Sigma}}_{t|t-1}$ is the sample covariance matrix (SCM) and also for the case in which the portfolio is formed according to the naive 1/n ($1/N$) strategy. To assess the out-of-sample performance of each strategy, we employ the following three criteria: (i) the out-of-sample Sharpe ratio ($SR = \frac{\mu_i}{\sigma_i}$), (ii) the Certainty-Equivalent Return ($CEQ = \mu_i - \frac{\gamma}{2}\sigma_i$), where γ is the risk-aversion coefficient, and (iii) the out-of-sample Treynor ratio ($TR = \frac{\mu_i}{\beta_i}$), where β_i is the portfolio's beta relative to the market portfolio. Following common practice, the CEQ return is defined to be the risk-free rate that an investor is willing to accept in order to be indifferent between choosing this riskless return and the return of the strategy. CEQ is calculated for various values of γ, with largely similar results (the reported ones correspond to $\gamma = 1$).

The results, reported in Table 9, may be summarized as follows:

(i) The SFMT-AR strategy dominates all the other strategies under any of the performance criteria mentioned above, although differences are marginal. For example, the SR of SFMT-AR is greater than that of SFMT-GARCH, SFMT-B-GARCH, SFMT-MGARCH, SFMT, SCM and $1/N$ by 6.24%, 105.50%, 5.00%, 11.77%, 14.58% and 456.48%, respectively. It is worth noting that all the statistical methods dominate the naive $1/N$ strategy by a wide margin. This piece of evidence runs counter to the view expressed in De Miquel, Garlappi and Uppal (2007) according to which no statistical method for forecasting the returns covariance matrix offers significant diversification gains over the $1/N$ strategy.

(ii) The SFMT-GARCH strategy comes second to SFMT-AR, offering some minor gains over the homoscedastic SFMT and the non-parametric SCM ones, but very significant gains over the naive $1/N$ strategy. It is also worth noting the exceptionally poor performance of SFMT-B-GARCH, which appears to be superior only to that of $1/N$ strategy.

Finally, it would be interesting to examine the differences between the forecasted covariance matrices produced by the two best performing models, namely SFMT-AR and SFMT-GARCH. To this end, we define a distance, $d_{AR-GARCH}$ between $\mathbf{\Sigma}_{t|t-1,(SFMT-AR)}$ and $\mathbf{\Sigma}_{t|t-1,(SFMT-GARCH)}$ and examine how this difference has evolved over the forecast period under consideration. Foerstner and Moonen (1999) define the distance between two symmetric semi-positive definite matrices as the sum of the squared logarithms of the properly defined eigenvalues, that is:

$$d_{AR-GARCH} = \sqrt{\sum_{i=1}^{n} \ln(\lambda_i(\mathbf{\Sigma}_{t|t-1,(SFMT-AR)}, \mathbf{\Sigma}_{t|t-1,(SFMT-GARCH)}))^2}$$

with the eigenvalues $\lambda_i(\mathbf{\Sigma}_{t|t-1,(SFMT-AR)}, \Sigma_{t|t-1,(SFMT-GARCH)})$, $i = 1, ..., n$ obtained from the solution of $\mid \lambda\mathbf{\Sigma}_{t|t-1,(SFMT-AR)} - \mathbf{\Sigma}_{t|t-1,(SFMT-GARCH)} \mid = 0$. The time evolution of $d_{AR-GARCH}$, presented in Figure 3 (Appendix B), suggests first that this distance ranges from 1.08 on 25/01/2008 to 4.05 on 19/12/2008. It also suggests that $d_{AR-GARCH}$ increased rapidly during the period of the recent financial crisis, returning to more normal levels after the first quarter of 2009. We obtain similar results when we use the Frobenius norm to calculate the distance, between $\mathbf{\Sigma}_{t|t-1,(SFMT-AR)}$ and $\mathbf{\Sigma}_{t|t-1,(SFMT-GARCH)}$, (see Figure 4 in Appendix B). Specifically, the results show a significant increase of the difference between the conditional covariance matrices produced by SFMT-AR and SFMT-GARCH during the one-year period that starts approximately at the bankruptcy date of Lehman Brothers (9/15/2008). Motivated by this observation, we examine the out-of sample performance of the models under consideration for the period 9/2008 - 8/2009. Because the annualized returns are negative for this period, the SR and TR statistics are not appropriate measures for comparing the models.[10]

Table 10 presents the results concerning CEQ, as well as the annualized returns and risk for each model. We observe that the strategy implied by $SFMT-AR$ combines the smallest (in absolute values) negative return with the lowest annualized risk. This results to a CEQ which is at least 11,85% higher than the corresponding value of the second best model (which, in terms of CEQ, is $SFMT-MGARCH$). Note that the calculation of CEQ in Table 10 retains the value of γ equal to 1. On the other hand, it is worth noting that during the period that followed the bankruptcy of Lehman Brothers, the risk aversion increased. This fact in combination with the best performance of $SFMT-AR$ in terms of both annualized return and risk, implies that the difference between the CEQ of $SFMT-AR$ and the CEQ of the second best model is actually bigger for this period.[11]

[10]For example, if two models produce comparable (in magnitude) annualized returns but the annualized standard deviation (beta) of the first is larger, the Sharpe (Treynor) ratio of the first model becomes smaller (less negative) than that of the second one.

[11]A natural extension of our empirical analysis would be to examine whether the combination of autoregressive betas with GARCH errors would yield better out of sample performance. To this end we repeated the out of sample study for the specific model ($SFMT-ARG$). The results, however, were not satisfactory. Specifically, $SFMT-ARG$ outperforms only $SFMT-B-GARCH$ and the equally weighted portfolio.

4 Conclusions

This paper has examined the in-sample and out-of-sample performance of several variants of the single factor model for stock returns. Attention was focused mainly on SFMT-GARCH in the context of which the idiosyncratic risk is conditionally heteroscedastic and SFMT-AR which assumes an autoregressive structure for the systematic risk and homoscedasticity for the non-systematic one. A large part of the paper dealt with the theoretical properties as well as the estimation issues of SFMT-AR. It was proved that SFMT-AR is capable of reproducing the most important stylized facts of individual stock returns, namely conditional heteroscedasticity and leptokurtosis. Interestingly enough, this result continues to hold even in the case in which the stochastic process generating the "unanticipated changes of the factor" is independent. The empirical results showed that none of the factor models under examination is fully empirically adequate, in terms of in-sample criteria. For example, SFMT-GARCH still suffers from substantial beta variation whereas SFMT-AR does not account fully for conditional heteroscedasticity.

However, these models offer significant gains for forecasting next period's covariance matrix of returns over the homoscedastic SFMT and the non-parametric method of forecasting second moments via their sample analogues. Moreover, these gains are maximized relative to the naive (1/N) allocation strategy, which in some recent studies was found to deliver the greatest portfolio diversification gains among a set of strategies that include various statistical methods (see, e.g. De Miguel et. al. 2007). Among the four conditionally heteroscedastic models under consideration, namely SFMT-AR, SFMT-GARCH, SFMT-B-GARCH and SFMT-MGARCH, the former was found to exhibit systematically the best out-of-sample performance, closely followed by SFMT-GARCH.

References

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003, Modeling and Forecasting Realized Volatility, *Econometrica* 71, 579-625.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold and Jin Wu, 2005, A Framework for Exploring the Macroeconomic Determinants of Systematic Risk, *American Economic Review* 95, 2, 398-404.

Anderson K., C. Brooks, A. Katsaris, 2010, Speculative bubbles in the S&P 500: Was the tech bubble confined to the tech sector?, *Journal of Empirical Finance* 17, 345–361.

Ang, Andrew, and Joseph Chen, 2007, CAPM over the long run: 1926-2001, *Journal of Empirical Finance* 14, 1-40.

Arditti, Fred D., 1967 Risk and the Required Return on Equity, *Journal of Finance* 22, 19-36.

Avramov, Doron, and Tarun Chordia, 2006, Asset Pricing Models and Financial Market Anomalies, *Review of Financial Studies* 19, 1001-1040.

Barndorff-Nielsen, Ole E., and Neil Shephard, 1998, Aggregation and model construction for volatility models, *Centre for Analytical Finance, University of Aarhus, Working Paper Series* No. 10.

Bali, Turan G., Hengyong Mo and Yi Tang, 2008, The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR, *Journal of Banking and Finance* 32, 269-282.

Beja, Avraham, 1972, On Systematic and Unsystematic Components of Financial Risk, *Journal of Finance* 27, 37-45.

Berk, Jonathan B., Richard C. Green and Vasant Naik, 1999, Optimal Investment, Growth Options, and Security Returns, *Journal of Finance* 54, 1553-1607.

Blattberg, Robert C., and Nicholas J. Gonedes, 1974, A Comparison of the Stable and Student Distributions as Statistical Models for Asset Prices, *Journal ofBusiness* 47, 244-280.

Blume, Marshall E., 1971, On the assessment of risk, *Journal of Finance* 26, 1-10.

——————, 1975, Betas and Their Regression Tendencies, *Journal of Finance* 30, 785-795.

Bollerslev, Tim, Robert F. Engle and Jeffrey M. Wooldridge, 1988, A Capital Asset Pricing Model with Time Varying Covariances, *Journal of Political Economy* 96, 116-131.

Bos, Theodore, and Paul Newbold, 1984, An emprical investigation of the possibility of stochastic systematic risk in the market model, *Journal of Business* 57, 35-41.

Bos, Theodore, and Thomas A. Fetherston, 1992, Market model nonstationarity in the Korean Stock Market, *Pacific Basin Capital Market Research*, 3.

————————————, 1995, Nonstationarity of the market model, outliers, and the choice of market rate of return, in Theodore Bos, Thomas A. Fetherston, eds.: *Advances in Pacific Basin Financial Markets, Vol. 1* (JAI Press, Greenwich, CT).

Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B., 1996, A test for independence based on the correlation dimension. *Econometrics Review* 15, 197–235.

Brock, William A., Hsieh, David A. and LeBaron Blake, 1991, Nonlinear dynamics, chaos and instability. The MIT Press, Cambridge, Massachusetts.

Castanias, Richard P. II, 1979, Macroinformation and the Variability of Stock Market Prices, *Journal of Finance* 34, 439-450.

Clark, Peter K., 1973, A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica* 41, 135-155.

———, 1987, *The Cyclical Component of U.S. Economic Activity, Quarterly Journal of Economics* 102, 797-814.

Collins, Daniel W., Johannes Ledolter and Judy Rayburn, 1987, Some Further Evidence on the Stochastic Properties of Systematic Risk, *Journal of Business* 60, 425-448.

DeMiguel V, Garlappi L, Uppal R., 2009, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies.* 22:1915–53

Fabozzi, Frank J., and Jack Clark Francis, 1977, Stability Tests for Alphas and Betas Over Bull and Bear Market Conditions, *Journal of Finance* 32, 1093-1099.

Faff, Robert W., John H. H. Lee and Tim R. L. Fry, 1992. Time stationarity of systematic risk: some Australian evidence. Journal of Business Finance & Accounting, 19, 253-270.

———, 1978. Beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, 13, 101-116.

Fama, Eugene F., 1965. The behavior of stock-market prices. *Journal of Business*, 38, 34-105.

———, 1968. Risk, Return and Equilibrium: Some Clarifying Comments. *Journal of Finance*, 23, 29-40.

Fama, Eugene F., and Kenneth R. French, 1988. Permanent and Temporary Components of Stock Prices. *Journal of Political Economy*, 96, 246-273.

———, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.

Fisher, Lawrence, and Jules H. Kamin, 1985, Forecasting Systematic Risk: Estimates of 'Raw' Beta That Take Into Account the Tendency of Beta to Change and the Heteroscedasticity of Residual Returns, *Journal of Financial and Quantitative Analysis* 20, 127-149.

Förstner, Wolfgang and Moonen Boudewijn, 1999, A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University.

Granger, Clive W. J., 1968, Some Aspects of the Random walk Model of Stock Market Prices, *International Economic Review* 9, 253-257.

Hagerman, Robert L., 1978, More Evidence on the Distribution of Security Returns, *Journal of Finance* 33, 1213-1221.

Harvey, Campbell R., 1989, Forecasts of Economic Growth from the Bond and Stock Markets, *Financial Analysts Journal* 45, 38-45.

Harvey, Campbell R., and Akhtar Siddique, 1999, Autoregressive Conditional Skewness, *Journal of Financial and Quantitative Analysis* 34, 465-487.

——————————, 2000a, Conditional Skewness in Asset Pricing Tests, *Journal of Finance* 55, 1263-1295.

——————————, 2000b, Time Varying Conditional Skewness and the Market Risk Premium, *Research in Banking and Finance* 1, 25-58.

Hsu, Der-Ann, Robert B. Miller and Dean W. Wichern, 1974, On the stable Paretian behavior of stock-market prices, *Journal of the American Statistical Association* 69, 108-113.

Jostova, Gergana, and Alexander Philipov, 2005, Bayesian Analysis of Stochastic Betas, *Journal of Financial and Quantitative Analysis* 40, 747-778.

Kendall, Maurice G., 1953, The Analysis of Economic Time-Series-Part I: Prices, *Journal of the Royal Statistical Society. Series A* 116, 11-34.

Kim, Myung J., Charles R. Nelson and Richard Startz, 1991, Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence, *Review of Economic Studies* 58, 515-528.

Kim, Chang-Jin., and Charles R. Nelson, 1999, State-Space Models with Regime Switching, MIT Press, London, England and Cambridge, Massachusetts.

Kon, Stanley J., 1984, Models of Stock Returns–A Comparison, *Journal of Finance* 39, 147-165.

Kon, Stanley J., and Frank C. Jen, 1978, Estimation of Time-Varying Systematic Risk and Performance for Mutual Fund Portfolios: An Application of Switching Regression, *Journal of Finance* 33, 457-475.

Kraus, Alan, and Robert H. Litzenberger, 1976, Skewness Preference and the Valuation of Risk Assets, *Journal of Finance* 31, 1085-1100.

Levenberg K., 1944, A method for the solution of certain problems in least squares, *Quarterly of Applied Mathematics*, Vol. 2, pp. 164-168.

Lim, Kian-Guan., 1989, A New Test of the Three-Moment Capital Asset Pricing Model, *Journal of Financial and Quantitative Analysis* 24, 205-216.

Lo, Andrew W., and A. Craig MacKinlay, 1988, Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies* 1, 41-66.

Madan, Dilip B., and Eugene Seneta, 1990, The Variance Gamma (V.G.) Model for Share Market Returns, *Journal of Business* 63, 511-524.

Mandelbrot, Benoit, 1963, The Variation of Certain Speculative Prices, *Journal of Business* 36, 394-419.

Mandelbrot, Benoit, and Howard M. Taylor, 1967, On the Distribution of Stock Price Differences, *Operations Research* 15, 1057-1062.

Mankiw, N. Gregory, David Romer and Matthew D. Shapiro, 1991, Stock Market Forecastability and Volatility: A Statistical Appraisal, *Review of Economic Studies* 58, 455-477.

McQueen, Grant, 1992, Long-Horizon Mean-Reverting Stock Prices Revisited, *Journal of Financial and Quantitative Analysis* 27, 1-18.

Mergner, Sascha and Jan Bulla, 2008, Time-varying Beta Risk of Pan-European Industry Portfolios: A Comparison of Alternative Modeling Techniques, *The European Journal of Finance*, vol. 14, no. 8, pp. 771–802.

Merton, R. C, 1973, "An Intertemporal Capital Asset Pricing Model." *Econometrica* 41 (5): 867–87.

Miller, Merton, and Myron Scholes, 1972, Rates of return in relation to risk: A reexamination of some recent findings, in Michael Jensen, ed.: *Studies in the theory of capital markets* (Praeger, New York).

Ng, V.K., R.F. Engle and M. Rothschild, 1992, "A Multi-Dynamic-Factor Model for Stock Returns," *Journal of Econometrics*, 52, 245-266.

Officer, Robert R., 1972, The Distribution of Stock Returns, *Journal of the American Statistical Association* 67, 807-812.

Ohlson, James, and Barr Rosenberg, 1982, Systematic Risk of the CRSP Equal-Weighted Common Stock Index: A History Estimated by Stochastic-Parameter Regression, *Journal of Business* 55, 121-145.

Osborne, Matthew F. M., 1959, Brownian Motion in the Stock Market, *Operations Research* 7, 145-173.

Pagan, A., 1980, "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients", *Journal of Econometrics* 13, pp.341-63.

Peligrad, Magda, and Sergey Utev, 2005, A new maximal inequality and invariance principle for stationary sequences, *The Annals of Probability* 33, 798-815.

Petkova, Ralitsa, and Lu Zhang, 2005, Is value riskier than growth? *Journal of Financial Economics* 78, 187-202.

Poterba, James M., and Lawrence H. Summers, 1988, Mean reversion in stock prices : Evidence and Implications, *Journal of Financial Economics* 22, 27-59.

Praetz, Peter D., 1972, The Distribution of Share Price Changes, *Journal of Business* 45, 49-55.

Richardson, Matthew, 1989, *Predictability of stock returns: statistical theory and evidence.* Ph.D. Thesis, Stanford University, Graduate School of Business.

Richardson, Matthew, and and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323-348.

Ross, Stephen, 1976, "The arbitrage theory of capital asset pricing". *Journal of Economic Theory* 13 (3): 341–360.

Schwartz, Robert A., and David K. Whitcomb, 1977, The Time-Variance Relationship: Evidence on Autocorrelation in Common Stock Returns, *Journal of Finance* 32, 41-55.

Shanken, J., 1990, Intertemporal asset pricing: an empirical investigation, *Journal of Econometrics* 45, 99–120.

Stout, William F., 1974. *Almost Sure Convergence* (Academic Press, New York, NY).

Sunder, Shyam, 1980, Stationarity of Market Risk: Random Coefficients Tests for Individual Stocks, *Journal of Finance* 35, 883-896.

Upton, David E., and Donald S. Shannon, 1979, The Stable Paretian Distribution, Subordinated Stochastic Processes, and Asymptotic Lognormality: An Empirical Investigation, *Journal of Finance* 34, 1031-1039.

Zhang, Lu, 2005, The value premium, *Journal of Finance* 60, 67-103.

APPENDIX A

Proof of Theorem 1:

(a) For notational simplicity, we drop the subscript i. First note that

$$
\begin{aligned}
(Var(r_t))^2 = & \left(\left(\beta^2+\sigma_\beta^2\right)\sigma_m^2+\mu^2\sigma_\beta^2+\sigma_u^2\right)^2 = \sigma_u^4+\mu^4\sigma_\beta^4+\beta^4\sigma_m^4+\sigma_m^4\sigma_\beta^4+ \\
& +2\left(\mu^2\sigma_m^2\sigma_\beta^4+\mu^2\sigma_u^2\sigma_\beta^2+\beta^2\sigma_m^2\sigma_u^2+\beta^2\sigma_m^4\sigma_\beta^2+\sigma_m^2\sigma_u^2\sigma_\beta^2+\mu^2\beta^2\sigma_m^2\sigma_\beta^2\right)
\end{aligned} \tag{14}
$$

Moreover, for the fourth central moment of r_t we have

$$
\begin{aligned}
E\left[(r_t-E[r_t])^4\right] &= E\left[(\beta w_t+\beta_t(\mu+w_t)+u_t)^4\right] \\
&= E\left[u_t^4+\mu^4\beta_t^4+w_t^4\beta^4+w_t^4\beta_t^4+6\mu^2u_t^2\beta_t^2+6\mu^2w_t^2\beta_t^4\right. \\
&\quad \left.+6u_t^2w_t^2\beta^2+6u_t^2w_t^2\beta_t^2+6w_t^4\beta^2\beta_t^2+6\mu^2w_t^2\beta^2\beta_t^2\right] \\
&= 3\left[\left(\sigma_u^4+\mu^4\sigma_\beta^4+\beta^4\sigma_m^4+\sigma_\beta^4\sigma_m^4\right)+2\left(\mu^2\sigma_u^2\sigma_\beta^2+3\mu^2\sigma_m^2\sigma_\beta^4\right.\right. \\
&\quad \left.\left.+\beta^2\sigma_u^2\sigma_m^2+\sigma_\beta^2\sigma_u^2\sigma_m^2+3\beta^2\sigma_\beta^2\sigma_m^4+\mu^2\beta^2\sigma_\beta^2\sigma_m^2\right)\right] \\
&= 3Var^2(r_t)+12\left(\mu^2\sigma_m^2\sigma_\beta^4+\beta^2\sigma_\beta^2\sigma_m^4\right)\ ,
\end{aligned} \tag{15}
$$

where we have used the fact that for the Gaussian distributions, the third moment is zero and fourth moment equals to three times the square of the second. Hence, the kurtosis coefficient of the unconditional distribution of stock returns is given by

$$
Kurt(r_t)=\frac{E\left[(r_t-E[r_t])^4\right]}{Var^2(r_t)}=3+\frac{12\left(\mu^2\sigma_m^2\sigma_\beta^4+\beta^2\sigma_\beta^2\sigma_m^4\right)}{Var^2(r_t)}. \tag{16}
$$

APPENDIX B

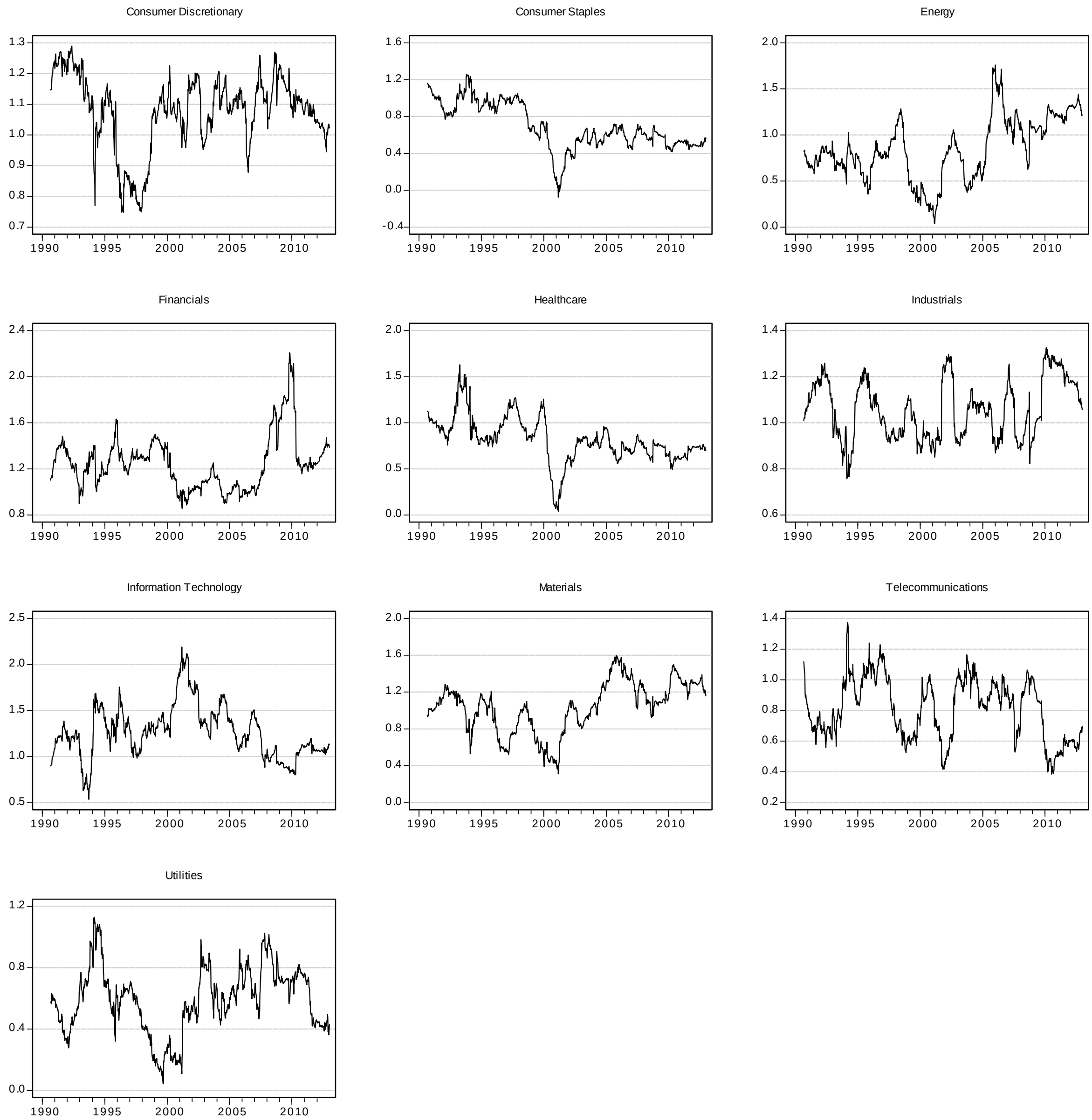

Figure 1: Rolling estimates for SFMT betas

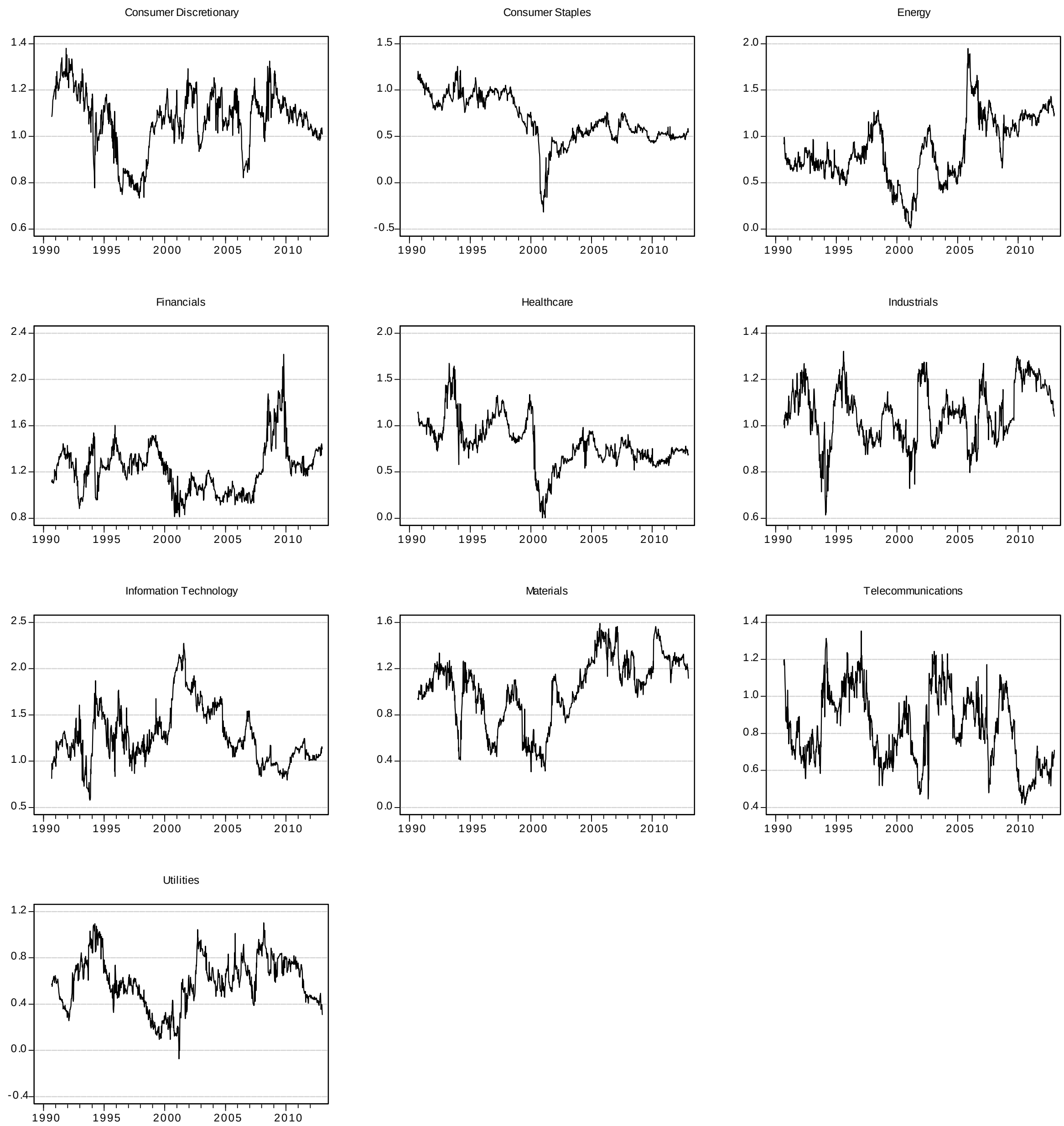

Figure 2: Rolling estimates for SFMT-GARCH betas

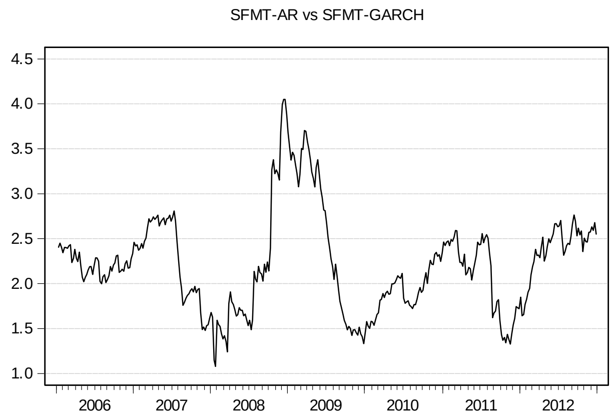

Figure 3: The time evolution of $d_{AR-GARCH}$.

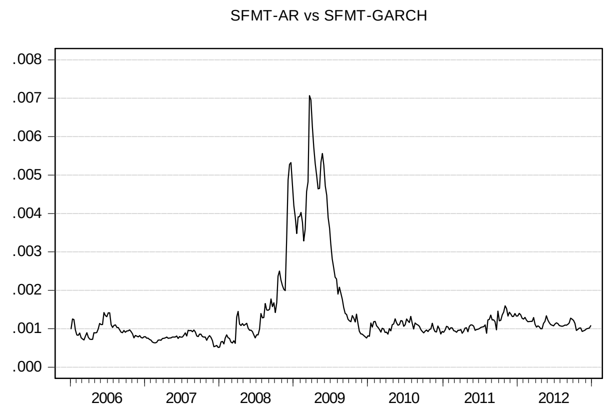

Figure 4: The time evolution of the difference in the covariance matrices, using the Frobenius norm.

Table 1: Monte-Carlo results

	Panel A: Persistent betas				
	α	β	$ln(\sigma_u^2)$	$ln(\sigma_\epsilon^2)$	ϕ
Case 1					
bias	0.00	−0.01	−0.62	0.39	−1.49
std	0.00	0.59	0.53	4.90	0.49
kurt	3.02	3.00	3.00	3.80	12.84
skew	0.02	0.01	−0.08	−0.41	−1.97
size(%)	4.54	7.26	5.62	5.84	5.14
Case 3					
bias	0.00	0.12	−0.52	7.57	−0.84
std	0.00	1.66	0.47	4.93	0.20
kurt	3.00	2.96	2.92	4.75	822.11
skew	−0.01	−0.04	−0.13	−0.20	−22.09
size(%)	5.30	16.46	4.66	4.58	0.84
	Panel B: Non-persistent betas				
	α	β	$ln(\sigma_u^2)$	$ln(\sigma_\epsilon^2)$	ϕ
Case 1					
bias	0.00	−0.01	−0.38	−5.01	−1.50
std	0.00	0.37	0.63	2.75	1.67
kurt	3.01	2.96	2.95	8.37	4.08
skew	0.03	−0.02	−0.07	−1.31	−0.41
size(%)	4.48	5.06	4.84	4.02	7.08
Case 4					
bias	0.00	0.04	−0.20	−23.70	−0.48
std	0.01	0.49	0.61	6.44	3.19
kurt	3.06	2.94	3.00	18.72	3.34
skew	0.01	−0.01	−0.01	−2.87	−0.10
size(%)	5.46	4.66	5.44	4.40	13.72

Note: Size denotes the empirical size of the corresponding t-statistic for the null hypothesis discussed above. Bias and size values are $\times 10^2$ and std $\times 10^1$.

Table 2: (%) empirical sizes for the BDS test

	m			
	2	3	4	5
Case 1	3.52	3.86	4.12	4.96
Case 2	3.90	4.46	5.02	5.94
Case 3	3.56	3.58	4.08	5.20
Case 4	2.94	3.46	4.02	4.60

Table 3: SFMT in-sample results

	Panel A: Coefficients									
	1	2	3	4	5	6	7	8	9	10
α	1.09	5.34	4.24	−4.84	5.54	0.026	2.52	−2.65	−6.66	−4.59
β	1.09***	0.61***	0.87***	1.35***	0.76***	1.05***	1.21***	1.00***	0.80***	0.57***
	Panel B: Statistics									
$Arch$	28.54***	84.11***	41.39***	198.76***	7.15***	100.16***	63.54***	22.52***	26.52***	35.76***
$BDS(3)$	8.24***	11.07***	5.95***	12.78***	10.37***	7.87***	11.12***	9.59***	4.20***	10.22***
$skew$	0.37	−0.37	−0.18	0.52	−0.25	0.41	−0.48	0.48	−0.17	−0.40
$kurt$	6.46	10.39	4.84	15.17	7.07	9.63	5.45	8.69	7.60	7.94

Notes: α values are $\times 10^4$. 1: Consumer Discretionary, 2: Consumer Staples, 3: Energy, 4: Financials, 5: Healthcare, 6: Industrials, 7: Information Technology, 8: Materials, 9: Telecommunications and 10: Utilities. * denotes statistical significance at 10%; ** denotes statistical significance at 5%; *** denotes statistical significance at 1%.

Table 4: SFMT-GARCH in-sample results

	Panel A: Coefficients									
	1	2	3	4	5	6	7	8	9	10
α	−0.00	6.46**	1.76	−1.47	7.13*	0.09	0.08	−4.13	−0.08	−0.16
β	1.07***	0.63***	0.94***	1.19***	0.78***	1.05***	1.11***	1.08***	0.77***	0.55***
c	3.78***	3.18***	7.20***	3.76***	3.66***	1.45***	3.37**	5.11***	3.58***	5.49***
γ	0.08***	0.10***	0.06***	0.11***	0.08***	0.08***	0.08***	0.06***	0.05***	0.07***
δ	0.89***	0.89***	0.92***	0.88***	0.91***	0.92***	0.91***	0.92***	0.95***	0.91***
	Panel B: Statistics									
$Arch$	0.09	1.08	0.63	4.82**	0.03	10.67***	0.06	2.41	0.19	0.19
$BDS(3)$	0.83	0.22	−1.85*	1.94*	0.90	0.60	−0.33	0.40	−0.42	1.50
$skew$	−0.14	−0.06	−0.11	0.06	−0.02	0.29	−0.33	−0.02	−0.09	−0.27
$kurt$	4.28	4.00	3.72	3.68	5.08	5.67	3.94	5.23	4.62	4.67

Notes: α values are $\times 10^4$ and c is $\times 10^6$. 1: Consumer Discretionary, 2: Consumer Staples, 3: Energy, 4: Financials, 5: Healthcare, 6: Industrials, 7: Information Technology, 8: Materials, 9: Telecommunications and 10: Utilities. * denotes statistical significance at 10%; ** denotes statistical significance at 5%; *** denotes statistical significance at 1%.

Table 5: SFMT-B-GARCH in-sample results

	Panel A: Coefficients									
	1	2	3	4	5	6	7	8	9	10
α	1.67***	1.66***	1.12***	1.45***	1.59***	1.36***	1.19***	1.06***	1.55***	1.16***
c	4.18***	0.57***	2.45***	1.71***	0.44***	0.99***	3.30***	1.40***	1.08***	2.06***
γ	0.11***	0.08***	0.08***	0.10***	0.07***	0.06***	0.08***	0.05***	0.07***	0.16***
δ	0.82***	0.91***	0.89***	0.89***	0.93***	0.93***	0.89***	0.93***	0.91***	0.81***
	Panel B: Statistics									
$Arch$	0.00	1.72	2.08	4.20**	0.37	0.77	0.06	0.03	0.00	0.01
$BDS(3)$	1.65*	2.46**	−0.53	1.63	2.09**	1.73*	1.23	0.67	−0.95	0.56
$skew$	−0.09	−0.57	−0.36	0.02	−0.55	−0.38	−0.49	−0.08	−0.20	−0.49
$kurt$	3.76	7.38	4.05	3.33	7.12	6.45	4.34	5.57	4.86	5.14

Notes: α values are $\times 10^3$ and c is $\times 10^5$. 1: Consumer Discretionary, 2: Consumer Staples, 3: Energy, 4: Financials, 5: Healthcare, 6: Industrials, 7: Information Technology, 8: Materials, 9: Telecommunications and 10: Utilities. * denotes statistical significance at 10%; ** denotes statistical significance at 5%; *** denotes statistical significance at 1%.

Table 6: SFMT-AR in-sample results

	Panel A: Coefficients									
	1	2	3	4	5	6	7	8	9	10
α	1.34	3.47	4.54	−8.36*	4.00	0.03	3.51	−3.51	−5.42	−2.06
σ_u^2	1.01***	1.82***	4.49***	2.00***	2.87***	0.85***	4.50***	3.52***	3.65***	2.57***
σ_ε^2	7.34***	0.78***	0.61***	24.13***	0.34***	7.52***	2.32***	0.19***	12.50***	20.46***
β	1.08***	0.68***	0.86***	1.28***	0.81***	1.05***	1.27***	1.02***	0.80***	0.57***
ϕ	0.20*	0.96***	0.98***	0.30***	0.97***	0.22**	0.92***	0.99***	−0.11	0.08
	Panel B: Statistics									
$Arch$	8.33***	44.12***	37.70***	48.03***	4..90**	31.18***	62.16***	10.52***	21.24***	23.59***
$BDS(3)$	6.47***	9.23***	5.45***	9.51***	8.74***	5.98***	10.11***	8.12***	2.79***	7.86***
$skew$	0.39	−0.77	−0.14	0.31	−0.29	0.26	−0.49	0.41	−0.38	−0.21
$kurt$	6.33	13.05	4.55	6.26	7.60	6.08	5.26	8.42	8.15	7.33

Notes: α and σ_u^2 values are $\times 10^4$ and σ_ϵ^2 is $\times 10^2$. 1: Consumer Discretionary, 2: Consumer Staples, 3: Energy, 4: Financials, 5: Healthcare, 6: Industrials, 7: Information Technology, 8: Materials, 9: Telecommunications and 10: Utilities. * denotes statistical significance at 10%; ** denotes statistical significance at 5%; *** denotes statistical significance at 1%.

Table 7: Correlation coefficients between SFMT-AR and SFMT-B-GARCH conditional betas

1	2	3	4	5	6	7	8	9	10
0.177	0.579	0.267	0.333	0.392	0.145	0.729	0.537	-0.128	0.157

Note: 1: Energy, 2: Material, 3: Industrials, 4: Consumer Discretionary, 5: Consumer Staples, 6: Healthcare, 7: Financials, 8: Information Technology, 9: Telecommunications and 10: Utilities.

Table 8: Correlation coefficients between SFMT-AR and SFMT-GARCH conditional variances

1	2	3	4	5	6	7	8	9	10
0.066	-0.230	-0.265	0.340	-0.248	0.034	0.678	-0.515	0.069	0.190

Note: 1: Energy, 2: Material, 3: Industrials, 4: Consumer Discretionary, 5: Consumer Staples, 6: Healthcare, 7: Financials, 8: Information Technology, 9: Telecommunications and 10: Utilities.

Table 9: Out-of-sample results (full sample)

Models	SR	CEQ(%)	TR	Ann. Ret.(%)	Ann. Risk(%)
$SFMT-AR$	0.138	0.95	0.033	2.10	15.16
$SFMT-GARCH$	0.130	0.80	0.031	2.10	16.10
$SFMT-B-GARCH$	0.067	-0.13	0.016	1.02	15.12
$SFMT-MGARCH$	0.132	0.82	0.031	2.16	16.40
$SFMT$	0.124	0.70	0.029	1.97	15.94
$Sample$	0.121	0.65	0.028	1.92	15.94
$1/n$	0.025	-1.59	0.010	0.51	20.50

Table 10: Out-of-sample results (9/2008-8/2009)

$Models$	$CEQ(\%)$	$Ann.Ret.(\%)$	$Ann.Risk(\%)$
$SFMT-AR$	-20.68	-16.40	29.27
$SFMT-GARCH$	-24.37	-19.12	32.40
$SFMT-B-GARCH$	-27.15	-22.64	30.02
$SFMT-MGARCH$	-23.46	-17.90	33.35
$SFMT$	-24.04	-19.20	31.13
$Sample$	-23.84	-18.98	31.17
$1/N$	-29.17	-21.29	39.69