Essays in Financial Econometrics

by Antypas J. Antonios

Department of Banking and Financial Management
University of Piraeus

Piraeus 2010

To my parents Vassiliki and Nakis

...Ithaca gave you the marvellous journey.
Without her you wouldn't have set out.
She has nothing left to give you now.
And if you find her poor, Ithaca won't have fooled you.
Wise as you will have become, so full of experience,
you'll have understood by then what these Ithakas mean.
Ithaka, (1911) K. Kavafis

Acknowledgements

Writing this thesis has been a long journey and a great many people have contributed to reaching its final destination. This is the best time and place to take the opportunity and thank all those people that have been my fellow travelers in this trip in science and knowledge.

I owe my deepest gratitude to my supervisor, Professor Nikitas Pittis. I had the honor to work with an advisor who his encouragement, guidance and support were present all the time. In many cases, he motivated me to explore my own ideas and he has always been there, patiently, to point out both the success and the failure.

I'd also like to give a heartfelt, special thanks to Professor Nikolaos Kourogenis and Professor Phoebe Koundouri. Working with them has been thought-provoking and efficient, since our long conversations included many insightful comments and constructive criticisms.

I am also very grateful to all my Professors at the Department of Banking and Financial Management who provided motivation, support and knowledge. I must refer separately to Professor Angelos Antzoulatos, Director of our Ph.D. program, who devoted a lot of his time to organize and supervise the progress of all Ph.D. students. His effort and insight will always be an inspiration.

My fellow Ph.D. students E. Alevizopoulou., P. Asimakopoulos., C. Bouras, E. Konstantinidi, D. Kyriakopoulou, C. Daskalaki, T. Syrmos , V. Malafouris, G. Karalas, and the elders V. Babalos, K. Lambrinoudakis, T. Stamatiou, G. Papanastasopoulos and C. Tsoumas were valuable `allies' in achieving this difficult task. I owe them my gratitude and friendship.

I also want to thank Liana Apostolou, Eirini Perantontaki, Thodora Christodoulou and Eleni Kareli. Their organizational skills minimized the `frictions' and gave me the opportunity to focus on my research.

My friends have been very supportive through these years and helped me to overcome many difficulties and keep my effort high. Akis A., Giorgos A., Manos D., Michalis K., Michalis P., Nikos K., Stefanos C., I greatly value our friendship and appreciate your belief in me.

The only way to conclude the acknowledgments is to express my eternal gratitude and love to my family that stood by me, every step in this journey: my mother Vassiliki, my brother George, my sister Gianna and Giouli. Without their love, understanding and support, all this would not have been possible.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Many econometric models that are commonly used in empirical financial and economic applications are linear. However, the existing literature provides several indications that nonlinear models may be more appropriate to describe relationships encountered in real phenomena. Excess kurtosis, volatility clustering, sensitivity of estimated parameters to alternative specifications of the estimation period are all points towards the presence of nonlinearities in financial and economic time series. Therefore, more complex models are needed in order to overcome the limitations of the linear regression models to deal with these stylized facts.

This thesis examines some non-standard parametric time series models which aim at capturing several features of time series of interest, not accounted for by the usual models of the econometric literature. The first class of models considered here are autoregressive models (AR) whose parameters are autoregressive or moving average (MA) processes themselves. Already in 1973 Belsley and Kuh (1973) argued: `The rationales for time varying parameter models are several. For one, the true coefficients themselves can often be viewed directly as the outcome of a stochastic process... Second, even when the underlying parameters are stable, situations arise in which a time-varying coefficient approach will prove to be effective. Such is the case when there are specification errors, such as excluded variables or linear approximations of curvelinear forms'.

The models studied here, can be considered as natural alternatives to the standard linear autoregressive models with constant parameters in cases where the estimated parameters of the AR models exhibit persistent time variation. Note that when the persistence in the variation of the stochastic coefficients is large, the random coefficient autoregressive models that have been analyzed to a sufficient extent in the literature fail to account for the empirical dynamics. Autoregressive models with AR or MA have already been examined in the statistical literature but many issues remained unresolved. These issues include the following: (a) What are the exact necessary and sufficient conditions for these processes to be second-order stationary? (b) What are the alternative estimation methods for estimating these models? How do these methods perform in small samples? (c) How do these models perform in out-of-sample forecasting when compared with other non-linear time series models? (d) What are the properties of typical realizations from these models for various values of their parameters? The first two questions are important for the feasibility of the models under study, in estimating and forecasting the mean and the volatility of financial series, while the next two issues are related to the efficiency and suitability of these models. All these issues are analyzed in the first chapter of the present thesis.

The second class of models analyzed here, aim at describing multivariate processes which exhibit unconditional heteroskedasticity (as opposed to the standard case of conditional heteroskedasticity modelled by GARCH models). These models are motivated by the empirical observations that the variance of the returns of certain financial assets exhibits time trends. As a result, we put forward a model in which the process generating asset returns exhibits trending second moments. The idea that the unconditional variance, σ_{it}^2, of the individual stock i is a linear, or more generally, polynomial function, $h_i(t)$, of time is interesting for both theoretical and practical reasons. Knowledge of the functional form, $h_i(\cdot)$, may be very useful in forecasting the future volatility level of stock i. It will also permit direct comparisons among the rates at which the volatilities of individual stocks increase with time. In addition, the assumption $\sigma_{it}^2 = h_i(t)$ implies that asymptotically the unconditional variance becomes infinite. This feature of the suggested model is affined with the possibility that the variance of stock returns is infinite, noted by Benoit Mandelbrot (see Mandelbrot 1966). The second chapter of the thesis, analyze the problem of defining optimal portfolios under the assumption of trending variances and covariances and show that the optimal weights at each point in time are known functions of time. It must be noted that the type of non-stationarity characterising these models is quite different from the usual unit-root type of nonstationarity exhibited by non-stable autoregressive models. In both cases the variance of the underlying process increases with time; however in the former case the source of the time heterogeneity is the infinite degree of persistence whereas in the latter case the non-stationarity may coincide with an independent process.

The third category of models that this thesis is dealing with, are models which exhibit conditional heteroskedasticity. Since the seminal papers of Engle (1982) and Bollerslev (1986), the class of generalized autoregressive heteroskedastic (GARCH) models has been widely used in modeling the conditional volatility of financial time series. GARCH models became very famous because they were designed to capture and model the volatility clustering phenomenon evidenced in time series, improving the accuracy of model's predictions. Lee and Hansen (1994) noted that GARCH models have become the `workhorse of the industry'.

We focus specifically in models which exhibit very strong-persistent conditional heteroskedasticity which is manifested as a unit root in the conditional variance. These models are usually referred to as Integrated GARCH (IGARCH) models and describe an infinite-variance process. Put it different, the unit root in the conditional variance implies that the unconditional variance is infinite. Although the IGARCH models are not covariance-stationary, it can be shown that they are strictly stationary and standard inference procedures remain asymptotically valid. Bollerslev et al. (1992) provide further investigations on the properties of IGARCH models. These models are analyzed in the context of an empirical puzzle that has been observed in the financial econometrics literature. In particular, it has been observed that asset returns observed at high frequencies (say daily) exhibit IGARCH effects and fat-tailed distributions. However, as we move to lower frequencies, the distribution of returns looks more and more like the normal distribution. The puzzle lies in the fact that the observed Aggregational Gaussianity cannot be established theoretically, since the infinite variance in the original high-frequency series makes the application of the Central Limit Theorem prohibitive. However, what we show in this chapter is that, despite their infinite-variance characteristics, the high-frequency IGARCH processes still belong to the domain of attraction of the Normal law, because all the moments of order less than two exist. In other words, the high-frequency processes are barely-infinite variance processes for which a (non-standard) Central Limit Theorem applies.

Finally, in the fourth chapter we consider nonlinear models to assess the success of a procedure that evaluates mutual funds, namely that of Morningstar star rating system. Morningstar evaluates mutual funds by assigning them a number of stars (1 star for the worst-performing funds, 5 stars for the best performing ones) depending on their risk-adjusted performance over previous periods. Traditional econometric approaches that use unconditional expected returns have been shown that they are unable to distinguish whether assets selection skills and market timing abilities of mutual fund managers are attributed to common time variations in returns and risk premia.

For this purpose, we analyze multivariate regression models in which the regression coefficients are not constant, but functions of a set of observable variables. These models are applied in cases where the systematic risk of a managed portfolio, as measured by the beta coefficient, changes with time as a result of portfolio re-balancing by the active manager (see Ferson and Schadt (1996)). Put it differently, the manager, reacting to new information, as reflected on a set of publicly available financial and economic variables, changes the composition of his portfolio, thus inducing a change in portfolio's beta. This in turn implies that when the performance of the manager is assesed in terms of risk-adjusted returns, the changing nature of the systematic risk together with its dependence on the set of observable variables should be taken into account. Modeling portfolio beta time variation efficiently, will result in estimating more reliable the skills of the portfolio manager. In the case of Morningstar, our findings suggest that the better performance of the higher star rated funds reflects superior stock selection rather than market timing abilities. Overall, the implication for the Morningstar ranking system is that this is most effective in identifying the worst-performing funds rather than the best-performing ones.

CHAPTER 2

The AR(1) Model with an AR(1) or MA(1) Coefficient

2.1 Introduction

Evidence of coefficient variation in linear regression models has led to increasing interest in models with stochastic coefficients. If there are no exogenous variables in the set of regressors, these models are usually referred to as `doubly stochastic time series' (see Tjostheim 1986). In particular, the random coefficient autoregressive model of order p (RCAR(p)), which assumes that the coefficients are serially independent, zero-mean, second-order stationary random vectors, has received regular attention by both time series analysts and econometricians, over the last thirty years or so. More specifically, Conlisk (1974, 1976) and Quinn and Nicholls (1981) have provided necessary and sufficient conditions for the RCAR(p) process to be stable. (see also Feigin and Tweedie 1985). Robinson (1978), Nicholls and Quinn (1980, 1981, 1982), Hwang and Basawa (1993,1998), Koul and Schick (1996) and Schick (1996) have studied the problem of asymptotic inference for these models. Nicholls and Quinn (1982), Ramanathan and Rajarshi (1994), Lee (1998) and Akharif and Hallin (2003) have investigated the issue of testing whether the autoregressive coefficients are, indeed, random. Finally, Bera, Higgins and Lee (1992) and Tsay (1987) have proved that the autoregressive conditional heteroskedastic (ARCH) processes may be viewed as autoregressive processes with zero-mean random coefficients, thus establishing a link between two areas of the econometrics literature which, until then, were being studied independently of each other.

The RCAR(p) model can be thought of as an autoregressive model whose coefficients are randomly perturbed. These `coefficient shocks' are independent over time, or in other words, they exhibit no memory. This feature of the RCAR(p) model is rather restrictive, thus limiting the extent to which this model generalizes the corresponding, constant-coefficient AR(p) model. Concerning forecasting, in particular, the differences between the forecasts produced by RCAR(p) and AR(p) are likely to be small, at least asymptotically. More interesting models may arise, quite naturally, by assuming that the intertemporal variation in the coefficients is not independent. Such models simply recognize the fact that the effects of random events, which cause the variation in the coefficients, may last for more than one period. In other words, although the RCAR(p) models allow for changes in the magnitude of autocorrelation across time, they do not permit these changes to exhibit any systematic pattern. However, it is quite natural to assume that the effects of the factors causing these changes are of persistent rather than purely random character.

The above mentioned considerations suggest that the most natural extensions of the RCAR(p) model, would be autoregressive models whose coefficients are stochastic process, exhibiting some degree of temporal dependence. Tjostheim (1986) and Meyn and Guo (1993) developed a general formulation for the univariate and multivariate case respectively, in which conditions for stability were obtained, without postulating a generating process for the stochastic coefficients. More specific assumptions on the stochastic coefficient process may produce stability conditions which are easier to test. The straightforward choices towards this direction would be to consider autoregressive models whose coefficients are, themselves, autoregressive or moving average processes. Weiss (1985) studied the simplest of these models, namely the AR(1) model, whose coefficient, say r_t, is the sum of a constant part, ρ, and a stochastic part, ρ_t, i.e. $y_t = (\rho + \rho_t)y_{t-1} + u_t$, with

$u_t \sim IID(0, \sigma_u^2)$ and the time-varying part is assumed to follow an AR(1) process itself, that is $\rho_t = \phi\rho_{t-1} + v_t$ with $v_t \sim IID(0, \sigma_\nu^2)$. Tjostheim (1986) studied the case where the stochastic part follows a MA(1) process, that is $\rho_t = \phi v_{t-1} + v_t$ with $v_t \sim IID(0, \sigma_\nu^2)$. The question to be answered in both cases, is what conditions quarrantee the asymptotic second-order stability of the process, $\{y_t\}$, in the sense that $E(y_t \mid y_0)$ and $E(y_t y_{t-s} \mid y_0)$, for fixed $s = 0, 1, ...$ converge to fixed values not depending on the initial value y_0. Weiss derived such a sufficient condition at a rather high price, involving both assumptions and approximations (see below). Moreover, the constraints on the model parameters that this sufficient condition implies are not explicit and depend on the sample size. Weiss also established a necessary condition for second-order stationarity of the AR(1)/AR(1) process, which is useful in the sense that its violation ensures that the underlying process is not second-order stable. Tjostheim derived necessary and sufficient conditions for the underline process y_t to be asymptotically second order stationary in the case where v_t are Gaussian, not allowing for deviations from symmetric and mesokurtic random errors.

Grillenzoni (1993) studies the more general case of $ARIMA(p, d, q)$ models with $ARIMA(P, D, Q)$ coefficients. This study analyzes in depth the identification-estimation problem of such models and shows that the Kalman filter is formally implementable and the ML estimates are, in general, consistent and asymptotically normal. This opens up the way for hypothesis testing on the model parameters, provided that the true values of these parameters lie in the interior of the parameter space. It also allows for the calculation of the usual objective model selection criteria. As far as the problem of deriving sharp sufficient conditions for the stability of these models, Grillenzoni (1993) acknowledges this to be a very difficult problem. In particular, he remarks that even in the simplest cases of an AR(1) model with an AR(1) or a MA(1) coefficient, the problem of deriving such conditions is "challenging".

The present study focuses on these two special cases mentioned above that is AR(1) / AR(1) and AR(1) / MA(1) models and explores three issues associated with these models: First it obtains useful sufficient conditions for second-order stability of the AR(1)/AR(1) process and necessary and sufficient conditions for the second-order stability of the AR(1)/MA(1) process. Second, it investigates the problem of model selection within a class of models containing apart from AR(1)/AR(1) and AR(1)/MA(1) the random and constant coefficient models as well. In particular, by means of Monte Carlo experiments it examines the frequency at which the usual information criteria such as the Akaike, (AIC), Swzartz (SIC) and the Hannan-Quinn (HQ) ones select the correct model under a variety of alternative data generation processes and sample sizes. As regards the third of the aforementioned issues, we provide evidence supporting the view that many financial time series are best described by either AR(1)/AR(1) or AR(1)/MA(1) models. This claim is supported by both in-sample criteria as well as by the superior forecasting performance of these model against that of constant or random coefficient ones.

The outline of this chapter is as follows: Sections II and III derive the sufficient condition for the stability of the AR(1)/AR(1) and AR(1)/MA(1) models, respectively. Section IV discusses briefly maximum likelihood (ML) estimation of the model parameters and presents the simulation results on the finite sample behaviour of these estimators and the performance of model selection criteria. Section (V) presents the empirical results, while Section (VI) concludes the chapter.

2.2 The AR(1) model with an AR(1) coefficient.

Let us consider the following process:

$$y_t = (\rho + \rho_t)y_{t-1} + u_t \ , t \geq 0 \ , \tag{2.1}$$

where

$$\rho_t = \phi\rho_{t-1} + v_t \tag{2.2}$$

and

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim NIID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}\right], \tag{2.3}$$

with the additional simplifying and rather innocuous assumption that $\sigma_u^2 = 1$. Weiss (1985) examined the model defined by (2.1)-(2.3) and focused on the quantity $W \equiv T^{1/2}S^2(\infty)^{-1/2}(X_T - R)$, where $X_k = \frac{1}{k}\sum_{i=1}^{k}\overline{\rho}_{t-i}^2$, $\overline{\rho}_t = (\rho + \rho_t)$, $S^2(\infty) = \lim_{k\to\infty} S^2(k)$,

$S^2(k) = k \cdot var(X_k)$, $R = E(\overline{\rho}_t^2)$, and first proved that $S^2(\infty) < \infty$. Then, the normality assumption for v_t enabled him to appeal to the Central Limit Theorem (theorem 2.4, in White and Domowitz 1984) in order to approximate W with a standard normal random variable. Conditional on this approximation, he adopted a second, rather more restrictive, approximation: He utilized a result by Granger and Andersen (1978), according to which the $n-th$ moment, $[E(X_n)^n]$, is approximately generated by a second-order difference equation. By doing so, he was then able to produce an approximate sufficient condition for the stability of y_t, namely $R + S^2(\infty) < 1$, whose validity, of course, depends on the accuracy and relevance of the above mentioned consecutive approximations. Moreover, one should also note that this condition is far from being necessary; in fact it is a rather strict one, thus excluding many stable cases that might be useful in practice. Weiss also obtains a useful necessary condition for second-order stability, which takes the form

$$R = \rho^2 + \frac{\sigma_\nu^2}{1-\phi^2} < 1 \tag{2.4}$$

Guyton, Zang and Foutz (1986), Pourhamadi (1986) and Grillenzoni (1993) consider the simpler model in which $\rho = 0$. This model assumes that the mean of the autoregressive coefficient process, $\{\rho_t\}$, is equal to zero. Under this restrictive assumption, it was shown that $\{y_t\}$ is second-order stationary, if $\{\rho_t\}$ belongs with probability probability 1 to the interval $(-1, +1)$ almost everywhere. This means that the realizations of $\{\rho_t\}$ are allowed to lie outside the stationarity region of a constant-coefficient AR(1) model only for finite periods. To the best of our knoweledge, no sufficient conditions for the stability of $\{y_t\}$ are available in the literature, for the more general case, $\rho \neq 0$, i.e. when the autoregressive coefficient process contains both a constant and a time-varying element.

To this end, we can write y_t as

$$y_t = y_0 \prod_{i=1}^{t} (\rho + \rho_i) + \sum_{k=1}^{t-1} u_k \prod_{i=k+1}^{t} (\rho + \rho_i) + u_t \ . \tag{2.5}$$

Then, we observe that strict stationarity of $\{y_t\}_{t\geq 1}$ depends on the stationarity of the terms

$$\prod_{i=1}^{t}(\rho+\rho_i) \text{ and } \sum_{k=1}^{t-1} u_k \prod_{i=k+1}^{t}(\rho+\rho_i) \ ,$$

which is guaranteed if and $\{\rho_t\}_{t\geq 1}$ and $\{u_t\}_{t\geq 1}$ are stationary and

$$|\rho+\rho_t|<1 \tag{2.6}$$

for almost every t (see Stout, 1974, p.184, and Grillenzoni, 1993). The stationarity of $\{\rho_t\}_{t\geq 1}$ and $\{u_t\}_{t\geq 1}$ is, in turn, guaranteed by (2.3) and the additional condition

$$|\phi|<1 \ . \tag{2.7}$$

Moreover, condition (2.6) guarantees the existence of all the moments for $\{y_t\}_{t\geq 1}$ since $\prod_{i=1}^{t}(\rho+\rho_i) \xrightarrow{p} 0$ as $t\to\infty$.

The preceding discussion motivates the derivation of sharp conditions that will guarantee the validity of (2.6). These conditions will concern the range of the random variables v_t's.

2.2.1 Exact Sufficient Conditions for the Stability of the AR(1) /AR(1) model, with $\rho\neq 0$.

In this section we obtain two alternative sets of conditions that guarrantee the second-order stability of $\{y_t\}$. The first set, $\mathbf{C}_1$, assumes that ϕ is positive whereas the second set, $\mathbf{C}_2$, assumes that ϕ is negative. These sets are defined as follows:

Conditions $\mathbf{C}_1$: a) $0\leq\phi<1$, b) $|\rho|<1$, c) $v_t\in[\alpha,\beta]$, and $-(1-\phi)(1+\rho)<\alpha\leq 0\leq \beta<(1-\phi)(1-\rho)$.

Conditions $\mathbf{C}_2$: a) $-1<\phi\leq 0$, b) $|\rho|<1$, c) $v_t\in[\alpha,\beta]$, and $-(1+\rho)-\phi(1-\rho)<\alpha\leq 0\leq\beta<1-\rho+\phi(1+\rho)$.

Proposition 1: *If conditions C_1 hold, then there exists a real number, $b: 0<b<1$, such that for every $t\geqslant 1$, $\rho+\rho_t\in[-b,b]$.*

Proof: The proof is obtained by induction. First note that since $\rho_1 = v_1$ and $v_1 \in [\alpha, \beta]$ we have

$$\rho + \rho_1 \in [\alpha + \rho, \beta + \rho]$$

From $\mathbf{C}_1$(c), there exists $\varepsilon > 0$ such that

$$\alpha + \rho > \varepsilon - (1-\phi)(1+\rho) + \rho = \varepsilon - 1 - \rho + \phi + \phi\rho + \rho = \phi(1+\rho) - 1 + \varepsilon > -1 + \varepsilon$$

and

$$\beta + \rho < -\varepsilon + (1-\phi)(1-\rho) + \rho = -\varepsilon + 1 - \rho - \phi + \phi\rho + \rho = 1 - \phi(1-\rho) - \varepsilon < 1 - \varepsilon\ .$$

Thus $\rho + \rho_1 \in [-1+\varepsilon, 1-\varepsilon]$. Let $\rho + \rho_i \in [-1+\varepsilon, 1-\varepsilon]$. Then

$$\begin{aligned}
\rho + \rho_{i+1} &= \rho + \phi\rho_i + v_{i+1} \\
&\Longrightarrow \rho + \rho_{i+1} \in [\rho + \phi(-1+\varepsilon-\rho), \rho + \phi(1-\varepsilon-\rho)] + [\alpha, \beta] \\
&\subseteq [\rho + \phi(-1+\varepsilon-\rho) + \varepsilon - (1-\phi)(1+\rho), \rho + \phi(1-\varepsilon-\rho) + \\
&(1-\phi)(1-\rho) - \varepsilon] \\
&= [-1+\varepsilon+\phi\varepsilon, 1-\varepsilon-\phi\varepsilon] \subseteq [-1+\varepsilon, 1-\varepsilon]\ .
\end{aligned}$$

Set $b = 1 - \varepsilon$ and the proof is complete.

For the case $\phi \leq 0$, we have Proposition 2:

Proposition 2: *If conditions* $\mathbf{C}_2$ *hold, then there exists* $0 < b < 1$ *such that* $\rho + \rho_t \in [-b, b]$.

Proof: The proof is similar to the one of Proposition 1 and, therefore, it is omitted.

Remark 1: Conditions $\mathbf{C}_1$ and $\mathbf{C}_2$ can be translated to restrictions on the range of values of the coefficient $\rho + \rho_t$. The following two examples demonstrate more clearly this effect. First consider the case $\phi = 0.6$ and $\rho = 0.7$. Then, in order to satisfy conditions $\mathbf{C}_1$, the values of v_t should belong to the interval $[-0.68, 0.12]$. Suppose that $\alpha = -0.66$

and $\beta = 0.1$. Then, one can replace these values in the proof of Proposition 1 and show that $\rho + \rho_t \in [-0.95, 0.95]$ for every $t \geqslant 1$. As another example, suppose that $\phi = -0.2$ and $\rho = 0.3$, which (according to conditions $\mathbf{C}_2$) implies that v_t must take values in the interval $[-1.16, 0.44]$. In this case, it can be shown that $\rho + \rho_t \in [-0.93, 0.93]$ for every $t \geqslant 1$.

Note that Propositions 1 and 2 do not make any distributional assumptions on v_t's apart from restricting their range. Therefore, the normality assumption in (2.3) can be dropped without affecting the results.

The arguments presented above may be thought of as a proof of the following theorem:

Theorem 2.2.1 *If $\{y_t\}_{t\geq 1}$ is generated by (2.1), (2.2) and conditions (2.7), (**??**) and C_1 or C_2 hold, then $\{y_t\}_{t\geq 1}$ is a strictly stationary process with finite second moments.*

2.3 The AR(1) model with a MA(1) coefficient.

In this section, we obtain necessary and sufficient conditions for the asymptotic second order stability of an AR(1)/MA(1) process:

$$y_t = (\rho + v_t + \theta v_{t-1}) y_{t-1} + u_t \ . \tag{2.8}$$

The case of and AR(1) model with a purely MA(1) coefficient was first analyzed by Tjostheim (1986). Tjostheim noted that `it seems very hard to obtain general conditions for second order stationarity' for doubly stochastic processes and discussed general techiniques towards this aim. He focused on the convergence in mean square of the term

$$\sum_{j=0}^{n} E\Big(\prod_{i=0}^{j-1} (\rho + \theta v_{t-1-i} + v_{t-i})^2\Big) \tag{2.9}$$

.By assuming normality, he used the characteristic function of $\{\rho + \theta v_{t-1-i} + v_{t-i}\}_{i\geq 0}$ and derived a coupled system of differential equations for the summand in (2.9). The sollution of this system provided the necessary and sufficient conditions for the asymptotic second order stationarity of an AR(1) process with a constant plus a MA(1) autoregressive coefficient. We proceed in deriving necessary and sufficient conditions for process (2.8) without making distributional assumptions.

Assumption 1: The random sequences $\{u_t\}_{t\geqslant 1}$ and $\{v_t\}_{t\geqslant 1}$ are neither serially nor contemporaneously correlated, with $E[u_t] = E[\nu_t] = 0$, $E[u_t^2] = \sigma_u^2 < \infty$, $E[v_t^2] = \sigma_v^2 < \infty$, $E[v_t^3] = \mu_3 < \infty$ and $E[v_t^4] = \mu_4 < \infty$, for every $t \geqslant 1$. Moreover, y_0 is a constant or a random variable, with finite first moment, uncorrelated with v_t, for $t \geqslant 0$.

We will obtain necessary and sufficient conditions for the stability of process (1).

By applying (4) to (1) we have

$$y_t = (\rho + v_t + \theta v_{t-1})y_{t-1} + u_t \ . \tag{2.10}$$

Set

$$r_t = \rho + \rho_t = \rho + v_t + \theta v_{t-1} \ .$$

Then

$$y_t = r_t y_{t-1} + u_t \tag{2.11}$$

$$\Rightarrow y_t = y_0 \prod_{i=1}^{t} r_i + \sum_{i=1}^{t-1} u_{t-i} \prod_{j=t-i+1}^{t} r_j + u_t \tag{2.12}$$

2.3.1 Necessary and Sufficient Conditions for First-Order Stability.

From equation (2.11) we have that

$$E[y_t] = E[y_0]E\left[\prod_{i=1}^{t} r_i\right] \ .$$

Let us focus on the quantity $E\left[\prod_{i=1}^{t} r_i\right]$:

$$\begin{aligned}
E\left[\prod_{i=1}^{t} r_i\right] &= E\left[r_t \prod_{i=1}^{t-1} r_i\right] = E\left[(\rho + v_t + \theta v_{t-1}) \prod_{i=1}^{t-1} r_i\right] \\
&= \rho E\left[\prod_{i=1}^{t-1} r_i\right] + \theta E\left[v_{t-1}(\rho + v_{t-1} + \theta v_{t-2}) \prod_{i=1}^{t-2} r_i\right] \\
\Rightarrow E\left[\prod_{i=1}^{t} r_i\right] &= \rho E\left[\prod_{i=1}^{t-1} r_i\right] + \theta \sigma_v^2 E\left[\prod_{i=1}^{t-2} r_i\right] \qquad (2.13)
\end{aligned}$$

Equation (7) is a difference equation and its convergence is guaranteed when both of the roots z_1, z_2 of the characteristic equation

$$z^2 - \rho z - \theta \sigma_v^2 = 0 \qquad (2.14)$$

lie in the interior of the unit circle. This restriction means that

$$\left| \frac{\rho \pm \sqrt{\rho^2 + 4\theta\sigma_v^2}}{2} \right| < 1 \ . \qquad (2.15)$$

Therefore, in order for the first moment of $\{y_t\}_{t\geq 1}$ to converge, the parameters ρ, σ_v^2 and θ must belong to the subset Δ_1 of the parametric space:

$$\Delta_1 = \left\{(\rho, \theta, \sigma_v) : -2 < \rho < 2 \text{ and } -1 < \theta\sigma_v^2 < 1 - |\rho|\right\} \ . \qquad (2.16)$$

Remark 2: Since equation (7) is a homogeneous difference equation, we conclude that when $(\rho, \theta, \sigma_v) \in \Delta_1$ then $E[y_t] \to 0$ as $t \to \infty$, for every y_0, defined in Assumption 1. To show that $(\rho, \theta, \sigma_v) \in \Delta_1$ is also necessary for first-order stability, we argue as follows: Assume that $(\rho, \theta, \sigma_v) \in \partial\Delta_1$ (i.e. the boundary of Δ_1). Then, we can distinguish the following two cases. The first one is when at least one of the roots z_1, z_2 is not a real number. Then, from the theory of difference equations, we have that the sequence defined by equation (7) does not converge. The second case is when both z_1 and z_2 are real numbers. Then, in order to have convergence of $E\left[\prod_{i=1}^{t} r_i\right]$, we should have $z_1 = 1$

(or $z_2 = 1$) and $z_2 \in (-1, 1]$ (or $z_1 \in (-1, 1]$). In such a case, $E[y_t]$ is asymptotically dependent on the initial value, y_0. Therefore, the parametric set Δ_1 defines both necessary and sufficient conditions on the parameters ρ, θ and σ_v, for the first-order stability of $\{y_t\}_{t\geq 1}$.

2.3.2 Necessary and Sufficient Conditions for Second-Order Stability.

Let us now examine the convergence of $E[y_t y_{t+s}]$, $s \geq 0$.

First note that

$$y_{t+s} = y_t \prod_{i=t+1}^{t+s} r_i + \sum_{i=1}^{s-1} u_{t+s-i} \prod_{j=t+s-i+1}^{t+s} r_j + u_{t+s} \tag{2.17}$$

From (11), for $s = 1$, we have

$$E[y_t y_{t+1}] = E[y_t^2 r_{t+1}] = \rho E[y_t^2] + \theta E[v_t y_t^2] \tag{2.18}$$

For $s \geq 2$, we have

$$\begin{aligned} E[y_t y_{t+s}] &= E[y_t^2 \prod_{i=t+1}^{t+s} r_i] = E[y_t^2(\rho + v_{t+1} + \theta v_t) \prod_{i=t+2}^{t+s} r_i] \\ &= \left(E[(\rho + v_{t+1}) \prod_{i=t+2}^{t+s} r_i] \right) E[y_t^2] + \theta E[\prod_{i=t+2}^{t+s} r_i] E[v_t y_t^2] \end{aligned} \tag{2.19}$$

Moreover

$$\begin{aligned} E[y_t^2] &= \sigma_u^2 + E[(\rho^2 + v_t^2 + \theta^2 v_{t-1}^2 + 2\rho\theta v_{t-1} + 2\rho v_t + 2\theta v_t v_{t-1}) y_{t-1}^2] \\ &= \sigma_u^2 + (\rho^2 + \sigma_v^2) E[y_{t-1}^2] + \theta^2 E[v_{t-1}^2 y_{t-1}^2] + 2\rho\theta E[v_{t-1} y_{t-1}^2] \end{aligned} \tag{2.20}$$

$$
\begin{aligned}
E[v_t y_t^2] &= E[v_t(\rho^2 + v_t^2 + \theta^2 v_{t-1}^2 + 2\rho\theta v_{t-1} + 2\rho v_t + 2\theta v_t v_{t-1}) y_{t-1}^2] \\
&= (2\rho\sigma_v^2 + \mu_3)E[y_{t-1}^2] + 2\theta\sigma_v^2 E[v_{t-1}y_{t-1}^2] \qquad (2.21)
\end{aligned}
$$

and

$$
\begin{aligned}
E[v_t^2 y_t^2] &= \sigma_v^2\sigma_u^2 \\
&\quad + E[v_t^2(\rho^2 + v_t^2 + \theta^2 v_{t-1}^2 + 2\rho\theta v_{t-1} + 2\rho v_t + 2\theta v_t v_{t-1}) y_{t-1}^2] \\
&= \sigma_v^2\sigma_u^2 + (\sigma_v^2\rho^2 + 2\rho\mu_3 + \mu_4)E[y_{t-1}^2] + 2\theta(\rho\sigma_v^2 + \mu_3)E[v_{t-1}y_{t-1}^2] \\
&\quad + \theta^2\sigma_v^2 E[v_{t-1}^2 y_{t-1}^2] \qquad (2.22)
\end{aligned}
$$

From equations (14), (15) and (16) we have that

$$
\begin{bmatrix} E[y_t^2] \\ E[v_t y_t^2] \\ E[v_t^2 y_t^2] \end{bmatrix} = \begin{bmatrix} \rho^2+\sigma_v^2 & 2\rho\theta & \theta^2 \\ 2\rho\sigma_v^2+\mu_3 & 2\theta\sigma_v^2 & 0 \\ \sigma_v^2\rho^2+2\rho\mu_3+\mu_4 & 2\theta(\rho\sigma_v^2+\mu_3) & \theta^2\sigma_v^2 \end{bmatrix} \begin{bmatrix} E[y_{t-1}^2] \\ E[v_{t-1}y_{t-1}^2] \\ E[v_{t-1}^2 y_{t-1}^2] \end{bmatrix} + \begin{bmatrix} \sigma_u^2 \\ 0 \\ \sigma_v^2\sigma_u^2 \end{bmatrix} \qquad (2.23)
$$

Set

$$
A = \begin{bmatrix} \rho^2 + \sigma_v^2 & 2\rho\theta & \theta^2 \\ 2\rho\sigma_v^2 + \mu_3 & 2\theta\sigma_v^2 & 0 \\ \sigma_v^2\rho^2 + 2\rho\mu_3 + \mu_4 & 2\theta(\rho\sigma_v^2 + \mu_3) & \theta^2\sigma_v^2 \end{bmatrix}, b = \begin{bmatrix} \sigma_u^2 \\ 0 \\ \sigma_v^2\sigma_u^2 \end{bmatrix} \qquad (2.24)
$$

$$
\text{and } Y_t = \begin{bmatrix} E[y_t^2] \\ E[v_t y_t^2] \\ E[v_t^2 y_t^2] \end{bmatrix}.
$$

Then

$$Y_t = AY_{t-1} + b = A^{t-1}Y_1 + \left(\sum_{i=0}^{t-2} A^i\right) b \; . \tag{2.25}$$

Therefore, the necessary and sufficient condition for the convergence of $\sum_{i=0}^{t-2} A^i$ and, consequently of A^{t-1}, is that all the eigenvalues ($\lambda_i,\ i = 1, 2, 3$) of A must lie inside the unit circle (i.e. $\|\lambda_i\| < 1$). Let us define the corresponding parameter set Δ_2, as follows: $\Delta_2 = \{(\rho, \theta, \sigma_v, \mu_3, \mu_4)$ such that $\max\{\|\lambda_i\|,\ i = 1, 2, 3\} < 1$. For $(\rho, \theta, \sigma_v, \mu_3, \mu_4) \in \Delta_2$, the matrix $I - A$ is invertible and therefore from (2.25) we have that

$$Y_t = A^{t-1}Y_1 + \left(I - A^{t-1}\right)(I - A)^{-1}b \Rightarrow Y_t \to (I - A)^{-1}b \tag{2.26}$$

as $t \longrightarrow \infty$.

Remark 3: For the special case of a random coefficient (i.e. when $\theta = 0$) the above condition coincides with the necessary and sufficient condition of Nichols and Quinn (1981). Indeed, since the only non zero eigenvalue of A is equal to $\rho^2 + \sigma_v^2$, the retriction $\rho^2 + \sigma_v^2 < 1$ (along with $\theta = 0$) is equivalent to saying that $(\rho, \theta, \sigma_v) \in \Delta_1$. Note that for this particular case μ_3 and μ_4 can take any value.

Remark 4: For the general case, $\theta \neq 0$, note that the limit of the sequence $\{Y_t\}_{t\geqslant 1}$ is independent of y_0. Let us define the following subset of the parameter space, $\Delta_3 \subseteq \mathbb{R}^3$ $\equiv \{(\rho, \theta, \sigma_v)$ such that $(\rho, \theta, \sigma_v, \mu_3, \mu_4) \in \Delta_2$ for some μ_3 and $\mu_4\}$. We also observe that if $\Delta_3 \smallsetminus \bar{\Delta}_1 \neq \emptyset$, where $\bar{\Delta}_1$ is the closure of Δ_1, then for some parameters (ρ, θ, σ_v) the second moments would be convergent, whereas the first moment would be explosive. Since such a configuration is impossible, we conclude that $\Delta_3 \subseteq \bar{\Delta}_1$**.** Therefore, the sufficient and necessary condition for the stability of the process $\{y_t\}_{t\geq 1}$ is that $(\rho, \theta, \sigma_v, \mu_3, \mu_4) \in \Delta \equiv \{(\rho, \theta, \sigma_v, \mu_3, \mu_4) \in \Delta_2$ such that $(\rho, \theta, \sigma_v) \in \Delta_3 \smallsetminus \partial\Delta_1\}$.

Remark 5: Note that if the v_t's are independent and have a symmetric distribution, then $\mu_3 = 0$. For the case of $v_t \sim N(0, \sigma_v^2)$, we have that $\mu_4 = 3\sigma_v^4$, thus reducing the dimension of the parameter space, Δ_2, from five to three. For this case, Figure (2.1) illustrates regions of second order stability in the $(\rho, \sigma_v)-$ plane, for some positive and negative values of θ. It is worth noticing that for negative values of θ, the stability regions include points that lie outside the unit circle, which is the boundary of the stability region

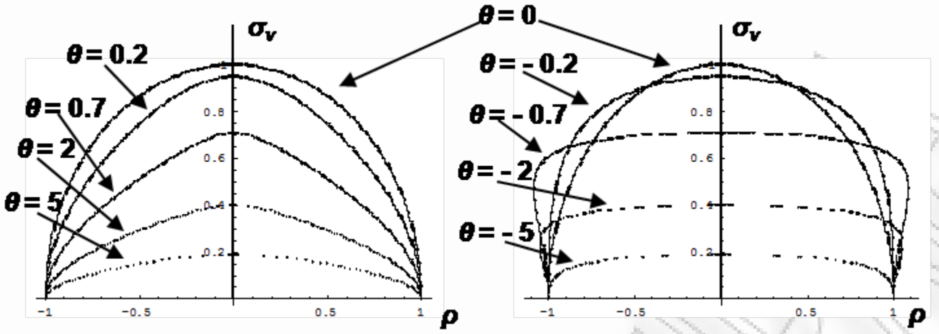


Figure 2.1: Regions of second order stability in the $(\rho, \sigma_v)-$ plane.

for $\theta = 0$. This in turn implies that the process $\{y_t\}_{t\geq 1}$ can be stable even if $|\rho|$ is slightly greater than unity. For example, for $\rho = 1.05$, $\theta = -2$ and $\sigma_v = 0.23$ a realization of $\{y_t\}_{t\geq 1}$ is illustrated in Figure (2.2).

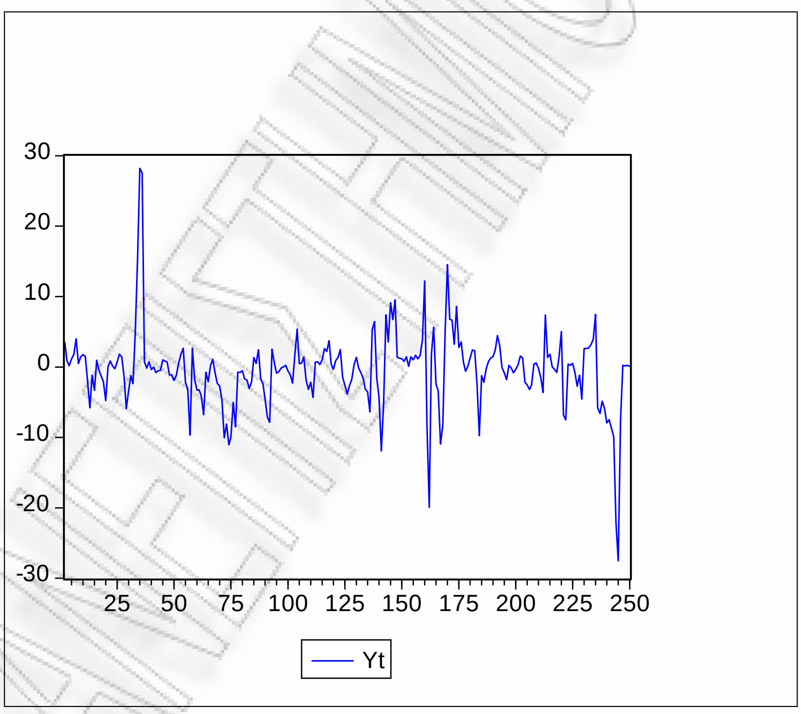


Figure 2.2: Realization of y_t

Under the necessary and sufficient stability condition, defined by the parameter space, Δ_2, the analytical expressions for the limit of $E[y_t y_{t+s}]$, $s \geq 0$ are as follows: For $s = 0$, we have

$$E[y_t^2] \to \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T (I-A)^{-1} b \text{ as } t \to \infty. \tag{2.27}$$

For $s = 1$, we utilize equation (12) to obtain,

$$E[y_t y_{t+1}] = \begin{bmatrix} \rho \\ \theta \\ 0 \end{bmatrix}^T Y_t \to \begin{bmatrix} \rho \\ \theta \\ 0 \end{bmatrix}^T (I-A)^{-1} b \text{ as } t \to \infty. \tag{2.28}$$

Finally, for $s \geq 2$, we utilize equation (13) to obtain,

$$E[y_t y_{t+s}] = \begin{bmatrix} E[(\rho + v_{t+1}) \prod_{i=t+2}^{t+s} r_i] \\ \theta E[\prod_{i=t+2}^{t+s} r_i] \\ 0 \end{bmatrix}^T Y_t$$

$$\Rightarrow E[y_t y_{t+s}] \to \begin{bmatrix} E[(\rho + v_1) \prod_{i=2}^{s} r_i] \\ \theta E[\prod_{i=2}^{s} r_i] \\ 0 \end{bmatrix}^T (I-A)^{-1} b \text{ as } t \to \infty \ , \tag{2.29}$$

since in the expression $E[(\rho + v_{t+1}) \prod_{i=t+2}^{t+s} r_i]$ only the variance of v_t appears, which is independent of t.

Remark 6: Substituting (2.24) in (2.27), we get the asymptoric constant variance of y_t :

$$var(y_t) \to \frac{\sigma_u^2(1 - 2\sigma_v^2\theta)}{G}$$

where

$$
\begin{aligned}
G &= 1-\mu_4\theta^2-2(\mu_3)^2\theta^3-2(\sigma_v^2)^3\theta^3-2\mu_3\rho\theta(1+\theta)- \\
&\quad -\rho^2(1+2\sigma_v^2\theta)+(\sigma_v^2)^2\theta(2+\theta+\theta^2)-\sigma_v^2(1+2\theta+\theta^2-2\mu_4\theta^3).
\end{aligned}
$$

If we set $\mu_3 = 0$ and $\mu_4 = 3\sigma_{v,}^2$ which are the corresponding moments of the normal distribution, then, after some manipulations, we get:

$$
G = (1-\rho^2-\sigma_v^2(1+\theta^2)-2(\sigma_v^2\theta)^2)(1-2\sigma_v^2\theta)-4\rho^2\sigma_v^2\theta
$$

which is the asymptotic variance derived by Tjostheim.

Remark 7: Note that the limit in (22) depends on the expectation of the product $\prod_{i=2}^{s} r_i$, which tends to zero as $s \to \infty$, (see the proof for the first order stability conditions). This in turn implies that a stable $\{y_t\}_{t\geq 1}$ is also asymptotically non-correlated.

2.4 Estimation Issues

The estimation problems under study may be thought of as special cases of the more general problem addressed by Grillenzoni (1993). This study shows that the procedure of maximizing a Gausssian likelihood function which is built on the prediction-error decomposition realized by the Kalman filter is `formally implementable and statistically efficient for doubly stochastic models' (Grillenzoni 1993, pp. 238). Furthermore, the paper demonstrates the derivation of the necessary Kalman filter equations in the case of AR(1)/AR(p) model. Tjostheim (1986) derived the Kalman filter equation for the AR(1)/MA(1) case for forecasting purposes but did not discussed the estimation procedure.

As a result, we do not discuss the present estimation problems in detail, but, instead we briefly explain how the Kalman filter equations and the likelihood function are defined for the cases of interest. Let us first define $\rho_{t|t-1} := E(\rho_t|\mathcal{F}_{t-1})$ and $P_{t|t-1} := var(\rho_t|\mathcal{F}_{t-1})$, with $\mathcal{F}_{t-1}$ being the information set available at time t-1.

2.4.1 The AR(1)-AR(1) Case

The AR(1)/AR(1) model defined by (2.1)-(2.2) is in state-space form, which allows the Gaussian log-likelihood function to be computed by the the Kalman filter.

Then, we have:

$$E(y_t|\mathcal{F}_{t-1}) = E(\rho y_{t-1} + \rho_t y_{t-1} + u_t|\mathcal{F}_{t-1}) = \rho y_{t-1} + \rho_{t|t-1} y_{t-1}$$

$$var(y_t|\mathcal{F}_{t-1}) = var(c + \rho y_{t-1} + \rho_t y_{t-1} + u_t|\mathcal{F}_{t-1}) = y_{t-1}^2 P_{t|t-1} + \sigma_u^2$$

and

$$\begin{aligned} cov(y_t, \rho_t|\mathcal{F}_{t-1}) &= cov(c + \rho y_{t-1} + \rho_t y_{t-1} + u_t, \rho_t|\mathcal{F}_{t-1}) \\ &= cov(\rho_t y_{t-1}, \rho_t|\mathcal{F}_{t-1}) = y_{t-1} P_{t|t-1}. \end{aligned}$$

Finally, by assuming joint normality for ρ_t and y_t, conditional on information set $\mathcal{F}_{t-1}$, we conclude that:

$$\left(\left|\begin{bmatrix}\rho_t \\ y_t\end{bmatrix}\right| \mathcal{F}_{t-1}\right) \sim N \begin{bmatrix} \rho_{t|t-1} \\ c + \rho y_{t-1} + \rho_{t|t-1} y_{t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & y_{t-1} P_{t|t-1} \\ y_{t-1} P_{t|t-1} & y_{t-1}^2 P_{t|t-1} + \sigma_u^2 \end{bmatrix} \tag{2.30}$$

Using the properties of the Gaussian joint distribution and (2.30) we obtain that $\rho_t|\mathcal{F}_t \sim N(\rho_{t|t}, P_{t|t})$,where:

$$\rho_{t|t} = E\{\rho_t|\mathcal{F}_t) = \rho_{t|t-1} + y_{t-1} P_{t|t-1} (y_{t-1}^2 P_{t|t-1} + \sigma_u^2)^{-1} (y_t - c - \rho y_{t-1} - \rho_{t|t-1} y_{t-1})$$

and

$$P_{t|t} = var(\rho_t|\mathcal{F}_t) = P_{t|t-1} - (y_{t-1} P_{t|t-1})^2 (y_{t-1}^2 P_{t|t-1} + \sigma_u^2)^{-1}$$

Once we have obtained the updating equations for the state variable and its variance, we can derive the so-called `prediction equations' which, for the case under study, take the form:

$$\rho_{t|t-1} = E(\rho_t|\mathcal{F}_{t-1})$$
$$= \phi\rho_{t-1|t-2} + \phi y_{t-2}P_{t-1|t-2}(y_{t-2}^2 P_{t-1|t-2} + \sigma_u^2)^{-1}\times \quad (2.31)$$
$$\times(y_{t-1} - c - \rho y_{t-2} - \rho_{t-1|t-2}y_{t-2}) \quad (2.32)$$

$$P_{t|t-1} = var(\rho_t|\mathcal{F}_{t-1})$$
$$= \phi^2 P_{t-1|t-1} + \sigma_v^2 = \phi^2 P_{t-1|t-2} - \quad (2.33)$$
$$-(\phi y_{t-2}P_{t-1|t-2})^2(y_{t-2}^2 P_{t-1|t-2} + \sigma_u^2)^{-1} + \sigma_v^2 \quad (2.34)$$

Under the assumption that the the state process is stationary, the filter may be initialised to the unconditional mean and variance of the state variable, that is $\rho_{1|0} = E(\rho_1) = 0$ and $P_{1|0} = var(\rho_1) = \frac{\sigma_v^2}{1-\phi^2}$. Nicholls and Quinn (1980,1981) discussed the estimation procedure of autoregressive random coefficient models and proposed an initial calculation for $var(\rho_1)$ (with $\phi = 0$) which can be utilized in our case.

Since, y_t conditional on $\mathcal{F}_{t-1}$ is normally distributed, the logarithm of the likelihood function is expressed as

$$\ln L(\theta; y_t) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln\left\{y_{t-1}^2 P_{t|t-1} + \sigma_u^2\right\} \quad (2.35)$$
$$-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \rho y_{t-1} - y_{t-1}\rho_{t|t-1})^2}{\left\{y_{t-1}^2 P_{t|t-1} + \sigma_u^2\right\}}. \quad (2.36)$$

Alternatively, we may adopt the so-called `Aitkenn formulation' of the likelihood function, which is based on the transformation of the original state space model to a single regression model characterised by conditional heteroskedasticity and serial correlation. Specifically, we rewrite **??** as follows:

$$y_t = \rho y_{t-1} + w_t \ , \tag{2.37}$$

where

$$w_t = \rho_t y_{t-1} + u_t = y_{t-1}[\phi^T \rho_0 + \sum_{i=0}^{T-1} \phi^i v_{t-i}] + u_t. \tag{2.38}$$

Hence,

$$E(w_t|\mathcal{F}_{t-1}) = E(y_{t-1}[\phi^T \rho_0 + \sum_{i=0}^{T-1} \phi^i v_{t-i}] + u_t|\mathcal{F}_{t-1}) = 0 \tag{2.39}$$

and

$$var(w_t|\mathcal{F}_{t-1}) = y_{t-1}^2[\phi^{2T}\frac{\sigma_v^2}{1-\phi^2} + \sigma_v^2 \sum_{i=0}^{T-1} \phi^{2i}] + \sigma_v^2 = \sigma_v^2 y_{t-1}^2[\frac{\phi^{2T}}{1-\phi^2} + \sum_{i=0}^{T-1} \phi^{2i}] + \sigma_u^2 \tag{2.40}$$

Combining (2.37), (2.39) and (2.40) we obtain

$$E(y_t|\mathcal{F}_{t-1}) = \rho y_{t-1} \tag{2.41}$$

and

$$var(y_t|\mathcal{F}_{t-1}) = \sigma_v^2 y_{t-1}^2[\frac{\phi^{2T}}{1-\phi^2} + \sum_{i=0}^{T-1} \phi^{2i}] + \sigma_u^2. \tag{2.42}$$

Based on (2.41) and (2.42), the logarithm of the likelihood function takes the form:

$$\begin{aligned} \ln L(\theta; y_t) \quad = \quad & -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left\{\ln \sigma_v^2 y_{t-1}^2 \left[\sum_{i=0}^{T-1} \phi^{2i} + \frac{\phi^{2T}}{1-\phi^2}\right] + \sigma_u^2\right\} \\ & -\frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \rho y_{t-1})^2}{\sigma_v^2 y_{t-1}^2 \left[\sum_{i=0}^{T-1} \phi^{2i} + \frac{\phi^{2T}}{1-\phi^2}\right] + \sigma_u^2} \end{aligned} \tag{2.43}$$

Remark 2: The maximum likelihood estimator of ρ, coincides with the GLS estimator,

$\widehat{\rho}_{GLS}$ of this parameter, obtained in the context of the AR(1) model (2.37). Specifically,

$$\frac{d\ln L}{d\rho} = -\frac{1}{2}\sum_{t=1}^{T}\frac{-2(y_t-\rho y_{t-1})y_{t-1}}{\sigma_v^2 y_{t-1}^2\left[\sum_{i=0}^{T-1}\phi^{2i}+\frac{\phi^{2T}}{1-\phi^2}\right]+\sigma_u^2} = 0 \Rightarrow$$

$$\Rightarrow \quad \widehat{\rho}(\sigma_u^2,\phi,\sigma_v^2) = \frac{\sum_{t=1}^{T}\frac{y_t y_{t-1}}{\sigma_v^2 y_{t-1}^2\left[\sum_{i=0}^{T-1}\phi^{2i}+\frac{\phi^{2T}}{1-\phi^2}\right]+\sigma_u^2}}{\sum_{t=1}^{T}\frac{y_{t-1}^2}{\sigma_v^2 y_{t-1}^2\left[\sum_{i=0}^{T-1}\phi^{2i}+\frac{\phi^{2T}}{1-\phi^2}\right]+\sigma_u^2}} = \widehat{\rho}_{GLS}$$

Remark 3: If $\phi = 0$, we obtain the log-likelihood function for the random coefficient model

$$\ln L(\theta; y_t) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln\left\{\sigma_v^2 y_{t-1}^2+\sigma_u^2\right\} - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\rho y_{t-1})^2}{\sigma_v^2 y_{t-1}^2+\sigma_u^2} \ ,$$

and if, in addition, $\sigma_v^2 = 0$, then

$$\ln L(\theta; y_t) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln\sigma_u^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\rho y_{t-1})^2}{\sigma_u^2} \ ,$$

which is the log-likelihood function of the constant coefficient AR(1) model. If, however, $\sigma_v^2 = 0$ and $\phi \neq 0$, then we end up with an AR(1) model whose coefficient varies deterministically. In this case the log-likelihood function does not provide an estimator for ϕ, since it attains the same maximum for any value $|\phi| < 1$. In other words, this particular parameter configuration causes identification failure for ϕ.

Pagan (1980) shows that the two alternative formulations presented above are equivalent as long as $|\phi| < 1$, and the Kalman filter is initialized to $E(\rho_0) = 0$ and $var(\rho_0) = \frac{\sigma_v^2}{1-\phi^2}$. Moreover, he 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 AR(1)-AR(1) model, these conditions amount to: (i) model identification (this exludes the case $\sigma_v^2 = 0, \phi \neq 0$), (ii) stationarity of the state process, that is $|\phi| < 1$, (iii) second-order stationarity of $\{y_t\}$ (see Theorem 1) and (iv) the model parameters taking values inside the permissible parameter space.

2.4.2 The AR(1)-MA(1) Case

The AR(1)/MA(1) model presented in (2.10) is not in state space form and we need to rewrite it slightly:

$$y_t = \rho y_{t-1} + \begin{bmatrix} y_{t-1} & \theta y_{t-1} \end{bmatrix} \begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} + u_t \tag{2.44}$$

$$\xi_{t+1} = \begin{bmatrix} v_{t+1} \\ v_t \end{bmatrix} = F\xi_t + \tilde{v}_{t+1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} v_{t+1} \\ 0 \end{bmatrix} \tag{2.45}$$

with $u_t \sim IID(0, \sigma_u^2)$ and $\tilde{v}_{t+1} \sim IID(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, Q = \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & 0 \end{bmatrix})$.

We define first $\hat{\xi}_{t|t-1} = E(\xi_t|\mathcal{F}_{t-1}) = E(\begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix}|\mathcal{F}_{t-1}) = \begin{bmatrix} 0 \\ \hat{v}_{t-1|t-1} \end{bmatrix}$ and $P_{t|t-1} = var(\begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix}|\mathcal{F}_{t-1})$. Gaussian maximum likelihood estimation requires the derivation of the conditional mean and variance of y_t:

$$\begin{aligned} E(y_t|\mathcal{F}_{t-1}) &= E(\rho y_{t-1} + \begin{bmatrix} y_{t-1} & \theta y_{t-1} \end{bmatrix} \begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} + u_t|\mathcal{F}_{t-1}) \\ &= \rho y_{t-1} + \theta y_{t-1}\hat{v}_{t-1|t-1} \end{aligned} \tag{2.46}$$

$$\begin{aligned} var(y_t|\mathcal{F}_{t-1}) &= var(\rho y_{t-1} + \begin{bmatrix} y_{t-1} & \theta y_{t-1} \end{bmatrix} \begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} + u_t|\mathcal{F}_{t-1}) = \\ &= \begin{bmatrix} y_{t-1} & \theta y_{t-1} \end{bmatrix} P_{t|t-1} \begin{bmatrix} y_{t-1} \\ \theta y_{t-1} \end{bmatrix} + \sigma_v^2. \end{aligned} \tag{2.47}$$

Calculation of the above moments requires the extrapolation of prediction equations

$\hat{v}_{t-1|t-1}$ and $P_{t|t-1}$.These equations can be derived, using the Kalman filter techniques as described in the previous section. Joint normality between ξ_t and y_t takes the following form:

$$\left(\left|\begin{bmatrix}\xi_t\\ y_t\end{bmatrix}\right|\mathcal{F}_{t-1}\right)\sim N\begin{bmatrix}\hat{\xi}_{t|t-1}\\ E(y_t|\mathcal{F}_{t-1})\end{bmatrix},\begin{bmatrix}P_{t|t-1} & P_{t|t-1}\begin{bmatrix}y_{t-1}\\ \theta y_{t-1}\end{bmatrix}\\ \begin{bmatrix}y_{t-1} & \theta y_{t-1}\end{bmatrix}P_{t|t-1} & var(y_t|\mathcal{F}_{t-1})\end{bmatrix} \tag{2.48}$$

The above assumption allows us to conclude $\xi_t|\mathcal{F}_t \sim N(\hat{\xi}_{t|t}, P_{t|t})$ where

$$\hat{\xi}_{t|t}=\hat{\xi}_{t|t-1}+\{P_{t|t-1}\times\begin{bmatrix}y_{t-1}\\ \theta y_{t-1}\end{bmatrix}\left[\begin{bmatrix}y_{t-1} & \theta y_{t-1}\end{bmatrix}P_{t|t-1}\begin{bmatrix}y_{t-1}\\ \theta y_{t-1}\end{bmatrix}+\sigma_u^2\right]^{-1}\times \tag{2.49}$$
$$\times[y_t-\rho y_{t-1}-\begin{bmatrix}y_{t-1} & \theta y_{t-1}\end{bmatrix}\hat{\xi}_{t|t-1}]\}$$

and

$$P_{t|t}=P_{t|t-1}-\{P_{t|t-1}\times\begin{bmatrix}y_{t-1}\\ \theta y_{t-1}\end{bmatrix}\times \tag{2.50}$$
$$\times\left[\begin{bmatrix}y_{t-1} & \theta y_{t-1}\end{bmatrix}P_{t|t-1}\begin{bmatrix}y_{t-1}\\ \theta y_{t-1}\end{bmatrix}+\sigma_u^2\right]^{-1}\times\begin{bmatrix}y_{t-1} & \theta y_{t-1}\end{bmatrix}P_{t|t-1}\}=$$
$$=\begin{bmatrix}\sigma_v^2 & 0\\ 0 & P^{(1,1)}_{t-1|t-1}\end{bmatrix}-\frac{1}{y_{t-1}^2[\sigma_v^2+\theta^2P^{(1,1)}_{t-1|t-1}]+\sigma_u^2}\begin{bmatrix}(\sigma_v^2)^2y_{t-1} & \theta\sigma_v^2y_{t-1}^2P^{(1,1)}_{t-1|t-1}\\ \theta\sigma_v^2y_{t-1}^2P^{(1,1)}_{t-1|t-1} & \theta^2y_{t-1}^2[P^{(1,1)}_{t-1|t-1}]^2\end{bmatrix}$$

Then, standard Kalman filter updating equations, allow us to move on, calculating $\hat{\xi}_{t+1|t}$ and $P_{t+1|t}$:

$$\begin{aligned}\hat{\xi}_{t+1|t} &= \begin{bmatrix} 0 \\ \hat{v}_{t|t} \end{bmatrix} = F\hat{\xi}_{t|t} = \\ &= \begin{bmatrix} 0 \\ \sigma_v^2 y_{t-1} \frac{y_t - \rho y_{t-1} - \theta y_{t-1}\hat{v}_{t-1|t-1}}{y_{t-1}^2(\sigma_v^2 + \theta^2 P_{t-1|t-1}^{(1,1)}) + \sigma_u^2} \end{bmatrix}\end{aligned} \tag{2.51}$$

and

$$\begin{aligned}P_{t+1|t} &= FP_{t|t}F' + Q = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} P_{t|t}^{(1,1)} & P_{t|t}^{(1,2)} \\ P_{t|t}^{(2,1)} & P_{t|t}^{(2,2)} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & 0 \end{bmatrix} = \\ &= \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & P_{t|t}^{(1,1)} \end{bmatrix}\end{aligned} \tag{2.52}$$

where $P_{t|t}^{(i,j)}$ is the (i,j) element of the mean square error matrix $P_{t|t}$, available from result (2.50).

Since y_t is, conditional on $\mathcal{F}_{t-1}$, normally distributed and the moments in (2.46) and (2.47) can be reproduced using results (2.49)-(2.52), we can use Rosenberg *(state space form)* representation of the likelihood function of y_t to estimate unknown parameters ρ, θ, σ_u^2 and σ_v^2 :

$$\ln L(\theta; y_t) = -\frac{T}{2}\ln 2\pi \tag{2.53}$$

$$-\frac{1}{2}\sum_{t=1}^{T} \ln\left\{ y_{t-1}^2\sigma_v^2[1 + \theta^2(1 - \frac{y_{t-2}^2}{y_{t-2}^2[\sigma_v^2 + \theta^2 P_{t-2|t-2}^{(1,1)}] + \sigma_u^2} + \sigma_u^2 \right\} \tag{2.54}$$

$$-\frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \rho y_{t-1} - \theta y_{t-1}\hat{v}_{t-1|t-1})^2}{\left\{ y_{t-1}^2\sigma_v^2[1 + \theta^2(1 - \frac{y_{t-2}^2}{y_{t-2}^2[\sigma_v^2 + \theta^2 P_{t-2|t-2}^{(1,1)}] + \sigma_u^2} + \sigma_u^2 \right\}}. \tag{2.55}$$

2.5 Monte Carlo Study

Next, we conduct a Monte Carlo study in order to achieve two tasks: The first is to examine the finite sample performance of the maximum likelihood estimator together with that of the related test statistics for various parameter settings of theoretical and empirical interest. The second task is to investigate the extent to which the AIC, SIC and HQ information criteria are capable of detecting the correct model from a set of models that contains apart from AR(1)/AR(1) and AR(1)/MA(1), the AR(1) model with constant coefficient (AR(1)) and the AR(1) model with random coefficient (RCAR(1)) as well. The theoretical properties of the above mentioned criteria, for the case of linear models, have been thoroughly studied in the literature. Shibata (1976) obtains the asymptotic distribution of the AIC- selected order for the case of an autorgressive model of order p. Hurvich and Tsai (1989) demonstrate that AIC does not provide a consistent estimate of p, whereas SIC and HQ do so at the cost of asymptotic efficiency. Beran, Bhanasali and Ocker (1998) examine the problem of model choice for the class of stationary and nonstationary, fractional and nonfractional autoregressive process. They show that AIC is of the same general form as for stationary autoregressive processes but with the fractional order, d, of integration of the underlying process being treated as an additional parameter. They also show that as in the stationary case, AIC, as opposed to SIC and HQ, does not provide a consistent estimate of the true order of the model. Inconsistency of AIC, however, is not always treated as an unpleasant feature of the selection procedure, especially in cases where the "true" model is not expected to belong to $\mathcal{M}$. According to Shibata (1983), "Inconsistency does not imply a defect of the selection procedure, but rather the inevitable concomitant of balancing underfitting and overfitting risks".

Despite the extentive research on AIC, SIC and HQ available for the case of linear models, the properties of above mentioned criteria in nonlinear models, such the present AR(1)-AR(1) model, are relatively unknown. Priestley (1981, Ch. 11) has applied AIC in various nonlinear models, such as bilinear and threshold autoregressive ones, without providing any justification for its use. Auestad and Tjostheim (1990) have offered a heuristic argument, accompanied with Monte Carlo simulations, for using information criteria in the general nonlinear case. These authors have also acknowledged that the theoretical analysis of the order determination problem for nonlinear time series models is prohibitively difficult. Consequently, Monte Carlo simulations seem to be the only feasible method of investigating the problem at hand.

In all the simulations that follow, we assume that $u_t \sim IIDN(0, \sigma_u^2), \nu_t \sim IIDN(0, \sigma_\nu^2)$, and also that $\{u_t\}$ and $\{\nu_t\}$ are mutually independent. The number of replications is equal to 2000 and the sample size, T, is set equal to 250, 500, and 1000. We explored many alternative parameter settings, covering the majority of cases that present either theoretical or empirical interest. We report the results from the following three representative cases: (i) $(\rho, \psi, \sigma_u^2, \sigma_\nu^2) = (0.4, -0.2, 0.07, 0.43)$, (ii) $(\rho, \psi, \sigma_u^2, \sigma_\nu^2) = (0.6, 0.3, 0.5, 0.1)$ and (iii) $(\rho, \psi, \sigma_u^2, \sigma_\nu^2) = (0.6, -0.5, 0.07, 0.70)$ where $\psi = \{\phi, \theta\}$ depending on the type of model that we simulated.

The values of the first case are representative of the corresponding ML estimates obtained for daily interest rate series (see next section). The second case is characterised by a relatively small state variance. This may cause problems to the ML estimator, since σ_ν^2 is close to the frontier of the parameter space. The parameters in the third case produce a value violates the necessary conditions for stabilitity, both in AR(1)/AR(1) case (R is equal to 1.293) and in AR(1)/MA(1) case (largest eigenvalue of matrix A is 1.175) The results for all cases are reported in the Tables presented in the Appendix. The first and second panels of these tables report the average bias, standard deviation, skewness and kurtosis coefficients of the ML estimators and the associated t-statistics respectively. In addition, the 5% empirical sizes of the t-statistics are also presented. Finally the last panels of the tables report the frequencies at which the information criteria select the correct model among the AR(1), AR(1)-RC, AR(1)-MA(1) and AR(1)-AR(1) ones.

The ML estimators of all the four parameters for both models of consideration, work remarkaby well for all the parameter configurations under study, including those which are near the boundary of the parameter space (Case ii) or those violating the necessary condition for stability (Case iii). 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 standard deviation of $\widehat{\phi}$ is equal to 0.112, 0.207 and 0.058 for

Cases (i), (ii) and (iii), respectively. In the case of AR(1)/MA(1) model, $\hat{\theta}$ slightly less well with standard deviations equal to 0.145, 0.348 and 0.116 for Cases (i) to (iii). When the sample size increases to $T = 1000$, the corresponding standard deviations decrease to 0.079, 0.153 and 0.039 for $\hat{\phi}$ and to 0.088, 0.306 and 0.072 for $\hat{\theta}$. Moreover, $\widehat{\rho}, \widehat{\phi}$ and $\hat{\theta}$ appear to be normally distributed for the majority of the cases under study. These results suggest that the ML estimators may perform well even in the absence of stationarity of $\{y_t\}$.

Turning to hypothesis testing on the parameters of the AR(1)-AR(1) model, the densities of the t-ratios of $\hat{\rho}$ and $\hat{\phi}$ estimators under study come reasonably close to the standard normal density, even for $T = 250$. This results in fairly accurate statistical inferences, as suggested by the fact that the empirical sizes of the tests are very close to their nominal values. It is interesting to note that the good behaviour of the t-tests is observed even for cases in which the state variance σ_ν^2 is relatively small (e.g. Case II). However, as the state variance decreases further, size distortions start to appear, which become stronger as σ_ν^2 approaches zero. For example, in the context of Case II and for $T = 250$, a decrease in the state variance from $\sigma_\nu^2 = 0.43$ to $\sigma_\nu^2 = 0.1$, results in a large increase in the empirical size of the t-test on ϕ, from 5.5% to 19.1%. In the case of AR(1)/MA(1) models, in all cases the t-ratios of $\hat{\rho}$ appear to behave as expected and have an empirical distribution close to the standard Normal even for small samples. However, the t-ratios and the empirical sizes of $\hat{\theta}$ work well only in the nonstationary y_t setting - Case (iii). In the other cases, it appears that t-ratios need larger samples to approach asymptotic Normality.

The performance of information criteria under study depends heavily on the size of the estimation sample. The ability of AIC,SIC and HQ criteria to detect the correct model increases as the we increase the sample size. However, these criteria although it appears to perform well in detecting AR(1)/AR(1) and AR(1)/MA(1) models even for small samples, they fail in distinguishing among each other. Monte Carlo results suggest that AIC exhibits the best performance followed by SIC. On the contrary, the performance of HQ is rather poor. These results agree with the ones established for linear cases, according to which when the "dimension" of the true model is large relative to those of its competitors, the probability of underfitting by means of SIC or HQ is rather high.

Specifically, for Case (ii) and $T = 250$, when the true model is AR(1)/AR(1), the percent selections of the correct model is 31%, 22% and 12% for AIC, SIC and HQ, respectively. These numbers increases in the case $T = 1000$ to 50%,43% and 26% respectively and if include the times that these criteria selects alternatively the AR(1) / MA(1) model, the cumulative performance of AIC, SIC and HQ in detecting random autoregressive coefficients that exchibit in general some type of dependence, increases since the percent selection of these models is 90.5%, 74.5 and 48.5% respectively. When the values of ϕ, θ or σ_ν^2 are relatively large (e.g. Case ii), the frequencies at which the true model is detected approach 100% even for $T = 500$.

Selecting between the AR(1)/MA(1) and AR(1)-AR(1) models may alternatively be accomplished by means of the t-test on ϕ and θ. It has already been established that the rejection frequencies of the true null hypothesis $\phi = 0$ are close to their nominal levels for the majority of cases under examination. When $\phi \neq 0$ the percent selections of the true AR(1)-AR(1) model are determined by the power properties of the t-test for the false hypothesis $\phi = 0$. These properties, in turn, depend on the true value of ϕ and the sample size. For example, for $(\rho, \phi, \sigma_u^2, \sigma_\nu^2) = (0.2, 0.2, 1, 0.4)$, the power of the t-test to reject the hypothesis $\phi = 0$ is 20% and 72% for $T = 250$ and 1000, respectively.

The remarkable performance of the ML estimates and the information criteria in cases where the conditions of stability are violated (e.g. Case iii) may seem surprising as first sight. However, it must be noted that this condition is too restrictive in the case of finite samples. Specifically, there might be cases in which necessary conditions are violated and the finite realizations of $\{y_t\}$ closely resemble those of a second-order stationary process. In order for a finite realization of $\{y_t\}$ to exhibit distinct non-stationary characteristics, there must be a consecutive number of time periods for which the values of $\rho + \rho_t$ are large. On the contrary if large values of $\rho + \rho_t$ occur only at isolated time periods, far apart from each other, then the realizations of the non-stationary $\{y_t\}$ are likely to be similar to those of a stationary process.

2.6 Empirical Results

This section presents empirical evidence suggesting that the AR(1) / AR(1) and AR(1) / MA(1) models describe many financial time series better than either the AR(1) or the AR(1)-RC models. This evidence is supported by both in-sample criteria and out-of-sample performance measures.

Specifically, we employ daily data for 9 stock indices, (namely DOW JONES COMP. AVG, S&P 100 INDEX, S&P 500 INDEX, NYSE COMPOSITE INDEX, NASDAQ COMPOSITE INDEX, DAX INDEX, NIKKEI 225, HANG SENG INDEX and ALL ORDINARIES INDX) obtained from Bloomberg Each series satisfy the minimum observation of at least 2000 observations prior to 01/01/1990 in order to have sufficient long history to achieve credible back-testing results

The four basic competing models, namely AR(1), AR(1)/RC, AR(1)/MA(1) and AR(1)/AR(1) but we include in the comparison the performance an AR(1)-GARCH(1,1) model which is widely used in financial time series. Starting in 01/01/1990 all models. are estimated for the logarithmic differences of each series Specifically, each model is estimated using data up to the period T and one-step-ahead forecasts are generated. Then $T+1$ data are added to the estimation sample, the models are re-estimated and new one-step-ahead forecasts are generated and compared with realized returns at $T+1$. This procedure continues until forecasts for the last period in our sample are produced. At the end,, the forecasting accuracy of each model is measured by the usual statistics, namely the mean absolute error (MAE) and the root mean squared error (RMSE). We also calculate the corresponding Akaike information criterion for all models under consideration and record which model was selected at each step (AR(1)-GARCH(1,1) is not included in this competition among the other models, because it is not nested with them). The results, reported in Tables (2.2) and (2.1) may be summarized as follows:

(i) AR(1)-MA(1) is selected as the best model by AIC,,most of the times for all stock indices under study.The constant AR(1) model is never chosen as the most appropriate model.

(ii) In terms of the Mean Absolute Error, all models under consideration, perform equivalently

(iii) Root Mean Square Error suggest cases where the best forecasting models are the autoregressive models with random coefficients that exhibits dependence. For example, the RMSE IS 0.95, 1.09 and 1.00 for AR(1), AR(1)-RC and AR(1)-GARCH models for SP500 Index, while AR(1)-MA(1) and AR(1)-AR(1) models have RMSE equal to 0.83 and 0.84 respectively.

Table 2.1: Forecasting Performance Evaluation of Studied Models, using Root Mean Square Error and Mean Absolute Error Measures

Stock Index		Estimated Model: AR	AR-RC	AR-MA	AR-AR	AR GARCH
DOW JONES	RMSE	1.46	1.44	1.48	1.47	1.45
	MAE	0.78	0.77	0.78	0.78	0.78
SP 100 INDEX	RMSE	1.12	1.21	0.99	1.00	1.15
	MAE	0.74	0.74	0.72	0.72	0.74
SP 500 INDEX	RMSE	0.95	1.09	0.83	0.84	1.00
	MAE	0.70	0.72	0.68	0.68	0.71
NYSE	RMSE	1.65	1.66	1.67	1.66	1.66
	MAE	0.84	0.84	0.85	0.85	0.84
NASDAQ	RMSE	2.44	2.87	2.29	2.36	2.69
	MAE	0.98	1.00	0.97	0.97	1.00
DAX INDEX	RMSE	1.26	1.28	1.24	1.24	1.25
	MAE	0.84	0.84	0.83	0.83	0.84
NIKKEI 225	RMSE	2.87	2.78	2.78	2.78	2.82
	MAE	1.24	1.22	1.22	1.22	1.23
HANG SENG	RMSE	2.30	2.34	2.37	2.36	2.33
	MAE	1.16	1.15	1.17	1.16	1.15
ALL ORDINARIES	RMSE	0.85	0.85	0.86	0.85	0.85
	MAE	0.67	0.68	0.67	0.67	0.67

2.7 Conclusions.

In this chapter we investigated the issue of stability of an AR(1) process with a stochastic, serially correlated coefficient, which was assumed to follow either an AR(1) or a MA(1) process. In the case of the AR(1) coefficient, it was shown that the problem of deriving empirically useful, exact sufficient conditions for the stability of the process is prohibitively difficult. However, we were able to derive the necessary and sufficient conditions for the

Table 2.2: Percentages of Model's Selection, using AIC

Stock Index	AR(1)	AR-RC	AR-MA	AR-AR
DOW JONES COMP	0	1.96	90.20	7.84
SP 100 INDEX	0	5.88	94.12	0.00
SP 500 INDEX	0	1.96	98.04	0.00
NYSE COMPOSITE INDEX	0	1.96	96.08	1.96
NASDAQ COMPOSITE INDEX	0	0.00	88.24	11.76
DAX INDEX	0	3.92	80.39	15.69
NIKKEI 225	0	14.00	72.00	16.00
HANG SENG INDEX	0	0.00	98.00	2.00
ALL ORDINARIES INDX	0	17.65	82.35	0.00

stability of the process for the case of the MA(1) coefficient.

2.8 Appendix

Table 2.3: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (i) , T=250, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.4,-0.2,0.07,0.43

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	-0.001	-0.013	0.002	-0.025
St.Dev.	0.071	0.170	0.015	0.116
Skewness	-0.256	-0.327	0.454	-0.047
Kurtosis	2.587	4.225	3.027	2.738

t-statistic of:	ρ	ϕ
Average	0.009	-0.100
St.Dev.	0.988	1.125
Skewness	-0.149	-0.963
Kurtosis	2.756	8.584

Empirical Size	4.00%	5.50%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	1.00%	5.00%
AR(1)-RC	48.00%	62.00%	72.00%
AR(1)-MA(1)	31.00%	22.00%	12.50%
AR(1)-AR(1)	21.00%	15.00%	10.50%

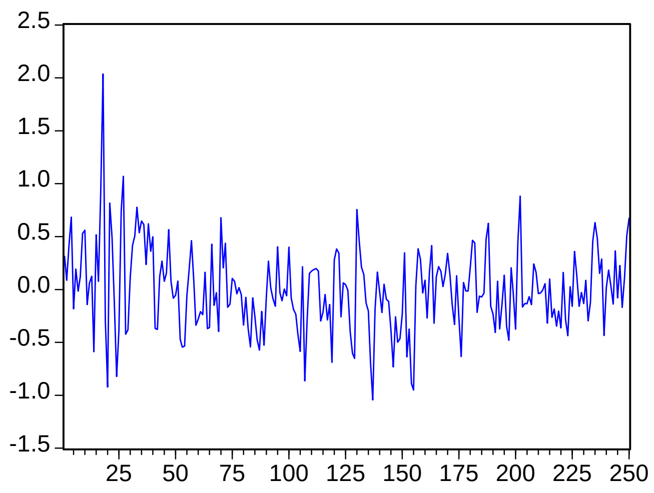


Figure 2.3: Realization of AR(1)/AR(1) Case (i) T=250

Table 2.4: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (i), T=500, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.4,-0.2,0.07,0.43

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	0.000	-0.010	0.000	-0.015
St.Dev.	0.047	0.112	0.010	0.083
Skewness	0.049	-0.285	0.289	-0.270
Kurtosis	2.941	3.468	3.373	2.938

t-statistic of:	ρ	ϕ
Average	0.007	-0.090
St.Dev.	0.941	1.004
Skewness	0.079	-0.177
Kurtosis	2.983	3.680

Empirical Size	3.00%	7.00%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	28.00%	45.00%	67.00%
AR(1)-MA(1)	37.50%	30.50%	19.00%
AR(1)-AR(1)	34.50%	24.50%	14.00%

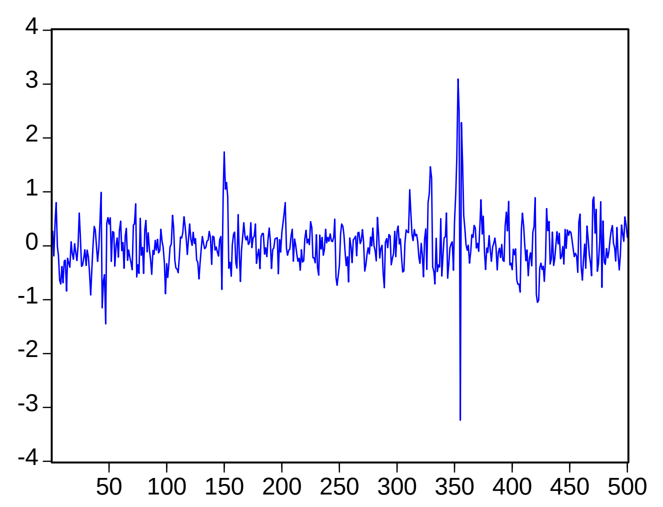


Figure 2.4: Realization of AR(1)/AR(1) Case (i) T=500

Table 2.5: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (i), T=1000, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.4,-0.2,0.07,0.43

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	-0.003	0.001	0.001	-0.014
St.Dev.	0.039	0.079	0.007	0.058
Skewness	-0.021	0.069	0.233	-0.259
Kurtosis	2.964	3.429	3.345	3.213

t-statistic of:	ρ	ϕ
Average	-0.080	0.011
St.Dev.	1.084	1.034
Skewness	0.037	-0.035
Kurtosis	2.938	3.360

Empirical Size	6.50%	7.50%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	9.50%	25.50%	51.50%
AR(1)-MA(1)	50.00%	43.00%	26.00%
AR(1)-AR(1)	40.50%	31.50%	22.50%

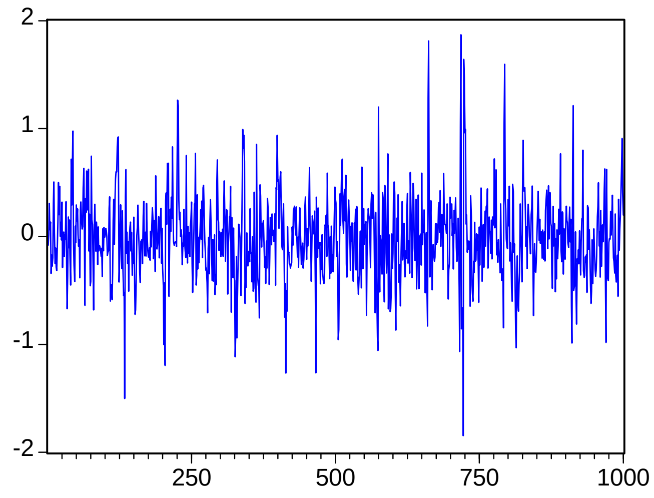


Figure 2.5: Realization of AR(1)/AR(1) Case (i) T=1000

Table 2.6: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (ii), T=250, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.6,-0.3,0.5,0.1

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	0.001	-0.024	0.002	-0.002
St.Dev.	0.063	0.260	0.061	0.050
Skewness	0.004	-0.064	0.281	0.285
Kurtosis	2.772	2.346	2.804	2.645

t-statistic of:	ρ	ϕ
Average	0.047	0.083
St.Dev.	0.922	0.947
Skewness	0.184	1.134
Kurtosis	2.988	5.639

Empirical Size	7.04%	19.1%

Model Selection	AIC	SC	HQ
AR(1)	33.93%	52.68%	71.88%
AR(1)-RC	36.61%	30.36%	23.21%
AR(1)-MA(1)	14.29%	9.38%	1.34%
AR(1)-AR(1)	15.18%	7.59%	3.57%

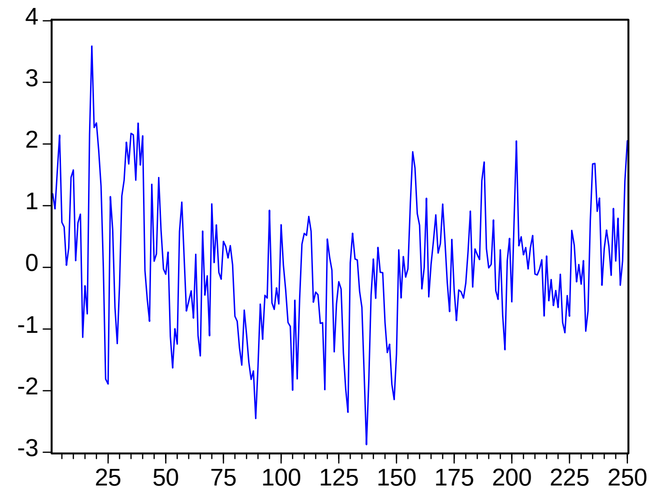


Figure 2.6: Realization of AR(1)/AR(1) Case (ii) T=250

Table 2.7: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (ii), T=500, $\{\phi, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,-0.3,0.5,0.1

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	0.001	-0.046	-0.001	-0.002
St.Dev.	0.048	0.207	0.042	0.038
Skewness	-0.278	-0.047	0.378	0.238
Kurtosis	2.412	2.621	3.370	2.975

t-statistic of:	ρ	ϕ
Average	0.067	-0.089
St.Dev.	0.995	0.933
Skewness	-0.092	0.641
Kurtosis	2.263	4.358

Empirical Size	5.50%	5.50%

Model Selection	AIC	SC	HQ
AR(1)	8.33%	25.00%	42.16%
AR(1)-RC	49.51%	53.92%	50.00%
AR(1)-MA(1)	17.65%	8.82%	3.43%
AR(1)-AR(1)	24.51%	12.25%	4.41%

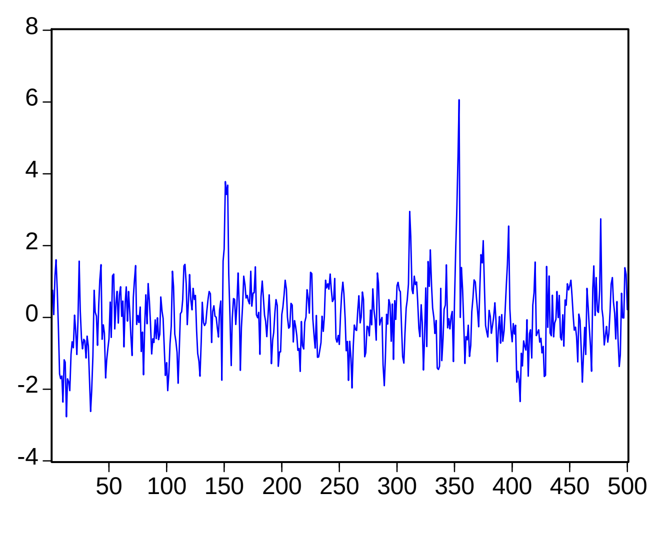


Figure 2.7: Realization of AR(1)/AR(1) Case (ii) T=500

Table 2.8: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (ii), T=1000, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.6,-0.3,0.5,0.1

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	0.002	-0.008	0.004	-0.004
St.Dev.	0.033	0.153	0.029	0.029
Skewness	-0.041	-0.247	0.048	0.227
Kurtosis	2.532	2.987	3.712	3.386

t-statistic of:	ρ	ϕ
Average	0.090	0.039
St.Dev.	0.965	1.006
Skewness	0.064	0.303
Kurtosis	2.553	3.124

Empirical Size	3.50%	7.50%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	4.50%	13.00%
AR(1)-RC	32.50%	51.00%	64.00%
AR(1)-MA(1)	29.00%	19.50%	8.00%
AR(1)-AR(1)	38.50%	25.00%	15.00%

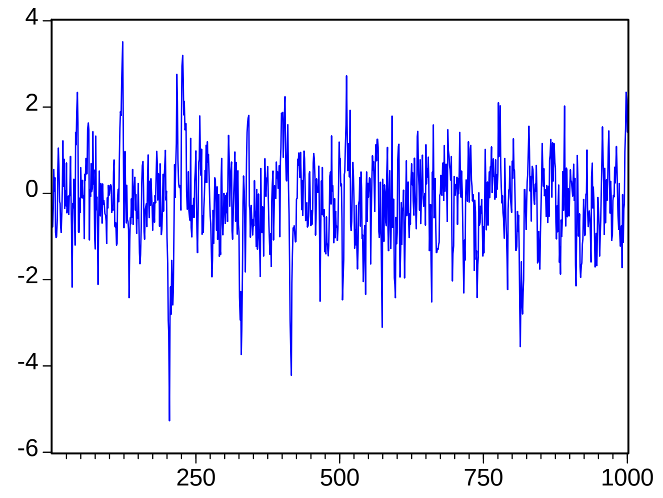


Figure 2.8: Realization of AR(1)/AR(1) Case (ii) T=1000

Table 2.9: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (iii), T=250, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.6,-0.5,0.07,0.70

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	0.003	0.005	0.001	-0.015
St.Dev.	0.060	0.078	0.012	0.110
Skewness	-0.237	0.343	0.135	-0.040
Kurtosis	2.908	2.923	2.720	2.927

t-statistic of:	ρ	ϕ
Average	0.066	-0.027
St.Dev.	0.982	0.947
Skewness	-0.088	-0.177
Kurtosis	2.875	3.342

Empirical Size	4.50%	5.50%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	0.00%	0.00%	0.00%
AR(1)-MA(1)	11.00%	11.00%	11.00%
AR(1)-AR(1)	89.00%	89.00%	89.00%

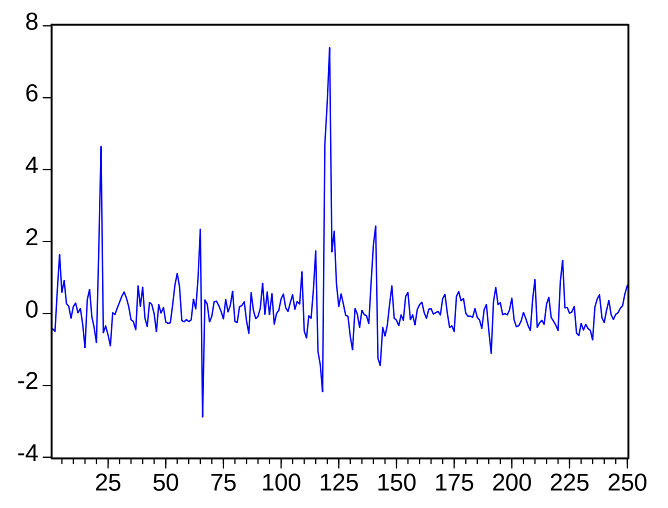


Figure 2.9: Realization of AR(1)/AR(1) Case (iii) T=250

Table 2.10: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (iii), T=500, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.6,-0.5,0.07,0.70

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	-0.003	-0.003	0.000	-0.009
St.Dev.	0.042	0.058	0.008	0.086
Skewness	0.063	0.204	0.528	0.075
Kurtosis	2.620	2.738	3.762	3.025

t-statistic of:	ρ	ϕ
Average	-0.052	-0.137
St.Dev.	1.003	1.027
Skewness	0.108	-0.124
Kurtosis	2.661	2.945

Empirical Size	6.00%	5.50%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	0.00%	0.00%	0.00%
AR(1)-MA(1)	97.50%	97.50%	97.50%
AR(1)-AR(1)	2.50%	2.50%	2.50%

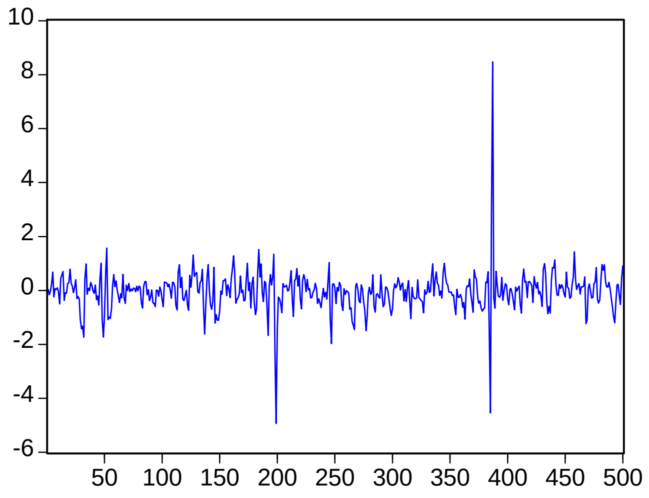


Figure 2.10: Realization of AR(1)/AR(1) Case (iii) T=500

Table 2.11: Monte Carlo Experiment: AR(1) / AR(1) Process - Case (iii), T=1000, $\{\rho, \phi, \sigma_v^2, \sigma_u^2\}$=0.6,-0.5,0.07,0.70

Bias of:	ρ	ϕ	σ_u^2	σ_v^2
Average	-0.004	0.001	0.001	-0.006
St.Dev.	0.032	0.039	0.006	0.055
Skewness	0.028	0.178	0.341	-0.178
Kurtosis	3.077	2.896	3.233	3.557

t-statistic of:	ρ	ϕ
Average	-0.124	-0.035
St.Dev.	1.076	1.000
Skewness	0.122	-0.114
Kurtosis	3.157	3.158

Empirical Size	7.00%	5.00%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	0.00%	0.00%	0.00%
AR(1)-MA(1)	0.50%	0.50%	0.50%
AR(1)-AR(1)	99.50%	99.50%	99.50%

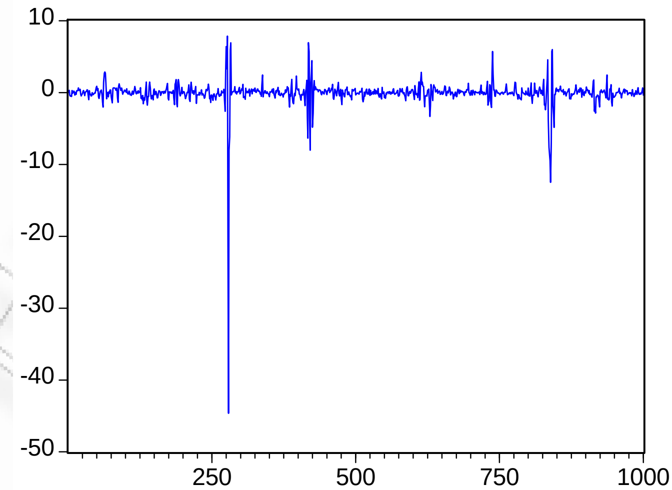


Figure 2.11: Realization of AR(1)/AR(1) Case (iii) T=1000

Table 2.12: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (i), T=250, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.4,-0.2,0.07,0.43

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	0.002	-0.035	0.001	-0.019
St.Dev.	0.071	0.209	0.014	0.111
Skewness	-0.251	-0.557	0.394	0.144
Kurtosis	2.781	4.619	2.827	2.416

t-statistic of:	ρ	θ
Average	0.048	0.040
St.Dev.	0.997	0.838
Skewness	-0.090	0.588
Kurtosis	3.011	2.812

Empirical Size	5.00%	1.50%

Model Selection	AIC	SC	HQ
AR(1)	0.46%	1.39%	5.56%
AR(1)-RC	47.22%	61.57%	73.61%
AR(1)-MA(1)	37.04%	25.46%	14.81%
AR(1)-AR(1)	15.28%	11.57%	6.02%

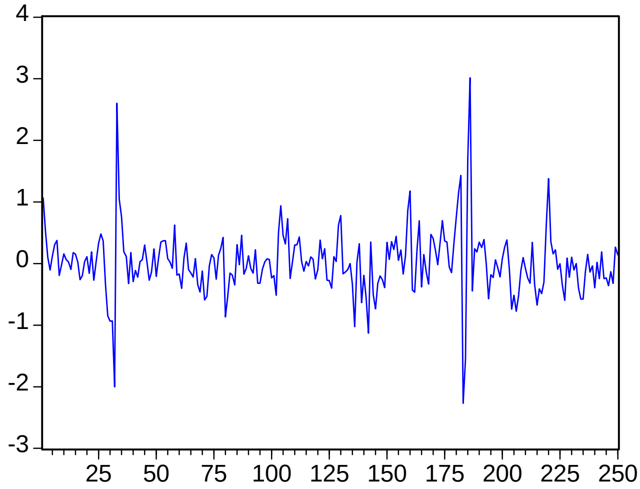


Figure 2.12: Realization of AR(1)/MA(1) Case (i) T=250

Table 2.13: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (i), T=500, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.4,-0.2,0.07,0.43

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	0.001	-0.031	0.000	-0.018
St.Dev.	0.048	0.145	0.010	0.084
Skewness	0.032	-0.936	0.190	-0.379
Kurtosis	2.951	4.487	3.404	3.265

t-statistic of:	ρ	θ
Average	0.024	-0.028
St.Dev.	0.951	0.882
Skewness	0.044	0.539
Kurtosis	2.980	3.111

Empirical Size	3.50%	2.00%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	30.20%	48.51%	68.81%
AR(1)-MA(1)	45.05%	37.13%	21.78%
AR(1)-AR(1)	24.75%	14.36%	9.41%

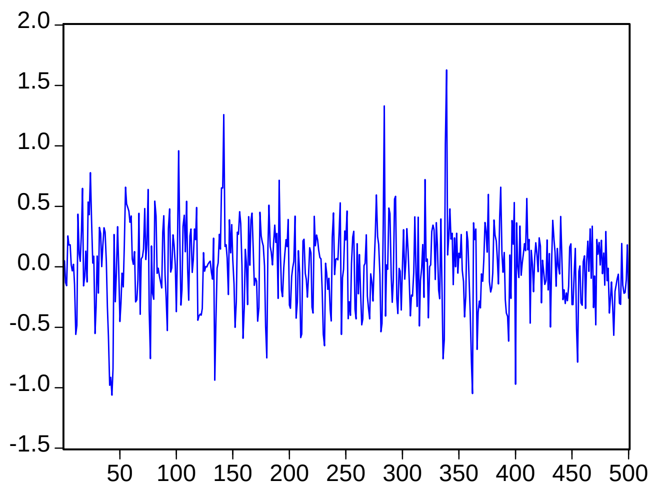


Figure 2.13: Realization of AR(1)/MA(1) Case (i) T=500

Table 2.14: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (i), T=1000, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.4,-0.2,0.07,0.43

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	-0.003	-0.001	0.001	-0.015
St.Dev.	0.039	0.088	0.007	0.061
Skewness	-0.012	-0.259	0.250	-0.349
Kurtosis	2.862	4.065	3.342	3.307

t-statistic of:	ρ	θ
Average	-0.065	0.058
St.Dev.	1.095	0.976
Skewness	0.059	0.316
Kurtosis	2.800	3.228

Empirical Size	7.00%	5.00%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	11.50%	29.00%	54.50%
AR(1)-MA(1)	54.50%	42.50%	28.50%
AR(1)-AR(1)	34.00%	28.50%	17.00%

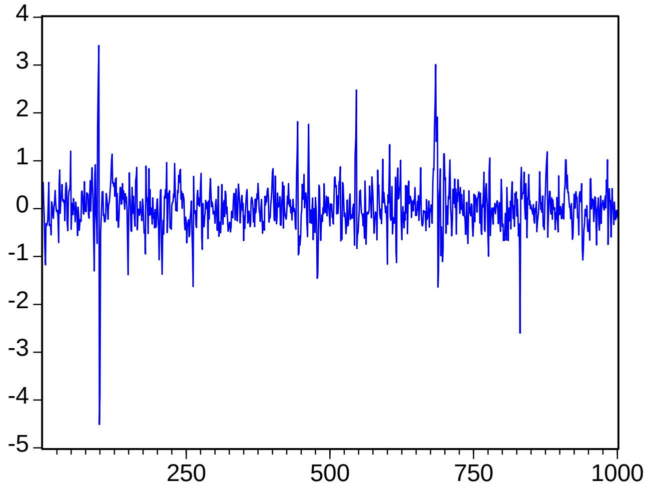


Figure 2.14: Realization of AR(1)/MA(1) Case (i) T=1000

Table 2.15: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (ii), T=250, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,0.3,0.5,0.1

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	0.000	0.115	0.002	-0.007
St.Dev.	0.063	0.448	0.061	0.051
Skewness	0.050	0.111	0.198	0.663
Kurtosis	2.907	1.641	2.628	3.187

t-statistic of:	ρ	θ
Average	0.040	-0.263
St.Dev.	0.945	0.502
Skewness	0.255	-0.966
Kurtosis	3.031	3.577

Empirical Size	6.50%	0.00%

Model Selection	AIC	SC	HQ
AR(1)	33.48%	54.35%	74.35%
AR(1)-RC	38.70%	31.74%	22.61%
AR(1)-MA(1)	16.09%	9.13%	0.87%
AR(1)-AR(1)	11.74%	4.78%	2.17%

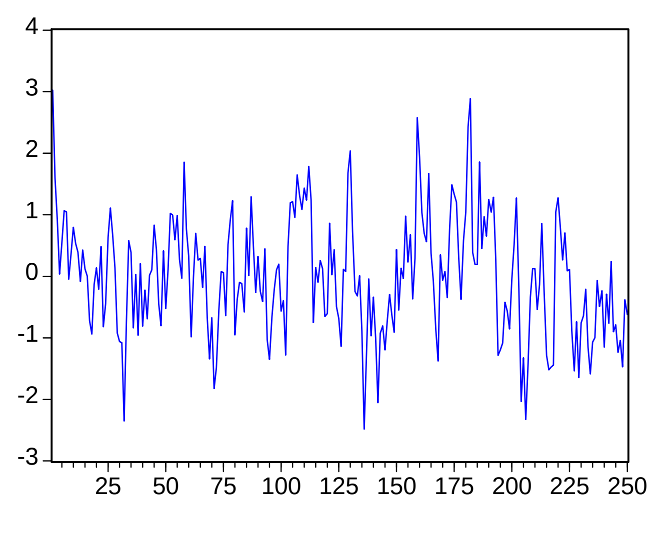


Figure 2.15: Realization of AR(1)/MA(1) Case (ii) T=250

Table 2.16: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (ii), T=500, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,0.3,0.5,0.1

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	0.004	0.081	0.000	-0.012
St.Dev.	0.046	0.387	0.042	0.039
Skewness	-0.048	0.347	0.152	-0.053
Kurtosis	3.043	2.118	3.280	2.276

t-statistic of:	ρ	θ
Average	0.115	-0.311
St.Dev.	0.982	0.636
Skewness	0.150	-1.000
Kurtosis	2.899	3.569

Empirical Size	5.50%	0.50%

Model Selection	AIC	SC	HQ
AR(1)	13.53%	26.57%	49.76%
AR(1)-RC	52.66%	58.45%	45.41%
AR(1)-MA(1)	20.77%	9.66%	2.90%
AR(1)-AR(1)	13.04%	5.31%	1.93%

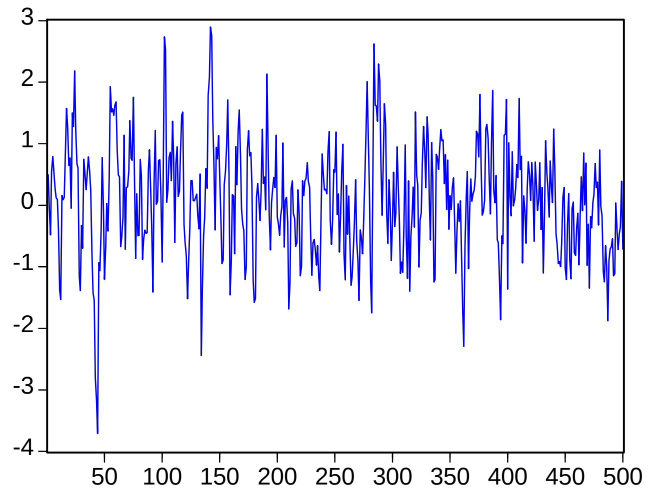


Figure 2.16: Realization of AR(1)/MA(1) Case (ii) T=500

Table 2.17: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (ii), T=1000, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,0.3,0.5,0.1

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	-0.002	0.076	0.002	-0.010
St.Dev.	0.034	0.306	0.030	0.030
Skewness	0.118	0.592	-0.025	0.003
Kurtosis	2.853	2.857	3.063	2.393

t-statistic of:	ρ	θ
Average	-0.035	-0.213
St.Dev.	1.019	0.753
Skewness	0.247	-1.042
Kurtosis	2.859	3.694

Empirical Size	5.50%	2.50%

Model Selection	AIC	SC	HQ
AR(1)	1.49%	5.47%	16.92%
AR(1)-RC	45.77%	55.72%	66.17%
AR(1)-MA(1)	18.41%	13.93%	4.48%
AR(1)-AR(1)	34.33%	24.88%	12.44%

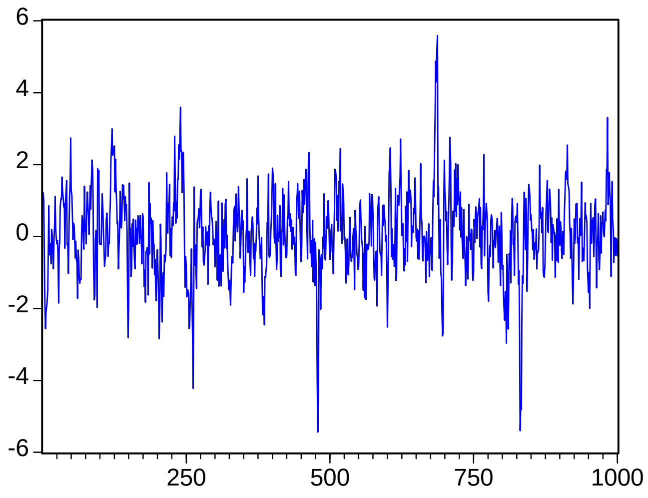


Figure 2.17: Realization of AR(1)/MA(1) Case (ii) T=1000

Table 2.18: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (iii), T=250, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,-0.5,0.07,0.70

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	-0.002	-0.087	0.001	-0.056
St.Dev.	0.060	0.222	0.012	0.155
Skewness	-0.156	-0.845	0.016	-0.080
Kurtosis	3.184	2.811	2.866	2.752

t-statistic of:	ρ	θ
Average	0.022	0.077
St.Dev.	1.122	0.814
Skewness	0.158	0.719
Kurtosis	3.603	3.457

Empirical Size	7.50%	4.50%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	0.00%	0.00%	2.36%
AR(1)-MA(1)	85.85%	85.85%	83.96%
AR(1)-AR(1)	14.15%	14.15%	13.68%

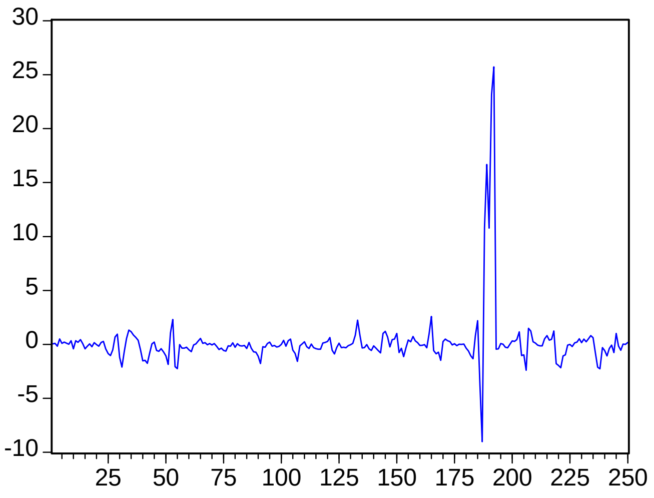


Figure 2.18: Realization of AR(1)/MA(1) Case (iii) T=250

Table 2.19: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (iii), T=500, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,-0.5,0.07,0.70

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	0.000	-0.032	0.001	-0.025
St.Dev.	0.039	0.116	0.008	0.097
Skewness	0.218	-1.055	0.363	-0.301
Kurtosis	3.083	5.806	3.007	3.480

t-statistic of:	ρ	θ
Average	0.011	-0.121
St.Dev.	1.026	0.952
Skewness	0.328	0.472
Kurtosis	3.377	3.142

Empirical Size	6.00%	5.00%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	0.00%	0.00%	0.00%
AR(1)-MA(1)	91.12%	91.12%	91.12%
AR(1)-AR(1)	8.88%	8.88%	8.88%

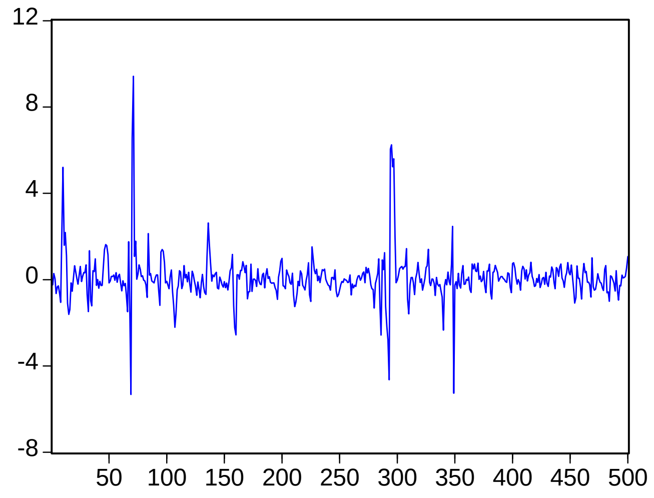


Figure 2.19: Realization of AR(1)/MA(1) Case (iii) T=500

Table 2.20: Monte Carlo Experiment: AR(1) / MA(1) Process - Case (iii), T=1000, $\{\rho, \theta, \sigma_v^2, \sigma_u^2\}$=0.6,-0.5,0.07,0.70

Bias of:	ρ	θ	σ_u^2	σ_v^2
Average	-0.004	-0.004	0.001	-0.011
St.Dev.	0.029	0.072	0.005	0.062
Skewness	-0.044	-0.494	0.288	-0.218
Kurtosis	2.915	4.035	3.454	2.884

t-statistic of:	ρ	θ
Average	-0.114	0.030
St.Dev.	1.031	1.023
Skewness	0.096	0.372
Kurtosis	2.784	3.619

Empirical Size	6.50%	7.00%

Model Selection	AIC	SC	HQ
AR(1)	0.00%	0.00%	0.00%
AR(1)-RC	0.00%	0.00%	0.00%
AR(1)-MA(1)	94.76%	94.76%	94.76%
AR(1)-AR(1)	5.24%	5.24%	5.24%

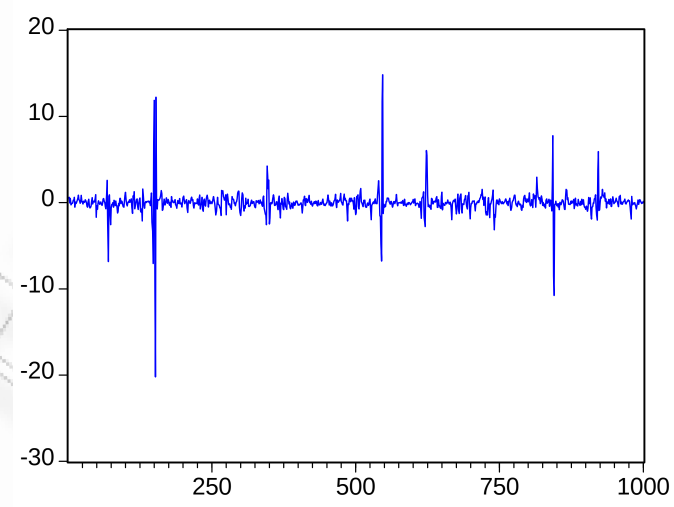


Figure 2.20: Realization of AR(1)/MA(1) Case (iii) T=1000

CHAPTER 3

Optimal Portfolios under Trending Second Moments of Asset Returns

3.1 Introduction

The behavior of volatility of stock returns over time has been extensively investigated by both the academic and investment communities in the last twenty five years or so. Currently, there is widespread agreement among researchers that this volatility has not remained constant over time. Various models for describing the time variation in volatility have been proposed in the literature, such as the well known GARCH and stochastic volatility models. These models treat the observed "volatility clustering" as non-linear dependence arising through the conditional variance of returns. This interpretation permits the underlying stochastic process, $\{R_t\}$, generating the returns to be strictly or even second-order stationary, since a time-varying conditional variance can coexist with a time-invariant unconditional variance. In other words, the observed time variation in the volatility of stock returns may be consistent with a stationary $\{R_t\}$, which exhibits second-order temporal dependence.

However, the aforementioned models are not capable of capturing all empirical characteristics the volatility of stock returns. For example, there are quite a few studies presenting evidence of variance breaks in $\{R_t\}$ (see, for example, Lamoureux and Lastrapes 1990, Stărică and Granger 2005). In fact, the high degree of persistence observed in the conditional variance process of stock returns may be the result of shifts in the unconditional variance of an otherwise locally stationary $\{R_t\}$, which (the shifts) have not been taken into account in the estimation of the conditional variance. The presence of variance breaks in $\{R_t\}$ imply that apart from conditional heteroscedasticity (non-linear dependence), the stock returns process is also characterized by unconditional heteroscedasticity (local time heterogeneity).

Campbell et al. (2001) suggest that the type of non-stationarity displayed by $\{R_t\}$ is more "global" than that implied by variance breaks. Specifically, these authors present evidence showing that the idiosyncratic component of the unconditional variance of the returns of individual firms exhibits a large positive linear trend over a 35-year period. The presence of such a trend is likely to dominate the behavior of the total firm volatility, thus producing a returns process which exhibits global non-stationarity. The latter is meant to imply that the marginal distributions of $\{R_t\}$ do not display intervals of time homogeneity (as in the case of local stationarity implied by variance breaks) but instead are continuously changing. This change, however, is not patternless but is governed by a systematic evolution of the variances of the marginal distributions of $\{R_t\}$.

In this chapter, we provide evidence showing that variance trends are present in other returns series, such as crops returns. Motivated by this piece of evidence, we formulate and solve the portfolio problem faced by an investor when the variances and the covariances of the returns of the available assets exhibit unbounded unconditional heteroscedasticity. In particular, we assume that the second moments of the joint distribution of the returns series are linear, or more generally, polynomial functions of time. Modeling volatility as a polynomial function of time may be thought of as providing a link between the bounded and infinite variance cases analyzed in the literature, since it permits the variances to be finite for any $t < \infty$, tending to infinity (not necessarily monotonically) as t grows larger.

3.2 Motivation

In this section, we show that the monthly percentage changes, r_{it}, in the price of three major crops, namely corn, soybeans and wheat, are characterized by trending variances. Figure 1 reports rolling estimates of the residual variance of an AR(1) model for r_{it}.

Figure 3.1: Rolling Estimation of Residuals Variance from an AR(1) Model. Starting Period 1990M1-1994M12 (60 Obs). Source: Bloomberg - S&P GS commodity indices - spot prices.

It can be seen that a clear upward, albeit non-monotonic trend is evident in all the three series under consideration. Moreover, we estimate the following auxiliary regression:

$$\varepsilon_{it}^2 = c + \sum_{j=1}^{4} \gamma_j \varepsilon_{it-j}^2 + \delta_{i1} t + \delta_{i2} t^2 + \nu_{it}$$

and test the joint hypothesis of no polynomial trends, i.e. $H_0 : \delta_{i1} = \delta_{i2} = 0$, $i = 1, 2, 3$ in the variance of the residuals ε_{it} from the AR(1) models. This hypothesis is strongly rejected for all the three cases under considerations with the p-values being less than 0.001 in all the cases.

3.2.1 Optimal Portfolios of Assets with Trending Volatilities

The next logical question concerns the implications of the trending variance hypothesis for optimal portfolio construction. In particular, assume that $\{\mathbf{R}_t\}$ denotes a $n-$dimensional vector stochastic process of the returns, R_{it}, $i = 1, 2, ..., n$ of n assets. The standard Markowitz procedure assumes that $\{\mathbf{R}_t\}$ is an independent and identically distributed (*iid*) process with mean vector, $\boldsymbol{\mu}$, and covariance matrix Σ. Based on the *iid* assumption, the portfolio $\mathbf{w} = [w_1, w_2, \ldots, w_n]'$ that minimizes the risk for a given level of expected return is time-invariant. The assumption of trending variances in stock returns violates the *iid* assumption, thus requiring a re-formulation of the optimization problem in the new framework. In the specification that follows, we shall retain the assumption of independence of $\{\mathbf{R}_t\}$ for reasons of simplicity. Specifically, we have,

$$\mathbf{R}_t = [R_{1t}, R_{2t}, \ldots, R_{nt}]' ,$$

$$E[\mathbf{R}_t] = \boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_n]' \ .$$

and

$$\mathbf{R}_t = \boldsymbol{\mu} + \mathbf{u}_t,$$

with $\mathbf{u}_t = [u_{1t}, u_{2t}, \ldots, u_{nt}]'$. The stochastic properties of $\{\mathbf{R}_t\}$ are determined by those of $\{\mathbf{u}_t\}$. In particular, we assume that $\{\mathbf{u}_t\}$ is an independent process with $E[\mathbf{u}_t] = \mathbf{0}$. Moreover, we assume that the covariance matrix, Q_t, of $\mathbf{u}_t$ changes with time according to

$$Q_t = (q_{ij,t})_{1\leq i,j\leq n} = F_t \bullet \Sigma \ , \tag{3.1}$$

$$F_t = (f_{ij}(t))_{1\leq i,j\leq n} \, , \, \Sigma = (\sigma_{ij})_{1\leq i,j\leq n}$$

where "$\bullet$" denotes the element-wise Hadamard product and $f_{ij}(t)$, $1 \leq i, j \leq n$ are functions of time, yet to be specified. This means that

$$q_{ij,t} = f_{ij}(t)\sigma_{ij},\ 1 \leq i, j \leq n.$$

Note that Q_t is also the covariance matrix of $\mathbf{R}_t$.

More specifically, we postulate the following model for the time-heterogeneity structure of $\mathbf{u}_t$:

$$\left.\begin{array}{c} \mathbf{u}_t = \mathbf{A}(t)\mathbf{v}_t\ , \\ \mathbf{A}(t) = diag\left\{\sqrt{f_1(t)}, \sqrt{f_2(t)}, \ldots, \sqrt{f_n(t)}\right\}, \\ f_i(t) = t^{k_i} + o\left(t^{k_i}\right),\ k_i \geq 0 \\ \mathbf{v}_t \sim iid(0, \Sigma) \end{array}\right\} \tag{3.2}$$

The model (3.2) implies that both the variances and the (absolute values of the) covariances of $\mathbf{R}_t$ are, in general, increasing functions of time. As a result, the optimal portfolio weights will vary over time as well. Specifically, assume that at period T, the typical investor wishes to determine the portfolio $\mathbf{w}_{pT} = [w_{1T}, w_{2T}, \ldots, w_{nT}]'$ that, for a given level of expected returns, minimizes the portfolio risk for period $T+1$. The solution of this optimization problem produces the following portfolio (vector of weights):

$$\mathbf{w}_{pT} = \left(\frac{C\mu_p - A}{D}\right) Q_{T+1}^{-1}\boldsymbol{\mu} + \left(\frac{B - A\mu_p}{D}\right) Q_{T+1}^{-1}\mathbf{1}\ , \tag{3.3}$$

where

$$\begin{aligned} A &= \mathbf{1}'Q_{T+1}^{-1}\boldsymbol{\mu}\ , \\ B &= \boldsymbol{\mu}'Q_{T+1}^{-1}\boldsymbol{\mu}\ , \\ C &= \mathbf{1}'Q_{T+1}^{-1}\mathbf{1}\ , \\ D &= BC - A^2 \text{ and} \\ \mathbf{1} &= [1, 1, \ldots, 1]' \in \mathbb{R}^n\ . \end{aligned}$$

For the practical implementation of solution (3.3), we need to obtain consistent estimates of $\boldsymbol{\mu}$ and Q_{T+1}.To this end, note that the standard sample covariance matrix estimator $\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t'$ diverges to infinity. The first step towards solving the estimation problem at hand, is to obtain a consistent estimator of k_i, $i=1,2,\ldots,n$. Such an estimator has been proposed by Kourogenis and Pittis (2008), namely:

$$\hat{k}_i = \frac{1}{\ln 2}\ln\left(\frac{\sum_{t=1}^{T}\hat{u}_{i,t}^2}{\sum_{t=1}^{[T/2]}\hat{u}_{i,t}^2}\right) - 1 \; , \tag{3.4}$$

where, for the case under study, $\hat{u}_{i,t}$ denotes the OLS residuals of the regression of $R_{i,t}$ on a constant. Next, we must obtain a consistent estimator of Σ. Let us denote by '$\xrightarrow{P}$' the convergence in probability. By utilizing Proposition 1 and Theorem 2 in Kourogenis and Pittis (2008), we obtain the following result:

Proposition 1: Under the specification (3.2), and if $k_i - \widehat{k}_i = o_p\left(\frac{1}{\ln T}\right)$, $i=1,2,\ldots,n$,

$$S_T \bullet \sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t' \xrightarrow{P} \Sigma \tag{3.5}$$

where

$$\widehat{\mathbf{u}}_t = \mathbf{R}_t - \widehat{\boldsymbol{\mu}} \; ,$$

$$\widehat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t$$

and

$$S_T = (s_{ij,T})_{1\le i,j\le n} = \left(\frac{\left(\widehat{k}_i+\widehat{k}_j\right)/2+1}{T^{\left(\widehat{k}_i+\widehat{k}_j\right)/2+1}}\right)_{1\le i,j\le n} .$$

Proof: Since

$$q_{ij,t} = t^{(k_i+k_j)/2}\sigma_{ij} + o(t^{(k_i+k_j)/2}) \; , \; 1\le i,j\le n \; ,$$

directly from Proposition 1 and Theorem 2 in Kourogenis and Pittis (2008) we have that and

$$\frac{1}{T^{\left(\widehat{k}_i+\widehat{k}_j\right)/2+1}}\sum_{t=1}^{T}\widehat{u}_{it}\widehat{u}_{jt} \xrightarrow{P} \frac{\sigma_{ij}}{(k_i+k_j)/2+1} \; , \; 1\le i,j\le n \; ,$$

which directly yields (3.5).

Proposition 1 allows us to estimate Σ, which in turn implies that a consistent estimate

$\widetilde{Q}_t$ of Q_t is feasible, in the sense that $\widetilde{Q}_t - Q_t \xrightarrow{P} 0$, provided that the exact form of F_t were known. However, we assume no a-priori knowledge of the exact functional forms f_i, $1 \leq i \leq n$, except from the fact that the degree k_i of the polynomial part, t^{k_i}, can be consistently estimated through (3.4). Since $f_i - t^{k_i} = o(t^{k_i})$, we are allowed to consider as an adequate approximation of $f(t+1)$ the value of $(t+1)^{\widehat{k}_i}$. A direct application of this approximation and Proposition 1 to (3.3) yields a feasible approximation of the optimal portfolio based on the extrapolated covariance matrix $\widehat{Q}_{T+1}$:

Corollary: The optimal portfolio, $\mathbf{w}_{pT}$, which minimizes the quantity

$$Var(r) = \mathbf{w}'_T Q_{T+1} \mathbf{w}_T \tag{3.6}$$

subject to

$$E[r] = \mathbf{w}'_T \boldsymbol{\mu} = \mu_p. \tag{3.7}$$

and

$$\mathbf{1}'\mathbf{w}_T = 1$$

is approximated by

$$\widehat{\mathbf{w}}_{pT} = \left(\frac{\widehat{C}\mu_p - \widehat{A}}{\widehat{D}} \right) \widehat{Q}_{T+1}^{-1} \widehat{\boldsymbol{\mu}} + \left(\frac{\widehat{B} - \widehat{A}\mu_p}{\widehat{D}} \right) \widehat{Q}_{T+1}^{-1} \mathbf{1} \ ,$$

where

$$\begin{aligned}
\widehat{Q}_{T+1} &= \left(\hat{q}_{ij,T+1}\right)_{1\le i,j\le n} \ , \\
\text{with } \hat{q}_{ij,T+1} &= \hat{\sigma}_{ij}\left(T+1\right)^{\left(\hat{k}_i+\hat{k}_j\right)/2} \\
&= \left(1+\frac{1}{T}\right)^{\left(\hat{k}_i+\hat{k}_j\right)/2} \frac{\left(\left(\hat{k}_i+\hat{k}_j\right)/2+1\right)}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt} \\
\widehat{A} &= \mathbf{1}'\widehat{Q}_{T+1}^{-1}\widehat{\boldsymbol{\mu}} \ , \\
\widehat{B} &= \boldsymbol{\mu}'\widehat{Q}_{T+1}^{-1}\widehat{\boldsymbol{\mu}} \ , \\
\widehat{C} &= \mathbf{1}'\widehat{Q}_{T+1}^{-1}\mathbf{1} \ , \\
\widehat{D} &= \widehat{B}\widehat{C}-\widehat{A}^2 \text{ and} \\
\mathbf{1} &= [1,1,\ldots,1]' \in \mathbb{R}^n \ .
\end{aligned}$$

Remarks:

1) The optimization problem defined above yields the one-period ahead optimal portfolio for a specific level of expected portfolio returns, thus determining the one-period ahead efficient frontier. Estimating Q_{T+1} by means of the sample covariance matrix $\widetilde{Q}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t'$ produces misleading estimates of the second moments of $\mathbf{R}_{T+1}$, since it ignores the presence of variance trends. In particular, if $k>0$ then $\widetilde{Q}_T \to \infty$.

2) The weights will not remain constant (and the optimal frontier too) due to time heterogeneity. This fact implies the importance of the stepwise recalculation of optimal weights. Moreover, since in the long run, the lower k(s) will yield significantly lower variances, an optimal portfolio chosen at time T with very long horizon, will consist only of the asset(s) that correspond to this lower k(s).

3) If $f_i(t) = t^{k_i} + o(1)$, $1 \le i \le n$, then from Proposition 1 we obtain $\widehat{Q}_{T+1} - Q_{T+1} \xrightarrow{P} 0$, which in turn implies that $\widehat{\mathbf{w}}_{pT} - \mathbf{w}_{pT} \xrightarrow{P} 0$ as $T \to \infty$.

4) It is easy to show that for $k_i < 1$, $1 \le i \le n$, we have $\widehat{\boldsymbol{\mu}} \xrightarrow{p} \boldsymbol{\mu}$ (see also Kourogenis and Pittis 2008).

3.3 The Modified Market Model

The Market Model, introduced by Fama (1976), is a model generated by statistical assumptions to relate the return of a security on the market portfolio. Fama, assumed joined Normality and IID asset returns to derive:

$$r_{it} = \alpha + \beta R_{mt} + u_{it}.$$

Normality is the crucial assumption in this derivation because it is characterized by the additivity property (sums of normal random variables are also normal) and the linearity of the conditional mean. Extensions of this simplified version of Market Model can be obtained if we relax the initial assumptions. As long as the distributional assumption, every other distribution that the above two properties are inherent, are valid candidates for generating a modified Market Model. Stable distributions, e.g., that allow for infinite variance, would generate a market model with no existing second moments. Heterogeneity and dependence assumptions will also result in Market Models with different features. We proceed now to examine what are the implications of the trending second moments assumption considered in this chapter in the properties of the Market Model.

We assume that we have n securities with the following distribution:

$$\begin{bmatrix} R_{1t} \\ R_{2t} \\ .. \\ R_{3t} \end{bmatrix} \sim N \begin{bmatrix} \mu_1 \\ \mu_2 \\ .. \\ \mu_n \end{bmatrix}, \begin{bmatrix} t^{k_1}\sigma_1^2 + o\left(t^{k_1}\right) & t^{\frac{k_1+k_2}{2}}\sigma_{12} + o\left(\frac{k_1+k_2}{2}\right) & \ldots \\ t^{\frac{k_2+k_1}{2}}\sigma_{21} + o\left(\frac{k_2+k_1}{2}\right) & t^{k_2}\sigma_2^2 + o\left(t^{k_2}\right) & \ldots \\ \ldots & \ldots & t^{k_n}\sigma_n^2 + o\left(t^{k_n}\right) \end{bmatrix}$$

Since the market portfolio is constructed by taking a weighted average of the n securities, the joint distribution of R_{it} , $i \in \{1, 2, ...n\}$ and $R_{mt} = \sum_{i=1}^{n} w_i R_{it}$ is bivariate Normal, by a known property of multivarate normality. This allows us to generate the regression and the skedastic function of R_{it} conditional on R_{mt}.

$$E(R_{it}|R_{mt}) = \left(\mu_i - \frac{cov(R_{it},R_{mt})}{var(R_{it)}}\right) + \frac{cov(R_{it},R_{mt})}{var(R_{it)}} R_{mt} = \alpha + \beta R_{mt}$$

$$var(R_{it}|R_{mt}) = var(R_{it,})(1 - corr^2(R_{it,}R_{mt}))$$

Therefore we can apply a statistical generating mechanism based on these results and produce a heteroskedastic market model where

$$R_{it} = \alpha + \beta R_{mt} + e_{it}$$

$$var(e_{it}) = var(R_{it}|R_{mt}) = var(R_{it)}(1 - corr^2(R_{it,}R_{mt})) =$$
$$= (t^{k_1}\sigma_i^2 + o\left(t^{k_1}\right))(1 - corr^2(R_{it,}R_{mt}))$$

. It is interesting to derive the properties of parameter β_{it}.

$$\begin{aligned}\beta_{it} &= \frac{cov(R_{it},R_{mt})}{var(R_{mt})} = \frac{cov(R_{it},\sum\limits_{i=1}^{n} w_i R_{it})}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{n} w_j w_i cov(R_{it},Rj)} = \\ &= \frac{\sum\limits_{j=1}^{n} w_j cov(R_{it},Rj)}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{n} w_j w_i cov(R_{it},Rj)} = \frac{\sum\limits_{j=1}^{n} w_j\left(t^{\frac{k_i+k_i}{2}} + o\left(\frac{ki+kj}{2}\right)\right)\sigma_{ij}}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{n}\left[w_j w_i\left(t^{\frac{k_i+k_j}{2}} + o\left(\frac{k_i+k_j}{2}\right)\right)\right]\sigma_{ij}}\end{aligned}$$

. Assume that $\max_{1\le j\le n}\{k_j\} = k^*$ and let $A = \{1 \le s \le n|k_s = k^*\}$ and $A' = \{1 \le s \le n|k_s \le k^*\}$. Then

$$\beta_{it} = \frac{\left(t^{k^*} + o\left(t^{k^8}\right)\right)\left[\sum\limits_{j\in A}^{n} w_j\sigma_{ij} + \sum\limits_{j\in A'}^{n} w_j o(1)\sigma_{ij}\right]}{\left(t^{k^*} + o\left(t^{k^8}\right)\right)\left[\sum\limits_{j\in A}^{n}\sum\limits_{i\in A}^{n} w_j w_i\sigma_{ij} + \sum\limits_{j\in A'}^{n}\sum\limits_{i\in A'}^{n} w_j w_i o(1)\sigma_{ij}\right]}$$

therefore

$$\beta_{it}^{*}=\frac{\sum_{j\in A}^{n} w_j\sigma_{ij}}{\sum_{j\in A}^{n}\sum_{i\in A}^{n} w_j w_i\sigma_{ij}} \underset{a}{\rightarrow} \begin{cases} 0 \text{ if } j\in A' \\ \beta_i^{*} \text{ if } j\in A \end{cases}$$

Remarks:

1) Trending unconditional second moments cause time varyibility for the parameter β. However, changes of β are not stochastic and follow a certain pattern to reach its limits.

2) In case of a common market shock that will cause similar trends for all securities, we can assume that set A will contain all n securities in which case $\beta_{it}^{*}=\beta_i$ (beta obtained from standard market models).

3) Stocks that exhibit abnormal increase in idiosyncratic volatility which are included in the construction of the market return, will cause movement of the level of betas for the rest of the stocks toward zero. This suggests that in case of increased volatility in some sectors of the market, we should construct seperate market portfolios to estimate betas for each group of different level of volatility,

3.4 Conclusions

In this chapter we have solved the portfolio problem for the case in which assets returns are characterized by unbounded heteroscedasticity. Specifically, it has been assumed that the second moments of returns evolve in a polynomial-like fashion, thus being asymptotically unbounded. It is shown that the optimal solution is a function of time depending on the orders k_i, $i=1,2,...,n$ at which the variances and covariances of asset returns grow over time. A feasible approximation to the optimal solution is obtained, which is based on the consistent estimator of k_i proposed by Kourogenis and Pittis (2008). This approximate solution is applicable in many cases of empirical interest, including that of farm planning, in which volatility trends are likely to be present.We also discussed the implications of trending second moments in the case of the market model were we noted that this kind of heteroskedasticity might generate time variation in parameter β of the market model.

CHAPTER 4

Aggregational Gaussianity and Barely Infinite Variance

4.1 Introduction

One of the most important questions in the financial literature concerns the distribution of financial prices. The interest for this question originated in the early 1950s with the detailed empirical study of Kendall (1953) on the statistical properties of a set of economic time series including commodity prices such as the Chicago wheat and New York cotton prices. This study was the first to notice that the empirical distributions of successive price changes deviate from normality mainly because they exhibit excess kurtosis. Then, the issue of leptokurtosis was taken up by Mandelbrot (1963) who put forward the idea that the observed leptokurtosis reflects the fact that the variance of commodity or stock price changes is infinite. More specifically, Mandelbrot observed that the logarithmic price changes within a specific period of time, say a day, is the sum of elementary logarithmic price changes, ξ_i, between transactions that occur in that day. He then assumed that the variance of these elementary price changes is infinite, which in turn implies that the Central Limit Theorem is not applicable. As a result, the sum of ξ_i's converges not to the normal distribution, but instead, to a Stable Paretian distribution. The latter is leptokurtic and has infinite variance.

An alternative explanation for the observed leptokurtosis in the empirical distributions of price changes was offered by, among others, Clark (1973), and Blattberg and Gonedes (1974). These studies attempt to explain leptokurtosis without sacrificing the finite-variance assumption. In particular, they put forward the idea that the transactions are not spread uniformly across time, which in turn implies that the underlying distribution of price changes is a mixture of normals.

The two competing explanations for leptokurtosis mentioned above bare different implications about the behavior of the distribution of logarithmic price changes as we move from higher (say daily) to lower (say monthly) frequencies of observations. In particular, it has been observed that as we move from higher to lower frequencies the degree of leptokurtosis diminishes and the empirical distributions tend to approximate normality. This stylized fact, referred to as “Aggregational Gaussianity”, can be accounted for only by the mixture of normals explanation of leptokurtosis and not by the infinite-variance alternative. Indeed, the stable-Paretian explanation is characterised by the property of “stability under addition” according to which if the daily price changes follow a stable Paretian distribution with characteristic exponent equal to a, then the monthly price changes also have to follow the same distribution. This in turn implies that the property of infinite variance cannot coincide with that of Aggregational Gaussianity.

In late 1980's, when a new class of models, namely the GARCH models, was put forward, the issue of the parallel existence of infinite variance and Aggregational Gaussianity re-emerged in the context of the estimates of the GARCH parameters. In particular, the estimation of GARCH models for commodity or stock price changes seemed to suggest (i) the presence of a unit root (or near-to-unit root) in the conditional variance, which gave rise to the so-called Integrated GARCH (IGARCH) models and (ii) the gradual declining of conditional heteroskedasticity and the associated leptokurtosis of the unconditional distribution as we move from higher to lower frequencies of observation (see Diebold 1988, Drost and Nijman 1993). Early enough, Nelson (1990) proved the stationarity of the IGARCH(1,1) model. Nevertheless, in vew of the fact that the presence of a unit root in the conditional variance implies that the unconditional distribution of IGARCH has infinite variance, a case in which the classical Central Limit Theorem (CLT) does not apply, the empirical studies seemed to suggest the simultaneous presence of two seemingly contradictory facts: Aggregational Gaussianity and infinite variance.

The above mentioned inconsistency between infinite variance and Aggregational Gaussianity motivated some authors to argue that the evidence of a unit root in the conditional variance was in fact spurious. It may arise either from structural breaks in the unconditional variance (see Diebold, 1986, Lamoureux and Lastrapes, 1990, and Diebold and Lopez, 1995) or from regime switching of the parameters of the conditional variance (see Fornari and Mele 1997, and Fong and See 2001, among others). Even in this case, however, an infinite variance stationary model may arise. For example, in a recent paper, Liu (2009) introduces the Integrated Markov Switching GARCH model, for which he proves stationarity and infinite variance. Liu's result implies that infinite variance can occur even if the conditional variance parameters in all-but-one regimes correspond to a finite variance model. Put differently, one and only regime (even with a very small probability of occurrence) with conditional variance parameters that do not correspond to a stable GARCH, can be the reason that a stationary and ergodic Markov Switching GARCH model has infinite variance.

In this chapter we aim at reconciling Aggregational Gaussianity and infinite variance without bringing in question the GARCH specification. We show that infinite variance and Aggregational Gaussianity can coexist, provided that all the moments of the unconditional distribution whose order is less than two exist. This moment condition is satisfied in the case of IGARCH processes, or put it differently, an IGARCH process is indeed a process with barely infinite variance (see Kourogenis and Pittis 2008). In other words, what we show in this chapter is that Aggregational Gaussianity can coexist with infinite variance, once the latter arises from a unit root in the conditional variance.

The chapter is organised as follows: In Section 2 we present evidence indicating that the price changes of six major crops, namely cocoa, coffee, corn, soybean, sugar and wheat, observed at high frequencies, seem to be characterised by both leptokurtosis and unit root in the conditional variance. We also show that both these effects tend to diminish as we move to lower frequencies. In Section 3 we explain why there is no paradox in admitting the simultaneous existence of Aggregational Gaussianity and IGARCH, by means of some limit theorems for mixing processes with barely infinite variance, developed in the probability theory over the last twenty years or so. In this Section we also discuss whether the mixing properties of an IGARCH process, obtained so far in the literature, conform to those assumed in the relevant limit theorems. In Section 4 we discuss some issues that arise in testing for Aggregational Gaussianity under infinite variance and present some additional empirical evidence supporting the coexistence of infinite variance and Aggregational Gaussianity. The last Section concludes the chapter.

4.2 Empirical Motivation: Distributional Characteristics Of Crop Price Changes

The motivation for this chapter derives from analyzing the dataset of spot crop prices obtained from S&P Goldman Sachs Commodity Indices for cocoa, coffee, corn, soybean, sugar and wheat. In this dataset, the inception date of each crop price index ranges from 12/31/1969 to 1/6/1984. Figure 1 reports the empirical distributions of logarithmic price changes for sugar at daily, weekly, monthly, quarterly semi-annual and annual frequencies (similar results are obtained for all the crops considered here).

We also estimate a GARCH(1,1) model for the daily logarithmic price changes of all the six crops under consideration (see Model (4.2)) of Section 3). The results may be summarized as follows:

(i) The sum of the maximum likelihood estimates of the GARCH(1,1) parameters is 0.994, 0.999, 0.997, 0.993, 0.995 and 0.994 for cocoa, coffee, corn, soybean, sugar and wheat daily price changes, respectively. These results suggest the presence of a near-to-unit root in the conditional variance of the daily series. Note that this sum decreases with the frequency of observation. For example, the sum of the GARCH parameters is 0.219, 0.4516, 0.752, 0.658, 0.885 and 0.776 for cocoa, coffee, corn, soybean, sugar and wheat semi-annual price changes, respectively. These results suggest that, on average, the

GARCH effects in semi--annual frequency are much weaker than the corresponding ones for daily frequency.

(ii) Visual inspection of the empirical distributions of the crop price changes under consideration suggests that these distributions are leptokurtic for daily, weekly and monthly frequencies. Overall, the degree of leptokurtosis seems to decrease as we move from daily to annual frequency at a slow rate. More specifically, the leptokurtosis does not seem to decrease substantially before we reach at least the quarterly frequency.

(iii) Overall, the combined evidence from (i) and (ii) above, suggests the simultaneous presence of a unit root in the conditional variance together with Aggregational Gaussianity for all the six series under consideration.

In the appendix of this chapter we provide further empirical evidence that aggregational gaussianity appears in additional financial time series of interest. We present the time plots and the corresponding histograms of daily, weekly, monthly and annual logarithmic returns for 3 stock indices, namely SP500 (US), DAX (DE) and NIKKEI (JP), and 3 US dollar spot rates with canadian dollar, british pound and japanese yen. Data were obtained from Bloomberg and cover the period from January of 1971 to December of 2009. Figures in the appendix demonstrates clearly that when we move fro high to lower frequencies of data, the aggregational procedure leads to the Normal distribution. Apart from the empirical evidence, in the appendix we included the same analysis for simulated daily returns series that follow a GARCH(1,1) process with ARCH (α) and GARCH (β) components adding to 0.95, 0.96, 0.97, 0.98, 0.99 and 1 (I-GARCH). This series of figures demonstrates clearly that even in the case of infinite variance (IGARCH), the distribution of lower frequency data (generated by the simulated daily returns) also converge to the Normal distribution.

4.3 Aggregational Gaussianity Under Barely Infinite Variance

Let R_t be the one-period (say daily) continuously compounded return on a crop, defined as $R_t = p_t - p_{t-1}$, where p_t is the natural logarithm of the price of the particular crop. In a similar fashion we define the k-period (say weekly or monthly) return $R_\tau(k)$ as:

$$R_\tau(k) = p_t - p_{t-k} = \sum_{i=1}^{k} R_{t-k+i}. \tag{4.1}$$

The new index, τ, is introduced for notational simplicity, representing the k-period interval, in terms of t. More specifically, since we consider non-overlapping returns, the series of k-period returns, produced by taking non-overlapping sums of the original one-period return series, will be of the form $\{\ldots, p_{t-k} - p_{t-2k}, p_t - p_{t-k}, \; p_{t+k} - p_t, \ldots\}$. This means that one unit in terms of τ will correspond to k units in terms of t.

Next, let us assume that the one-period returns, follow an Integrated GARCH(1,1) (IGARCH(1,1)) process:

$$\begin{gathered} R_t = h_t \nu_t \\ \nu_t \sim NIID(0, \sigma_\nu^2) \text{ and } h_t^2 = c + b h_{t-1}^2 + \gamma \nu_{t-1}^2, \text{ with} \\ c > 0,\ 0 \le b < 1,\ 0 \le \gamma < 1 \text{ and } b + \gamma = 1. \end{gathered} \tag{4.2}$$

We shall attempt to answer the following question: Given that R_t follows an IGARCH process with infinite variance, how does the distribution of $R_\tau(k)$ behave as the returns horizon k increases? To answer this question, we must examine whether the probabilistic properties of R_t are such that enable the application of a relevant limit theorem. To this end, let us first briefly discuss the case of a stable GARCH process, that is when $b+\gamma < 1$. It is well known that under the restriction $b + \gamma < 1$, R_t is a second-order stationary process whose unconditional variance is equal to $\sigma_{R_t} = c/(1-(b+\gamma))$. This process is also $\beta-$mixing with exponential decay (see Carrasco and Chen 2002 and Francq and Zakoian 2006). Since a $\beta-$mixing process is also $\alpha-$mixing, we can appeal to the central limit theorem of Ibragimov (1962) and conclude that as $k \to \infty$, the sequence $R_\tau(k)$ converges in law to the normal distribution. Alternatively we may say that the distribution function of R_t belongs to the domain of attraction of the normal law. Moreover, in this case, the standardizing sequence is given by $\sigma_{R_t}\sqrt{k}$, which enables us to say that the distribution function of R_t belongs to the domain of *normal* attraction (DNA) of the normal law (see Ibragimov and Linnik 1971). A similar result in a different context was obtained by Diebold (1988) who showed that the GARCH effects tend to disappear under temporal aggregation.

Let us now focus attention on the case under study, that is when $b + \gamma = 1$ in which case, the variance of R_t is infinite. In this case we cannot apply the central limit theorem mentioned above. Moreover, the results of Diebold (1988) are derived under the assumption $b + \gamma < 1$ which means that they do not cover the IGARCH case. Therefore, we cannot say anything about the temporal aggregation properties of IGARCH processes. The presence of infinite variance seems to suggest that we must move away from the central limit theorem into limit theorems developed for the case of random variables with infinite variances. Historically, the problem described above was first dealt with by Lévy (1935) in the context of independent and identically distributed (*iid*) random variables and later by Ibragimov and Linnik (1971) for the case of mixing random variables (see Kourogenis and Pittis 2010 for an extensive discussion). Given the infinite variance of R_t, it seems reasonable to assume that R_t belongs to the domain of non-normal attraction of a stable law with exponent a. If this were the case, it would have implied two things: (a) the limiting distribution of $R_\tau(k)$ is a stable distribution (but *not* the normal distribution);(b) the sequence by which the partial sum process, $R_\tau(k)$, is standardised cannot be $\sigma_{R_t}\sqrt{k}$.

However, the case of IGARCH is different: An IGARCH process exhibits barely infinite variance meaning that all the moments $E\left|u_1\right|^\delta$ for every δ, $0 \leq \delta < 2$ are finite (see Corollary 1 in Kourogenis and Pittis 2008). In such a case, despite having infinite variance, the R_t's belong to the domain of non-normal attraction of the normal law. In other words, there exists a sequence $\{\delta_k\}$, which necessarily has the form $\delta_k = L(k)\sqrt{k}$, such that:

$$\frac{R_\tau(k)}{\delta_k}$$

weakly converges to the normal distribution. The function $L(k)$ is of particular interest: it is usually referred to as "slowly varying (at infinity)" meaning that $\frac{L(tx)}{L(x)} \to 1$ as $x \to \infty$ for every $t > 0$. The limit theorems that ensure this result, are produced by Bradley (1988) or Peligrad (1990) for $\rho-$mixing and $\phi-$mixing sequences, respectively (see Kourogenis and Pittis 2008, 2010). These results show that the finite variance assumption is not necessary for the central limit theorem. More specifically, for strictly stationary sequences, (as is the IGARCH case considered here) the central limit theorem amounts to the truncated moment function, defined by:

$$H(x) = ER_1^2\mathbf{I}_{|R_1|\leq x} \ ,$$

being slowly varying as $x \to \infty$, that is:

$$H(x) \text{ is slowly varying as } x \to \infty \tag{4.3}$$

In fact, the condition of slow variation of $H(x)$ is both necessary and sufficient for R_t to lie in the domain of attraction of the normal distribution (see Ibragimov and Linnik 1971). The requirement that $H(x)$ is a slowly varying function is equivalent to the condition:

$$E\left|R_1\right|^{\delta} < \infty, \ 0 \leq \delta < 2. \tag{4.4}$$

The latter condition amounts to saying that the R_t's have *just barely* infinite variance (see Bradley 1988). This implies that the central limit theorem may hold even in cases that the variance of the R_t's is infinite, provided that all the moments of order $\delta < 2$ are finite.

The preceding discussion suggests that the empirical features of Aggregational Gaussianity and Infinite Variance in crop price changes can coincide due to the limit theorems for mixing sequences with barely infinite variance mentioned above. However, one word of caution is in order. In order to apply the central limit theorem of Bradley (1988) or that of Peligrad (1990) we must ensure that an IGARCH process is either $\rho-$mixing or $\phi-$mixing, respectively. As far as we know, the relevant literature is yet to produce such a result. Having said this, it is worth mentioning Francq and Zakoian's (2006) relevant result, which proves that an IGARCH process is $\beta-$mixing with exponential decay. However since there is no proof to date that $\beta-$mixing implies either $\rho-$mixing or $\phi-$mixing, the use of the above mentioned theorems should be exercised with caution.

4.4 Testing for Aggregational Gaussianity Under IGARCH

The preceding discussion must have made clear that Aggregational Gaussianity is allowed to coincide with the assumption that the returns over the shortest horizon (say daily) follow an IGARCH process with barely infinite variance. However to establish this fact empirically, using formal statistical methods is rather tricky. The usual procedure for evaluating whether a given empirical distribution is normal involves estimating the sample skewness and kurtosis coefficients, α_3 and α_4, respectively. To this end, establishing Aggregational Gaussianity would imply to estimate these coefficients over various frequencies, and observe that α_3 and α_4 tend to 0 and 3, respectively, as the frequency of observation (returns horizon) decreases (increases). However, this strategy does not work in the case under study, because the returns over the shortest horizon (one-period), R_t, are assumed to follow an IGARCH process. In this case, the population skewness and kurtosis coefficients are infinite, which in turn implies that the corresponding sample estimates, $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ will diverge to infinity as the sample size (of daily observations) increases.

Let us examine more closely the behavior of the estimated kurtosis coefficient, $\widehat{\alpha}_4$, of $R_\tau(k)$ as k increases under the assumption that the one-period returns, R_t, is an IGARCH process. To this end, we conduct a small Monte Carlo experiment. Specifically, we generate 1000 near-to- IGARCH(1,1) series of length equal to 10056 which is the number of daily observations in our sample. The conditional variance parameters were set equal to $b = 0.059299$ and $\gamma = 0.935634$, which are the average values of the estimated parameters across the three crops under consideration. For each of these 1000 replications, we generate five more series, $R_\tau(k)$, $k = 5$, 20, 60, 120, and 240 according to (4.1), corresponding to weekly, monthly, quarterly, semi-annual and annual frequencies. Note that the number of observations decreases with k; in particular we end up with 2011, 503, 168, 84 and 42 observations for $k = 5$, 20, 60, 120, and 240, respectively. Then, for each replication, we estimate the kurtosis coefficient for all the available frequencies, namely $k = 0$, 5, 20, 60, 120, and 240 and take the average (referred to as $\widehat{\alpha}_{4,MC}$) across the 1000 replications for each frequency. The results are reported in Figure (4.2), together with the corresponding average estimated kurtosis coefficients (referred to as $\widehat{\alpha}_{4,D}$) across the six crops under consideration.

The results may be summarised as follows:

(i) The Monte Carlo kurtosis coefficient, $\widehat{\alpha}_{4,MC}$ appears to exhibit a pattern similar to that observed for the kurtosis coefficient, $\widehat{\alpha}_{4,D}$, of the real data. In particular, $\widehat{\alpha}_{4,MC}$ increases temporarily as we move from $k = 1$ to $k = 5$ and then decreases with k.

(ii) The behaviour of $\widehat{\alpha}_{4,MC}$ reported above is typical for IGARCH (or near-to-IGARCH) processes. On the contrary for GARCH parameters safely inside the stationarity region the behaviour of $\widehat{\alpha}_{4,MC}$ is exactly that predicted by CLT, namely $\widehat{\alpha}_{4,MC}$ converges monotonically to 3 as the returns horizon increases.

(iii) The behaviour of $\widehat{\alpha}_{4,MC}$ reported above may be due to the following reasons: First, as k increases there are two opposite forces at work: The first one stems from the fact that R_t does belong to the domain of attraction of the normal law, which means that as k increases, the corresponding processes $R_\tau(k)$ become "more normal". This force creates a tendency for the estimates of the kurtosis coefficient to approach the value of 3. However, as k increases, the number of observations on the corresponding $k-$horizon returns, available in a given time period, decreases. For example, for the time period 29/12/1969 to 12/11/2009 we have 10056 daily observations but only 2081 weekly, 480 monthly, 161 quarterly, 81 semi annual and 41 annual observations.

The smaller number of observations makes it harder for CLT to take effect, thus creating a tendency for $\widehat{\alpha}_{4,MC}$ to deviate from 3. The second reason, which may explain the non-monotonicity in the behavior of $\widehat{\alpha}_{4,MC}$ is related to the rate of convergence of $R_\tau(k)$ to normal. In the absence of a finite second moment, the rate of convergence to the normal distribution is expected to be much slower than the corresponding one for the finite-variance case. This property combined with the fact that the number of observations decreases with k may explain the slow and non-monotonic way by which $\widehat{\alpha}_{4,MC}$ approaches the value of 3 as k increases. To this end, it is also interesting to note that the rate of convergence to normality in the presence of a barely infinite variance as suggested by the normalizing sequence, δ_k, is $L(k)\sqrt{k}$ with $L(k)$ being a slowly-varying and possibly non-monotonic function.

4.5 Conclusions

Motivated by empirical evidence indicating that the price changes of six major crops, when observed at high frequencies, seem to be characterised by both leptokurtosis and unit root in the conditional variance, while both of these effects tend to diminish as one moves to lower frequencies, we explain why there is no paradox in admitting the simultaneous existence of Aggregational Gaussianity and infinite variance. In particular, we show that Aggregational Gaussianity and infinite variance can coexist, provided that all the moments of the unconditional distribution whose order is less than two exist. Our theoretical explanation derives from limit theorems for mixing processes with barely infinite variance, developed in the probability theory literature. More specifically, we suggest that the limit theorems of Bradley (1988) or that of Peligrad (1990) for mixing sequences with barely infinite variance, for $\rho-$mixing and $\phi-$mixing sequences respectively, ensure the coincidence of the empirical features of Aggregational Gaussianity and Infinite Variance in crop price changes. Finally, we discuss some issues that arise in testing for Aggregational Gaussianity under infinite variance and present some additional empirical evidence supporting the coexistence of IGARCH effects in high frequency data and Aggregational Gaussianity.

4.6 Appendix

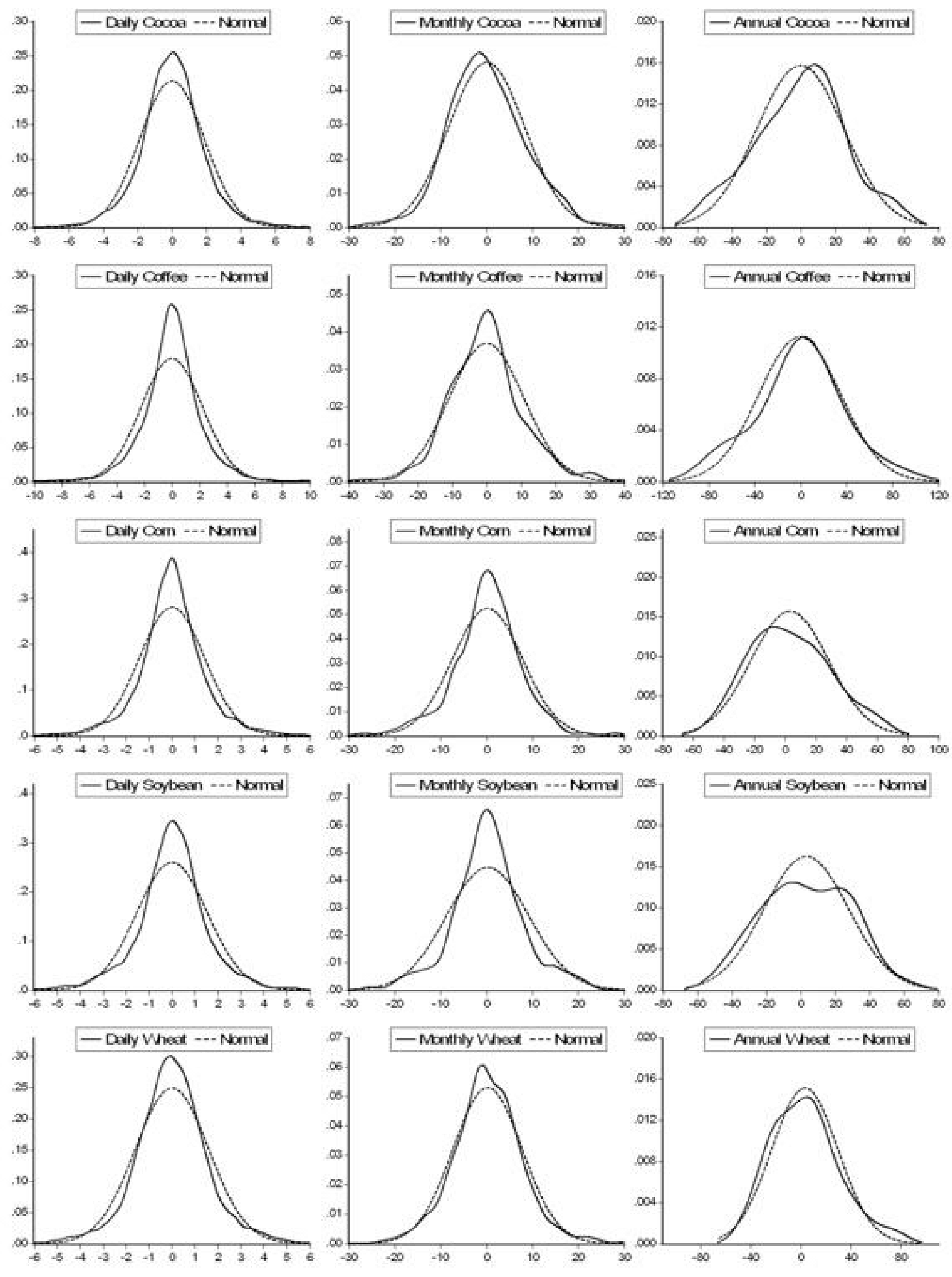


Figure 4.1: Empirical distributions of crop returns

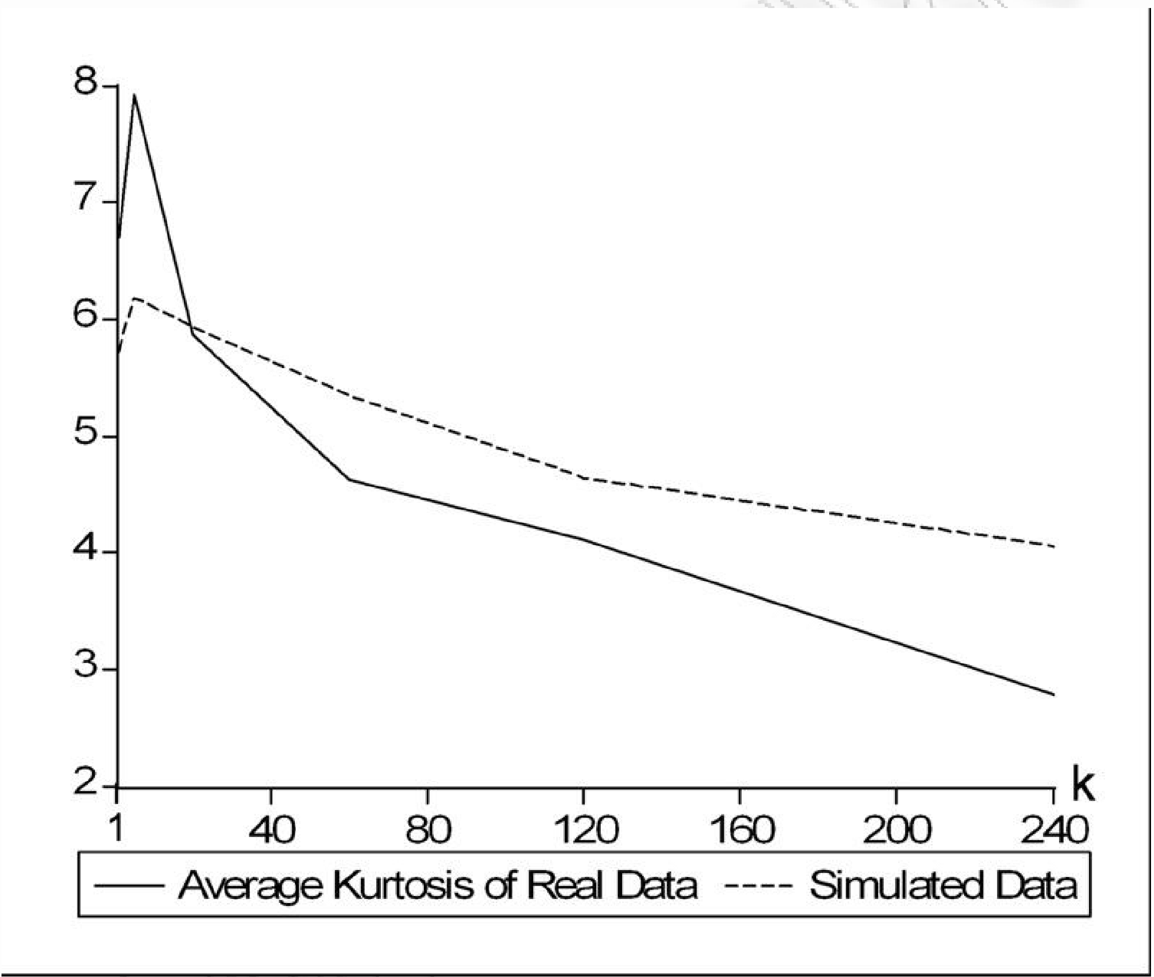


Figure 4.2: Sample kurtosis of simulated and crop returns.

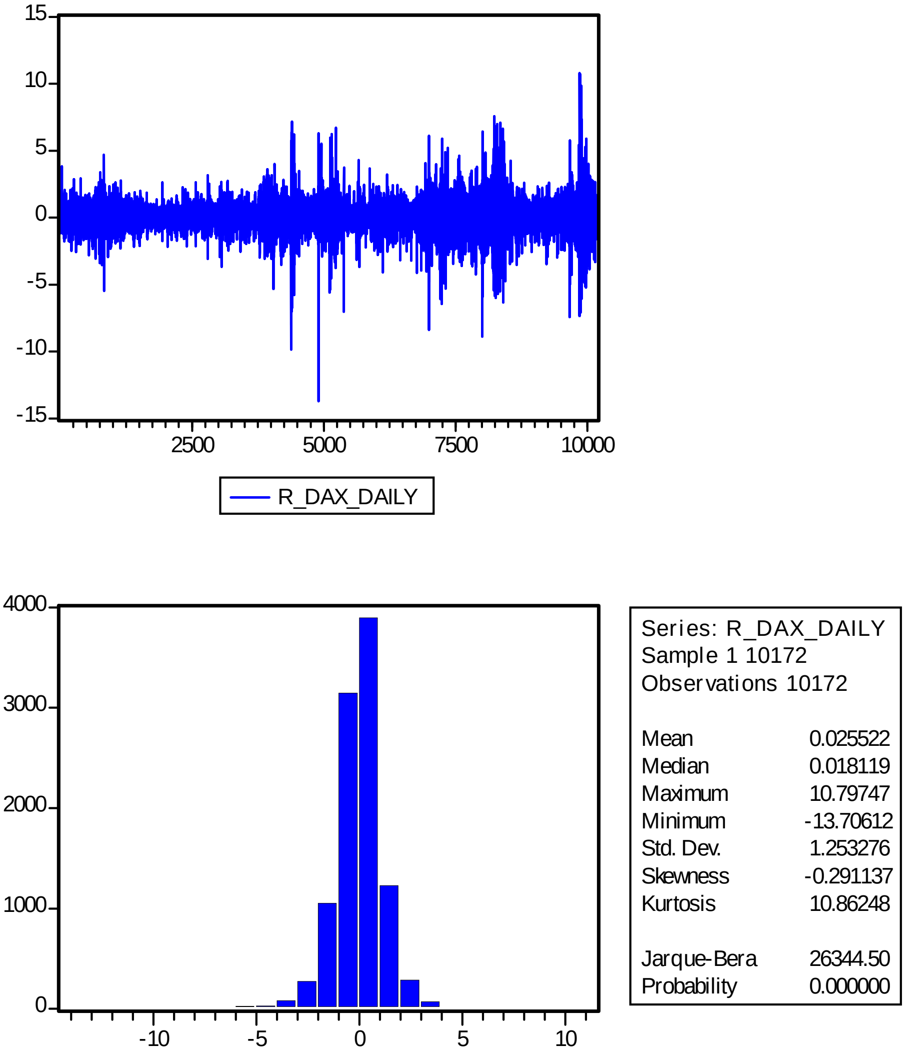


Figure 4.3: Time Plot and Histogram of DAX Index Daily Returns

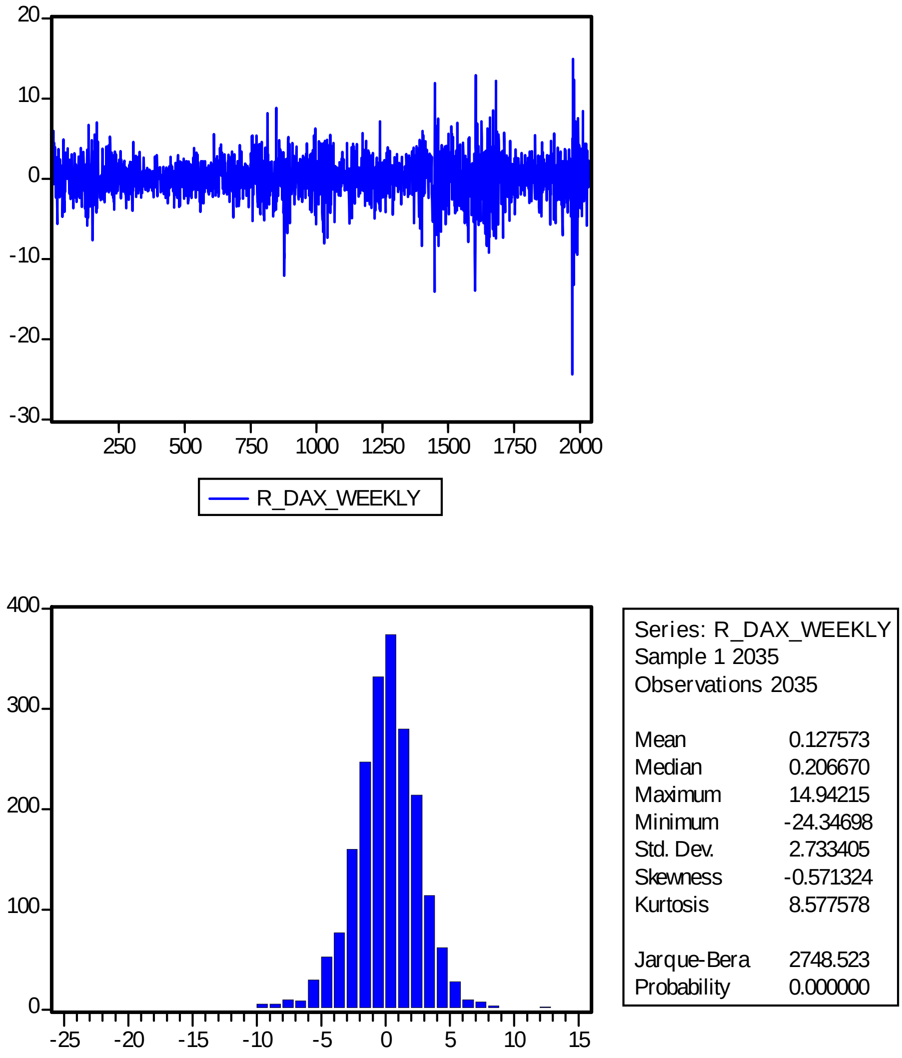


Figure 4.4: Time Plot and Histogram of DAX Index Weekly Returns

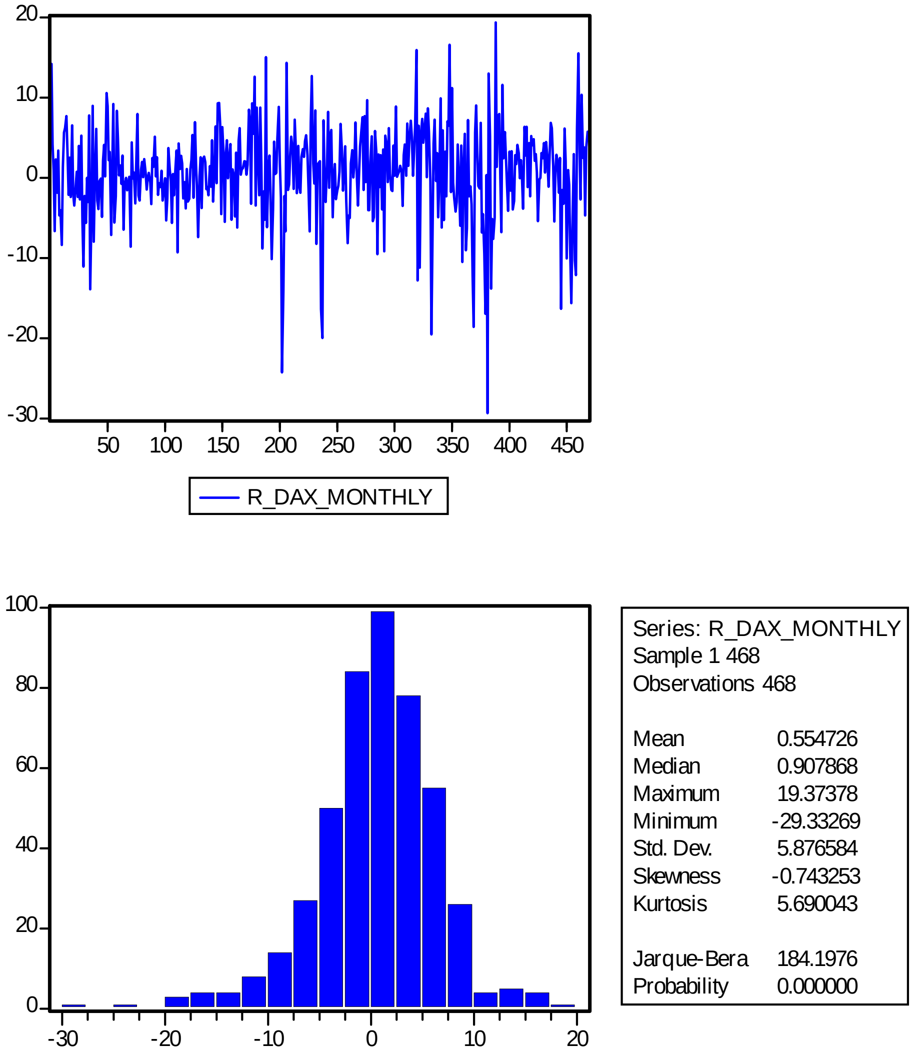


Figure 4.5: Time Plot and Histogram of DAX Index Monthly Returns

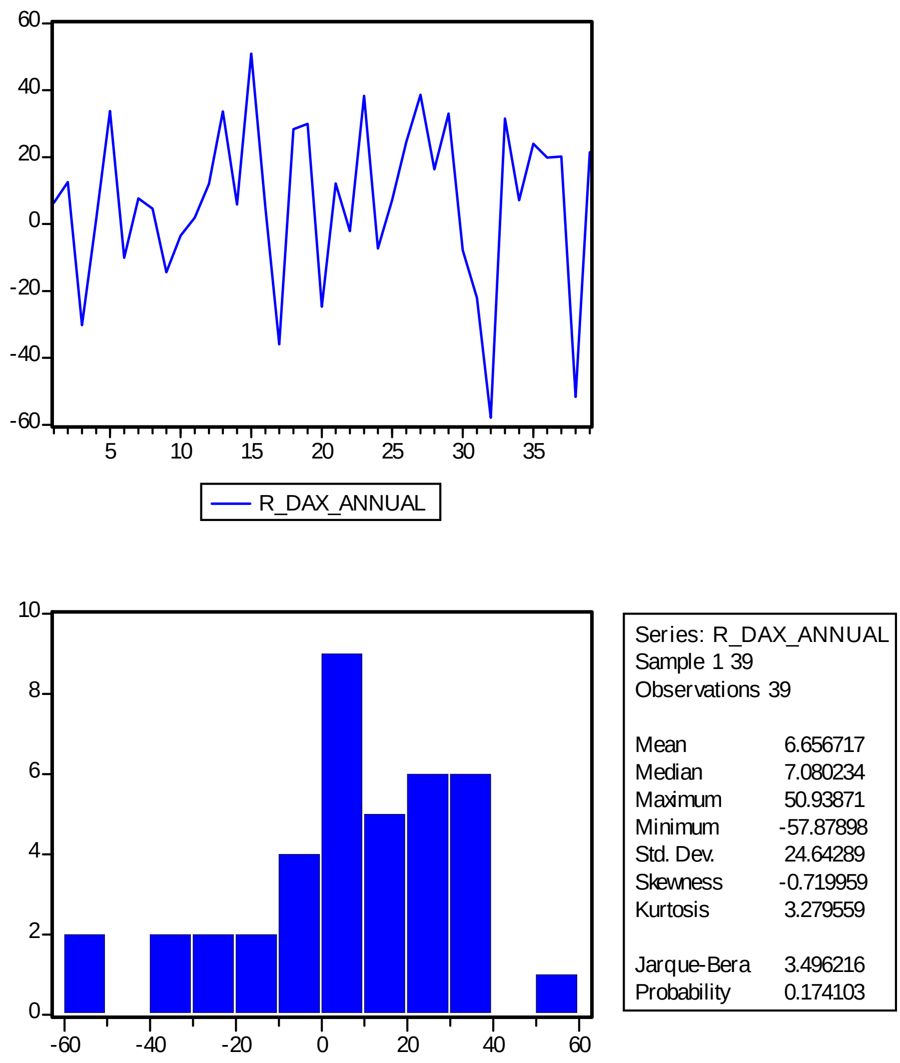


Figure 4.6: Time Plot and Histogram of DAX Index Annual Returns

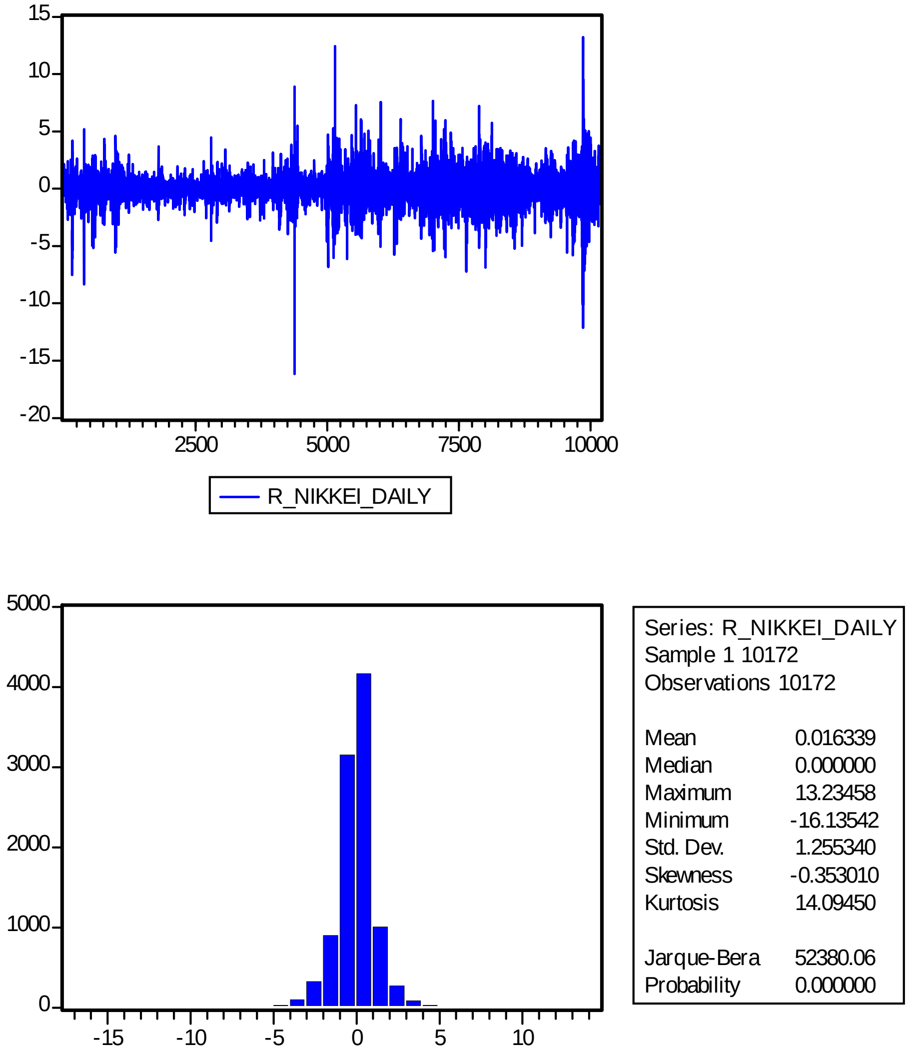


Figure 4.7: Time Plot and Histogram of NIKKEI Index Daily Returns

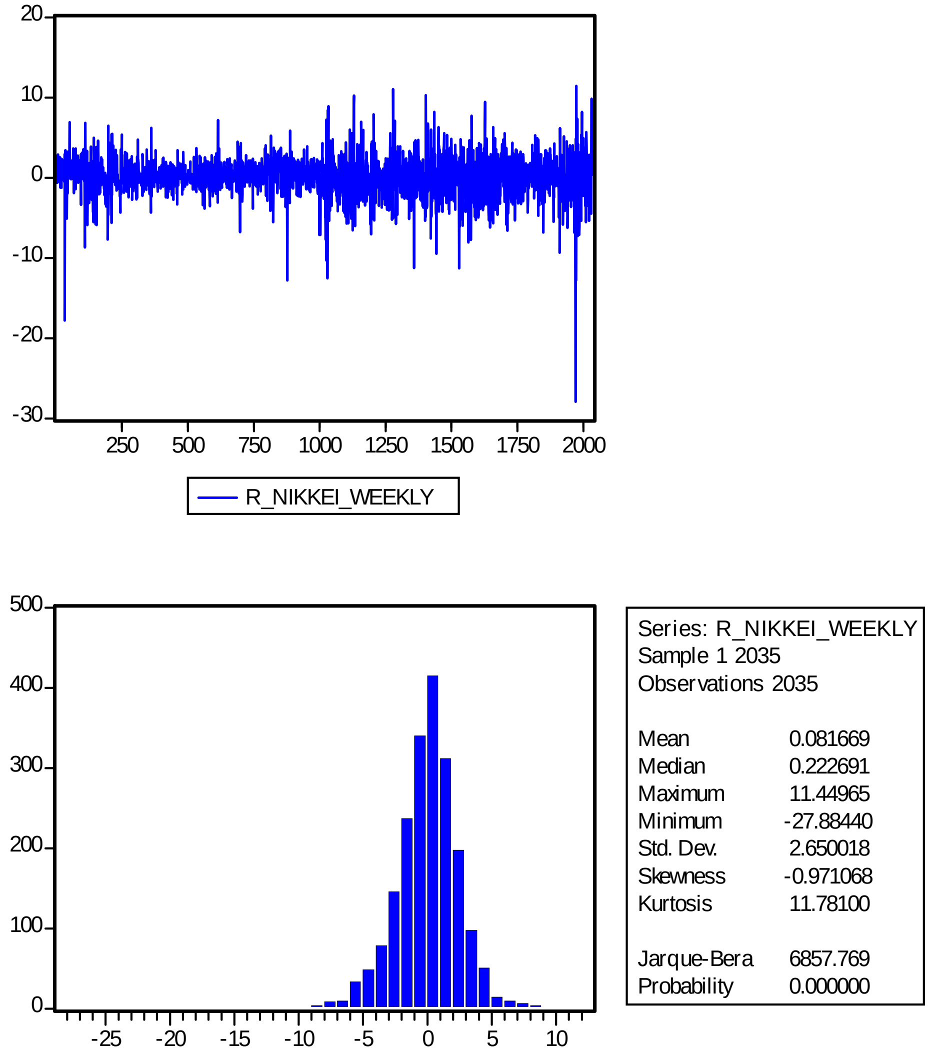


Figure 4.8: Time Plot and Histogram of NIKKEI Index Weekly Returns

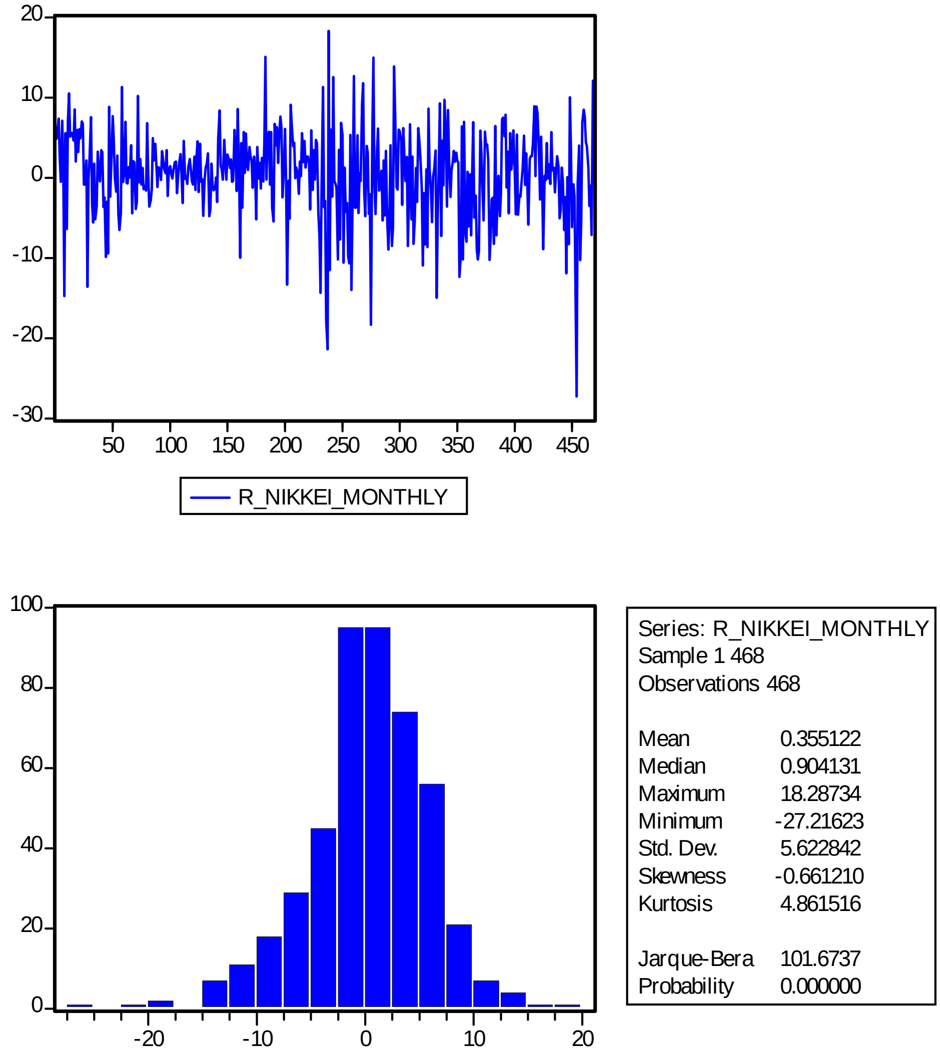


Figure 4.9: Time Plot and Histogram of NIKKEI Index Monthly Returns

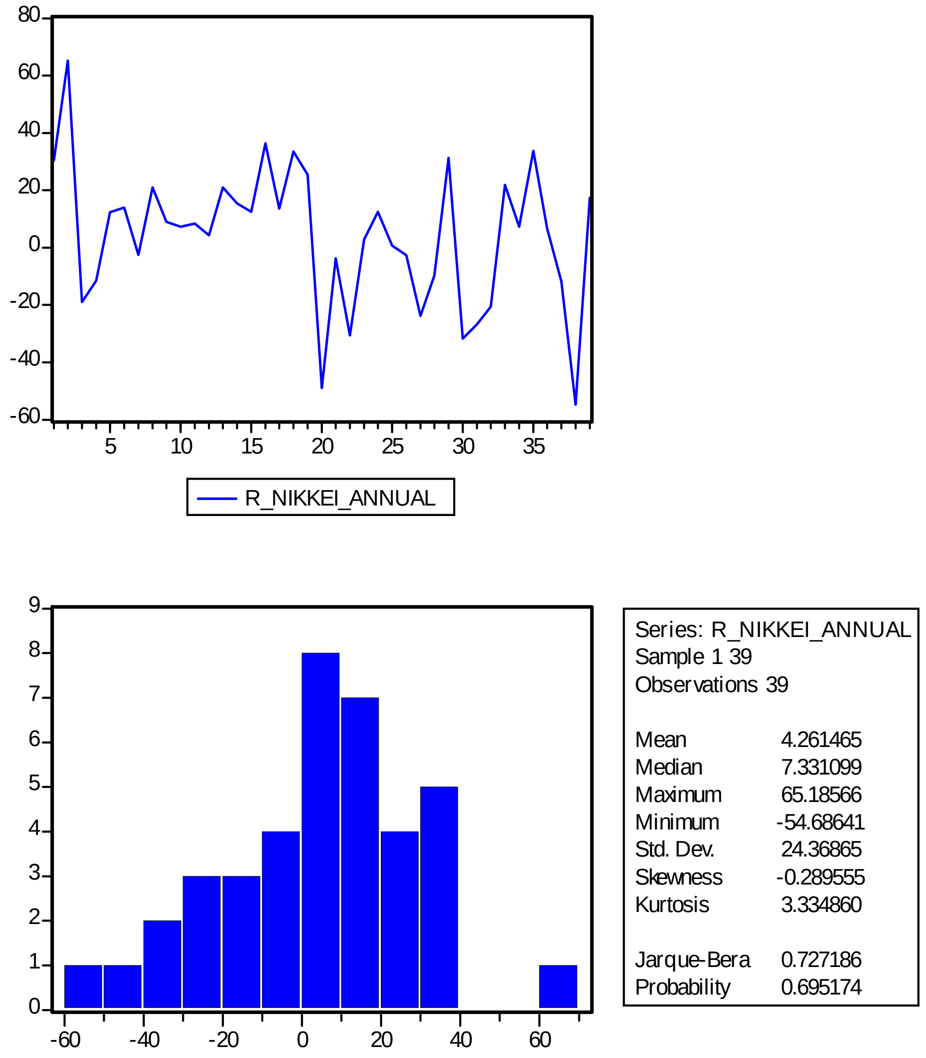


Figure 4.10: Time Plot and Histogram of NIKKEI Index Annual Returns

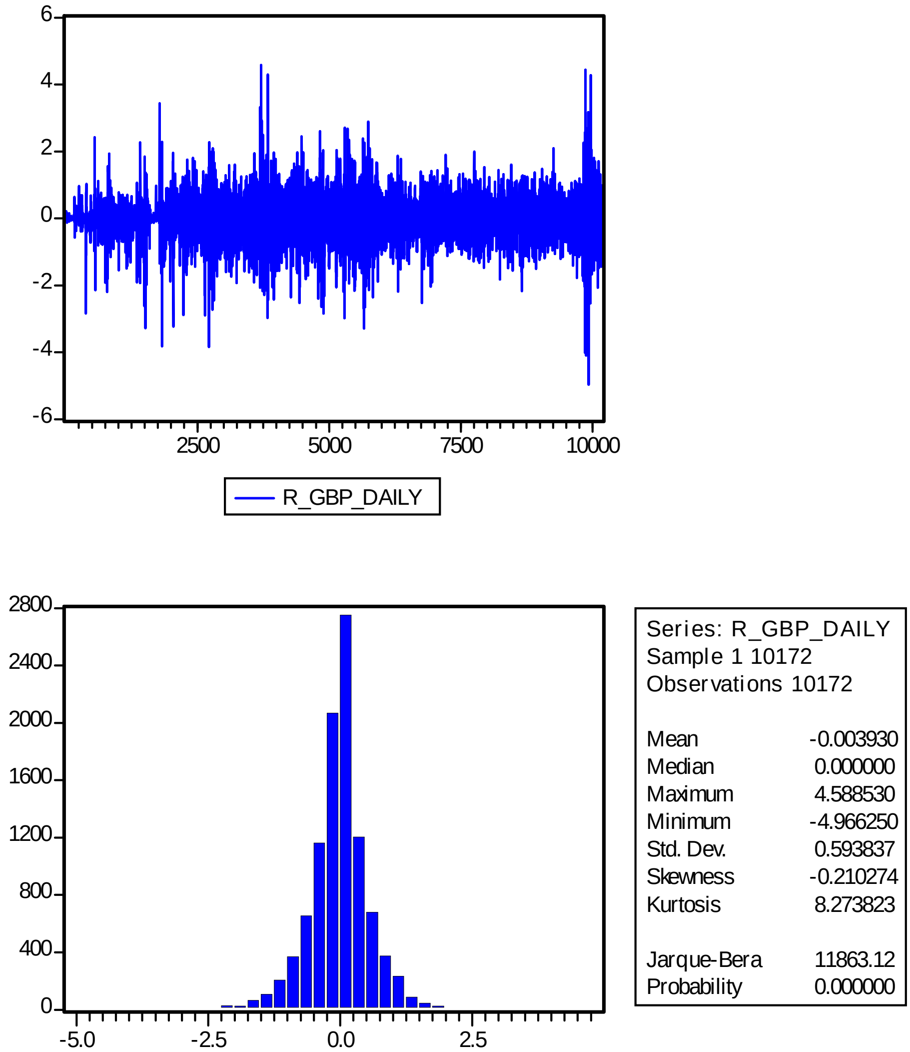


Figure 4.11: Time Plot and Histogram of GBP/USD Daily Returns

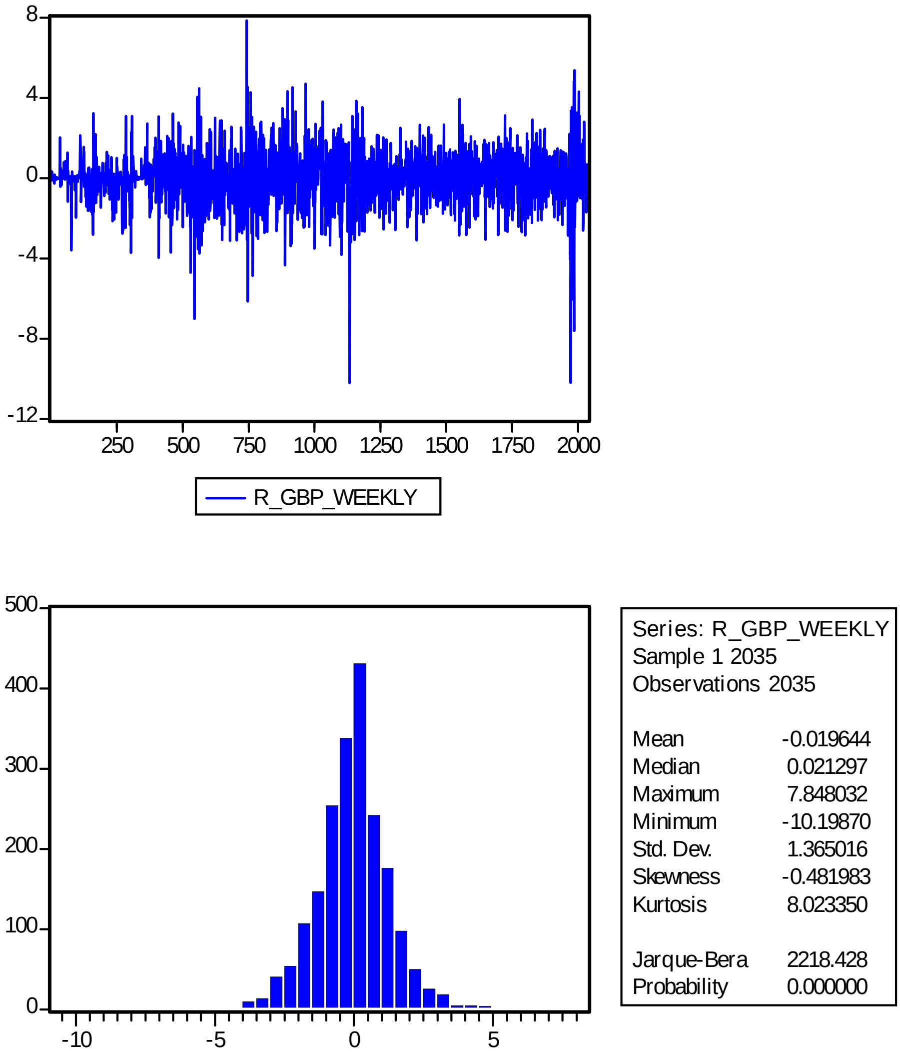


Figure 4.12: Time Plot and Histogram of GBP/USD Weekly Returns

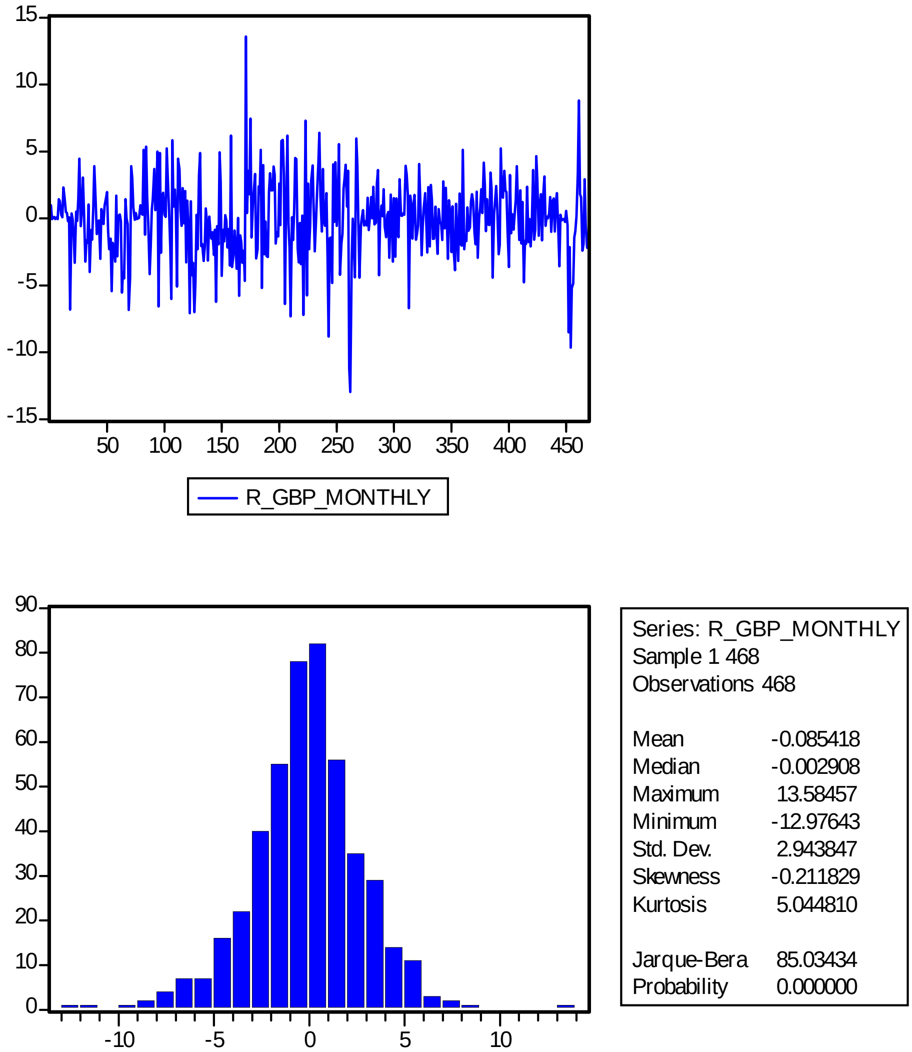


Figure 4.13: Time Plot and Histogram of GBP/USD Monthly Returns

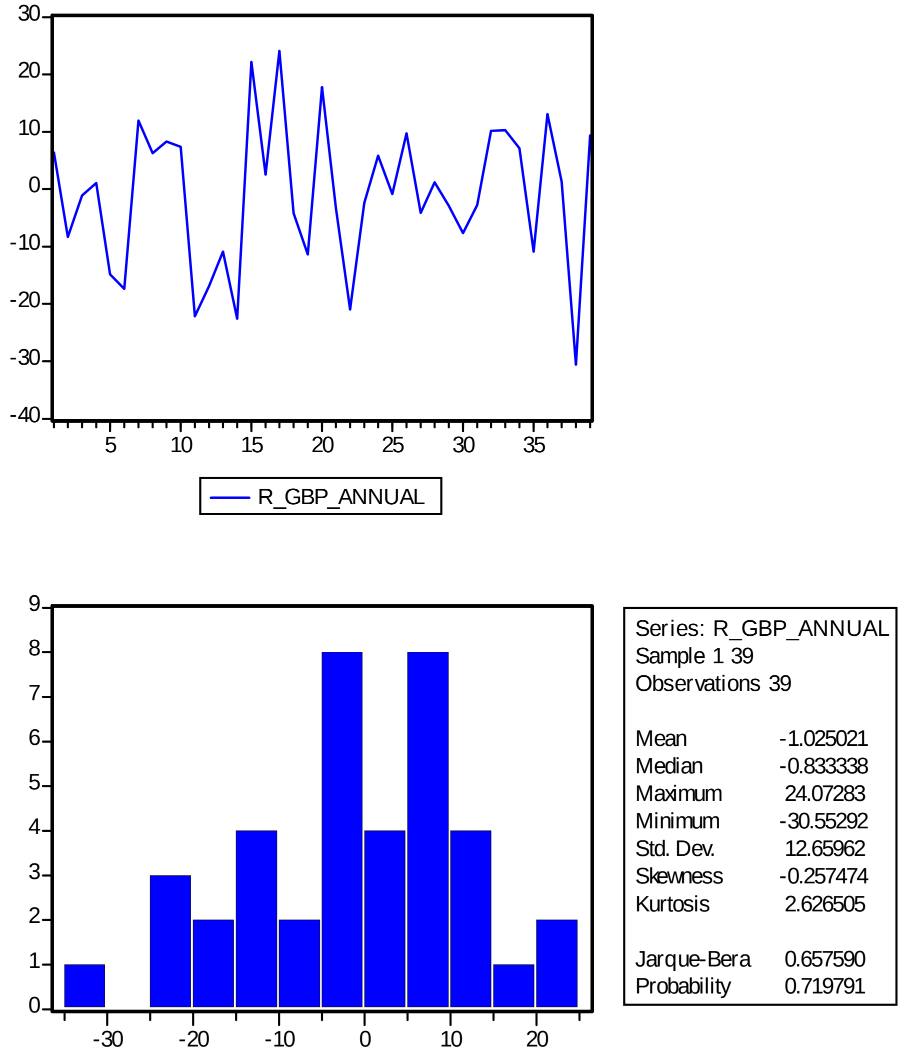
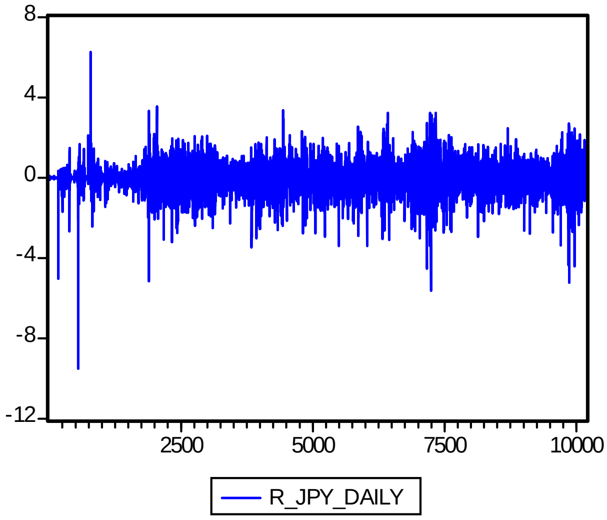
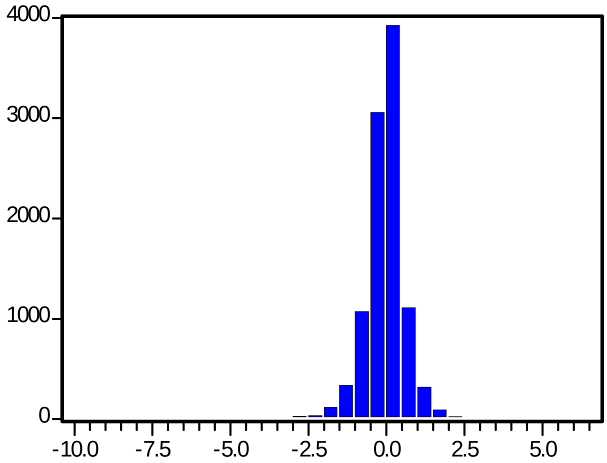


Figure 4.14: Time Plot and Histogram of GBP/USD Annual Returns





Series: R_JPY_DAILY
Sample 1 10172
Observations 10172

Mean	-0.013280
Median	0.000000
Maximum	6.255581
Minimum	-9.504765
Std. Dev.	0.642145
Skewness	-0.809967
Kurtosis	13.77288
Jarque-Bera	50300.17
Probability	0.000000

Figure 4.15: Time Plot and Histogram of JPY/USD Daily Returns

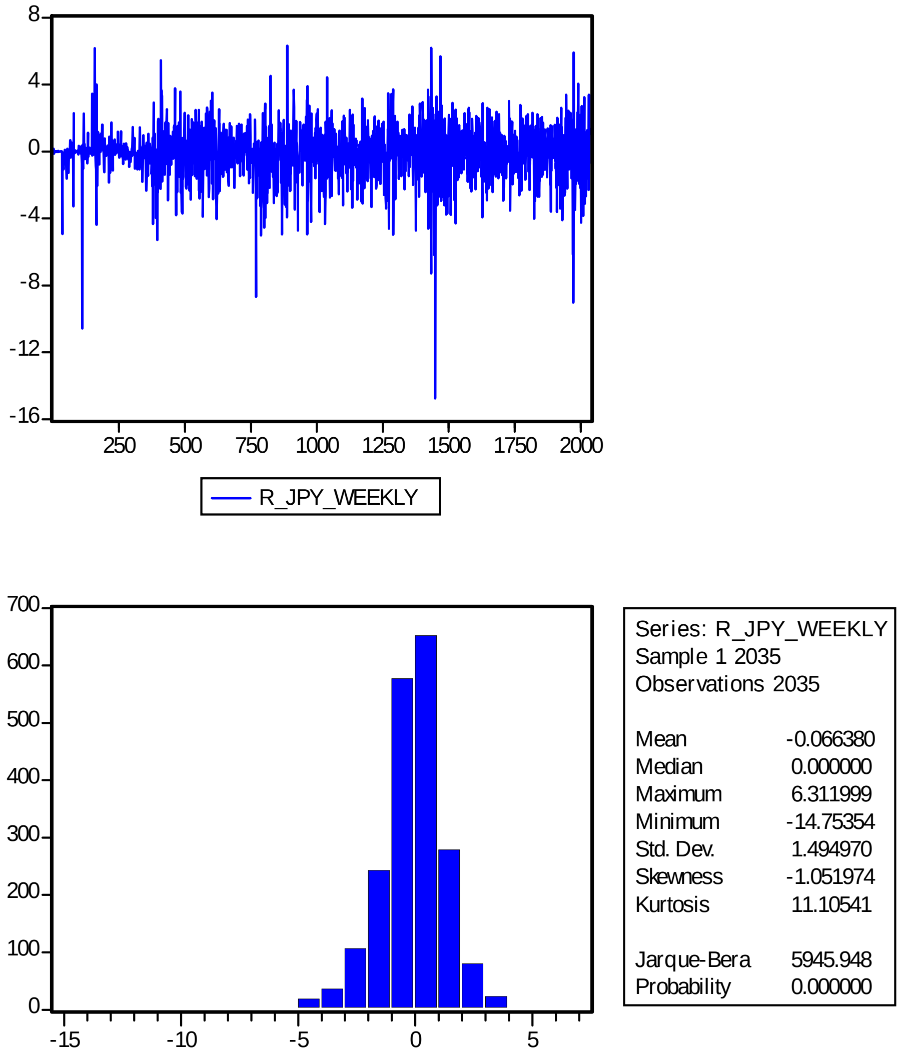


Figure 4.16: Time Plot and Histogram of JPY/USD Weekly Returns

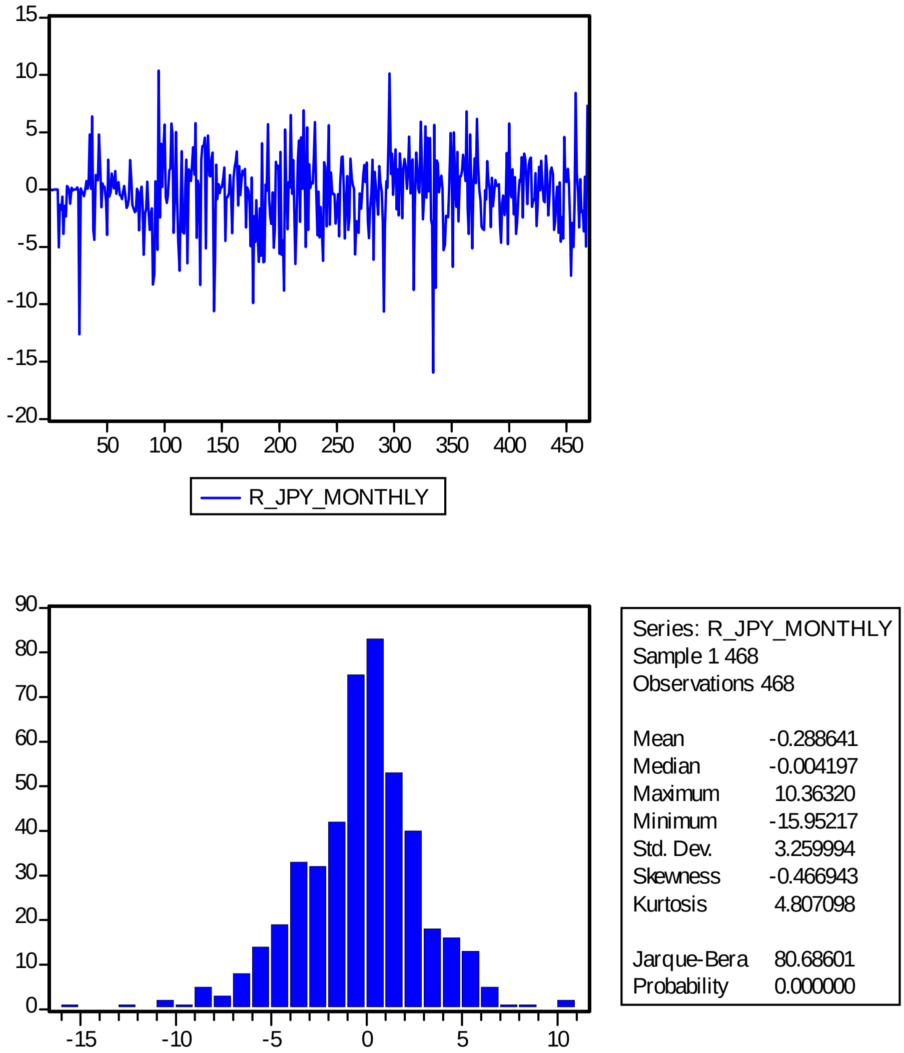


Figure 4.17: Time Plot and Histogram of JPY/USD Monthly Returns

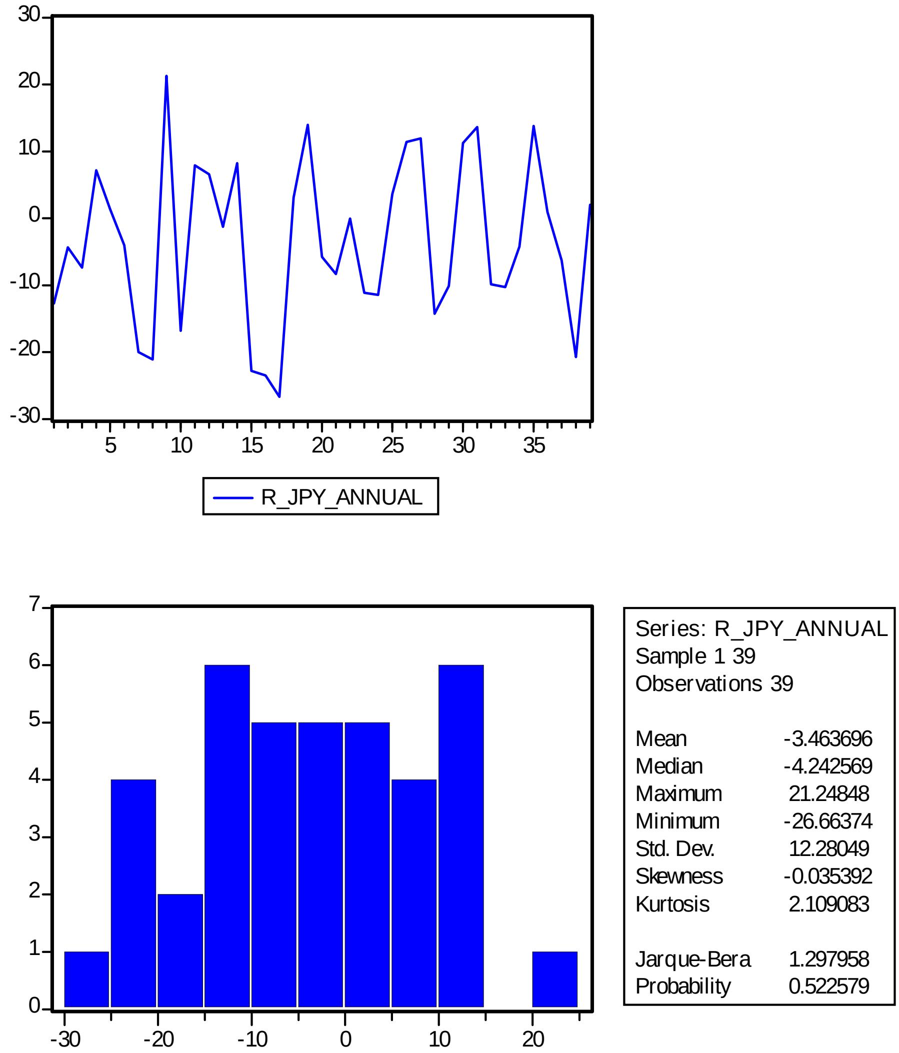


Figure 4.18: Time Plot and Histogram of JPY/USD Annual Returns

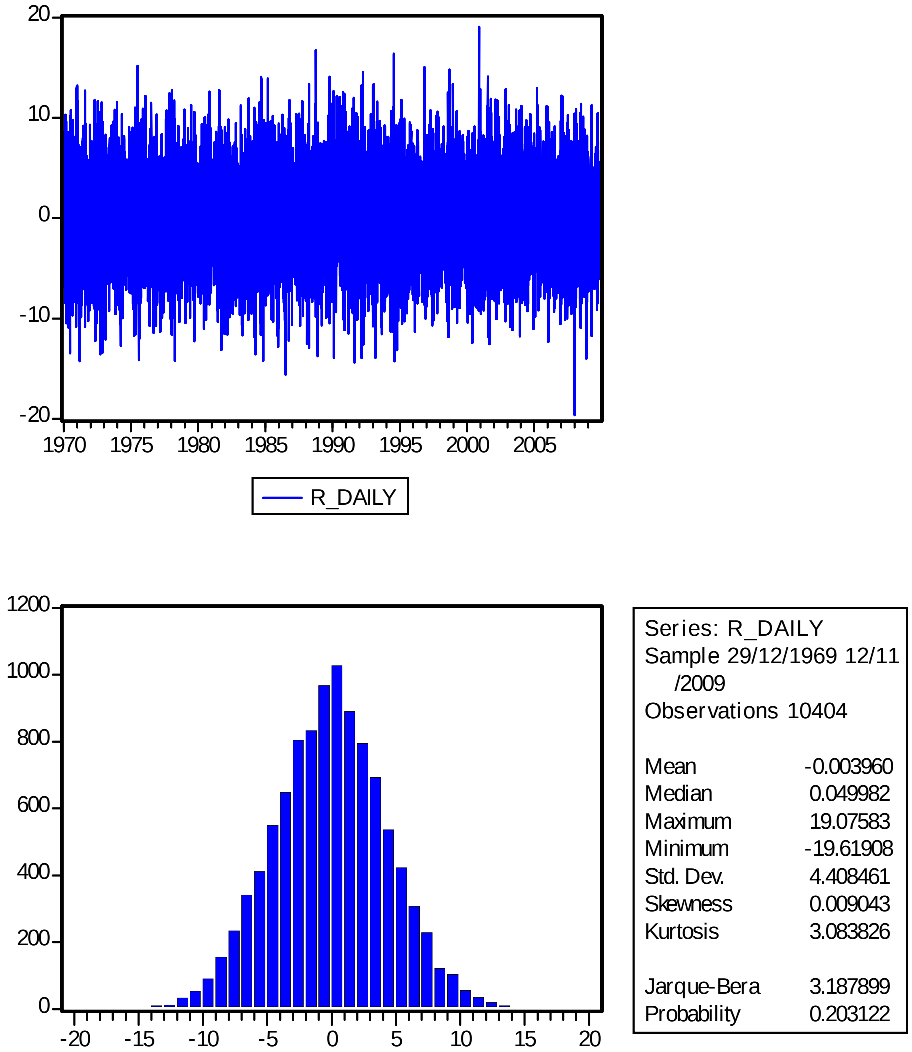


Figure 4.19: Time Plot and Histogram of Simulated GARCH Daily Returns,$\alpha = 0.05, \beta = 0.90, \alpha + \beta = 0.95$

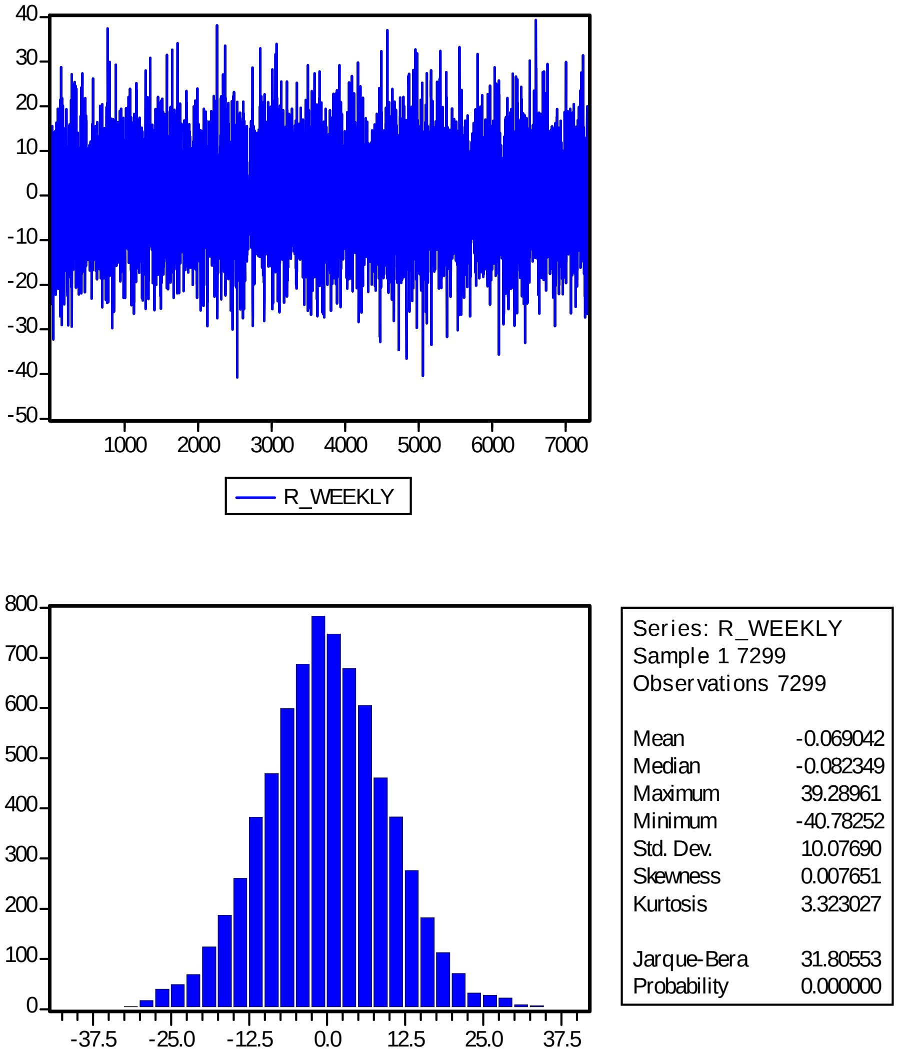
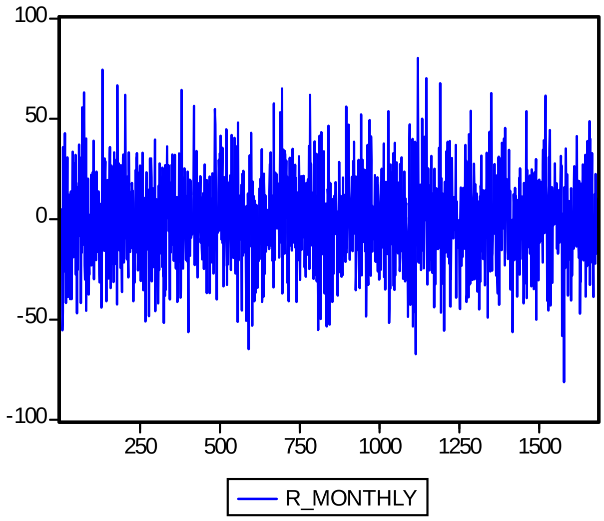


Figure 4.20: Time Plot and Histogram of Simulated GARCH Daily Returns, Weekly Aggregation, $\alpha = 0.05, \beta = 0.90, \alpha + \beta = 0.95$



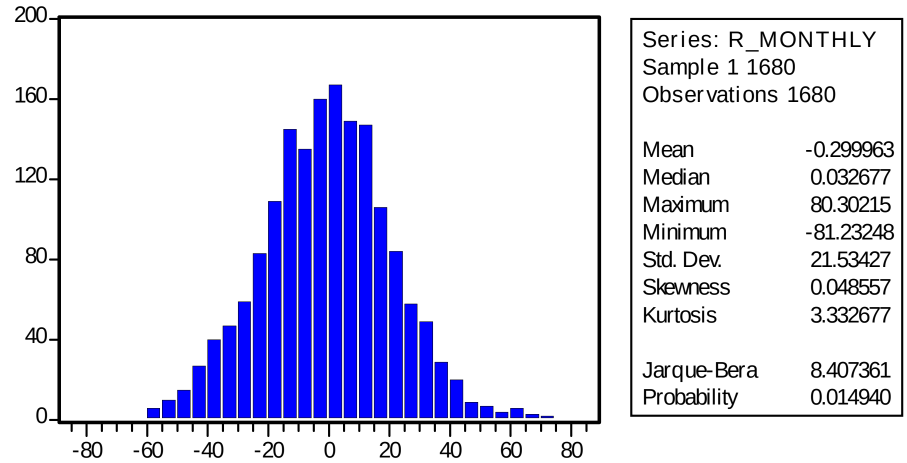


Figure 4.21: Time Plot and Histogram of Simulated GARCH Daily Returns, Monthly Aggregation,$\alpha = 0.05, \beta = 0.90, \alpha + \beta = 0.95$

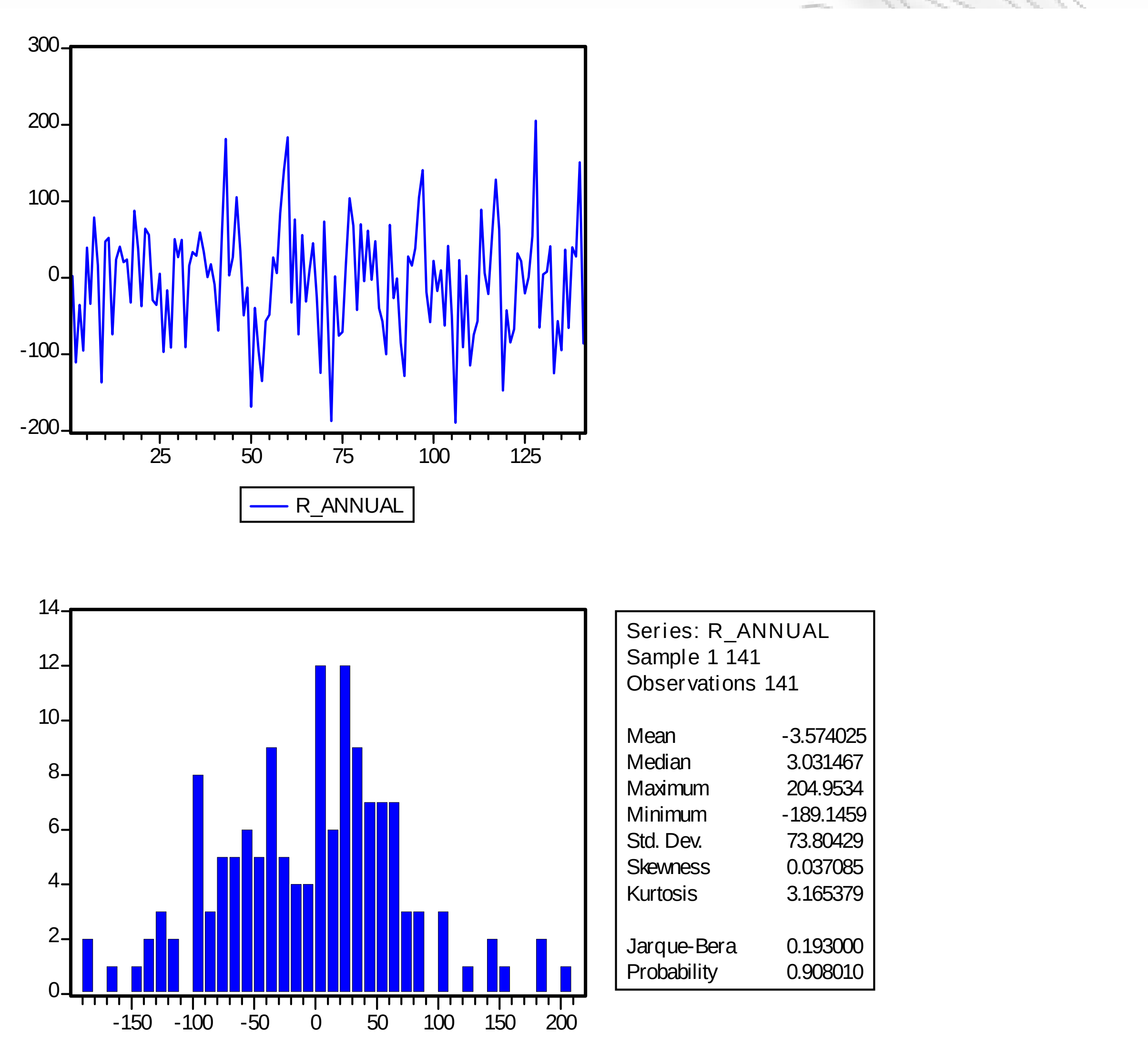


Figure 4.22: Time Plot and Histogram of Simulated GARCH Daily Returns, Annual Aggregation, $\alpha = 0.05, \beta = 0.90, \alpha + \beta = 0.95$

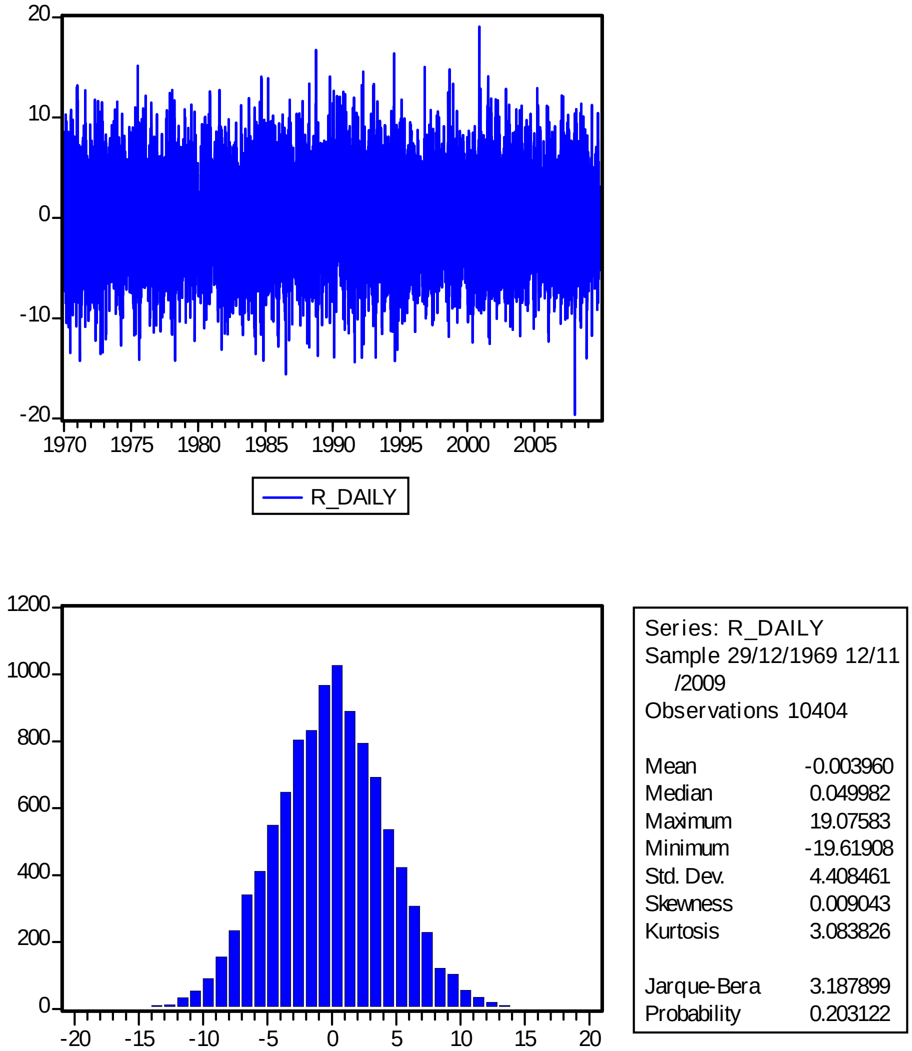


Figure 4.23: Time Plot and Histogram of Simulated GARCH Daily Returns,$\alpha = 0.05, \beta = 0.91, \alpha + \beta = 0.96$

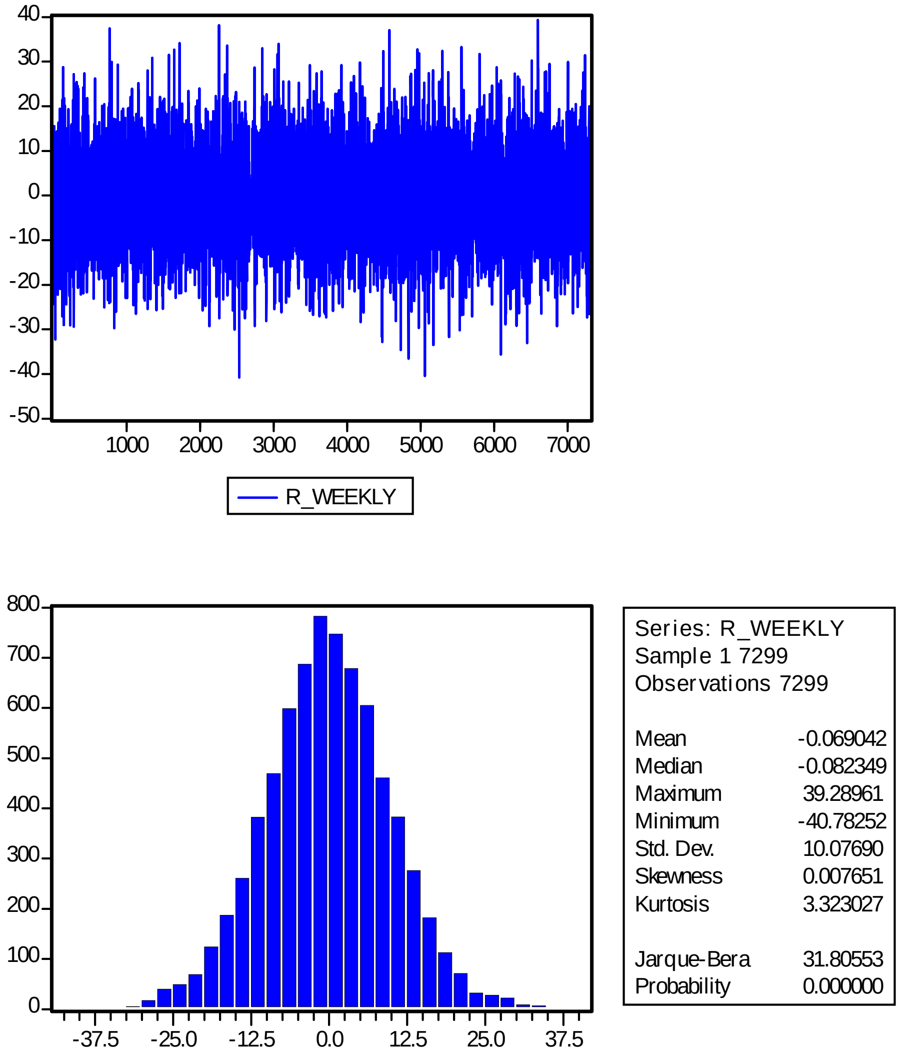


Figure 4.24: Time Plot and Histogram of Simulated GARCH Daily Returns, Weekly Aggregation, $\alpha = 0.05, \beta = 0.91, \alpha + \beta = 0.96$

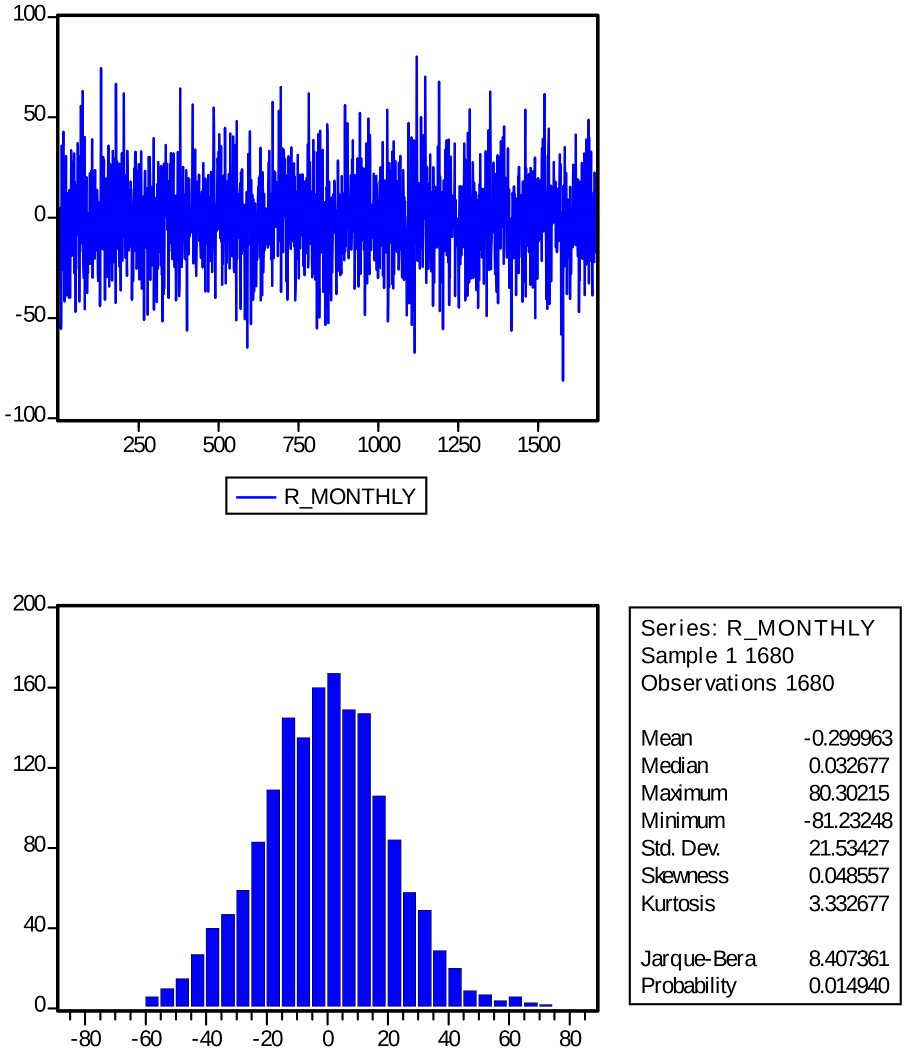


Figure 4.25: Time Plot and Histogram of Simulated GARCH Daily Returns, Monthly Aggregation,$\alpha = 0.05, \beta = 0.91, \alpha + \beta = 0.96$

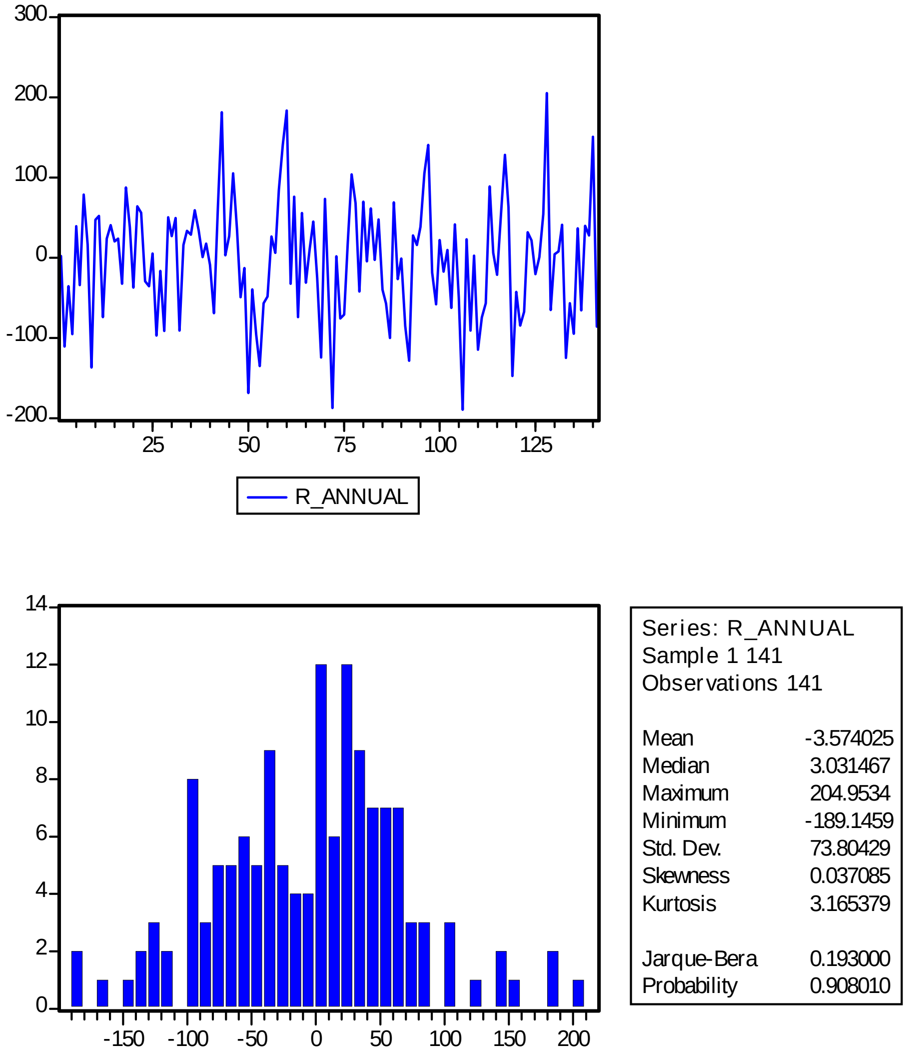


Figure 4.26: Time Plot and Histogram of Simulated GARCH Daily Returns, Annual Aggregation, $\alpha = 0.05, \beta = 0.91, \alpha + \beta = 0.96$

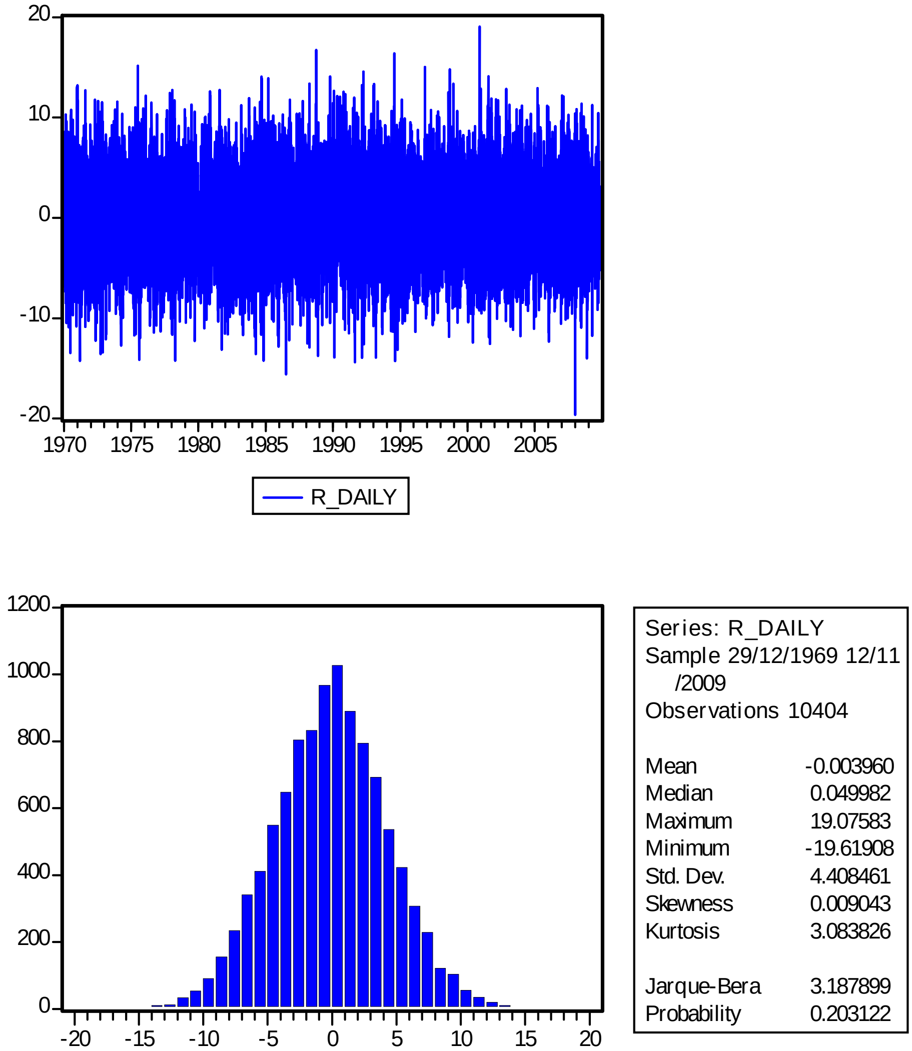
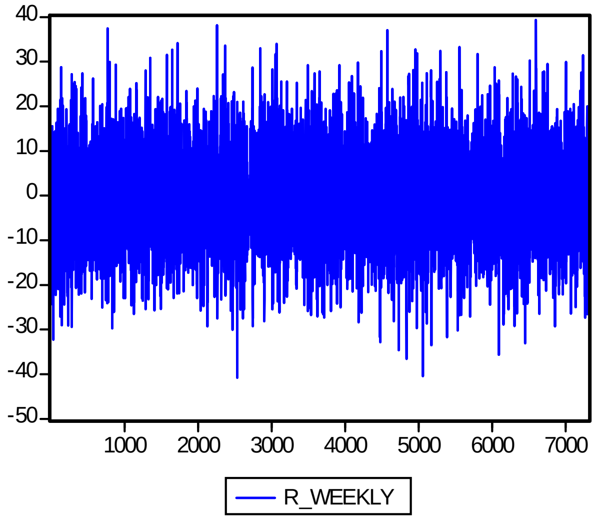


Figure 4.27: Time Plot and Histogram of Simulated GARCH Daily Returns,$\alpha = 0.05, \beta = 0.92, \alpha + \beta = 0.97$



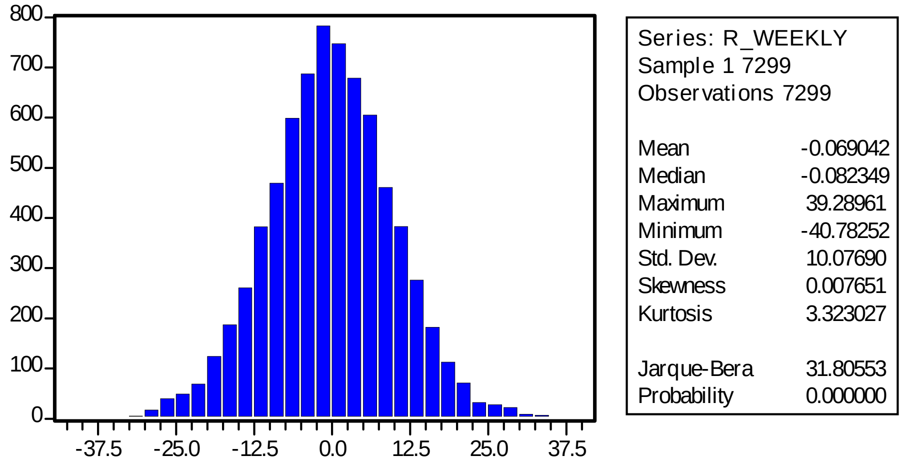
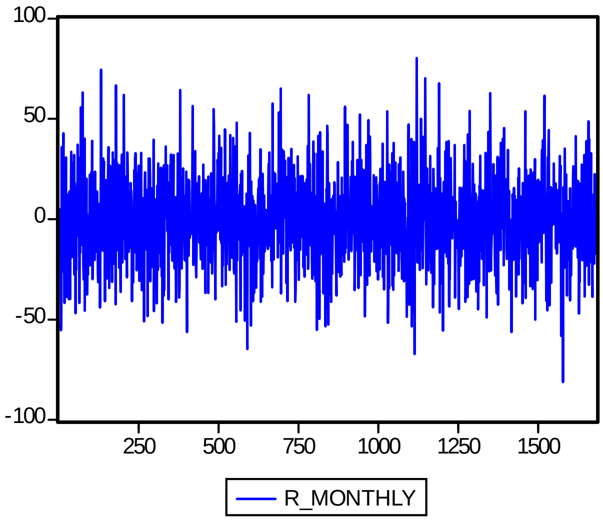


Figure 4.28: Time Plot and Histogram of Simulated GARCH Daily Returns, Weekly Aggregation, $\alpha = 0.05, \beta = 0.92, \alpha + \beta = 0.97$



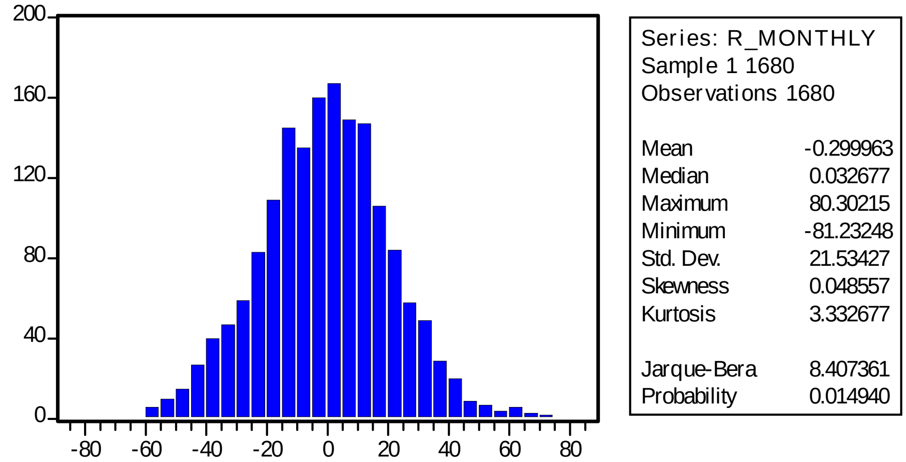


Figure 4.29: Time Plot and Histogram of Simulated GARCH Daily Returns, Monthly Aggregation,$\alpha = 0.05, \beta = 0.92, \alpha + \beta = 0.97$

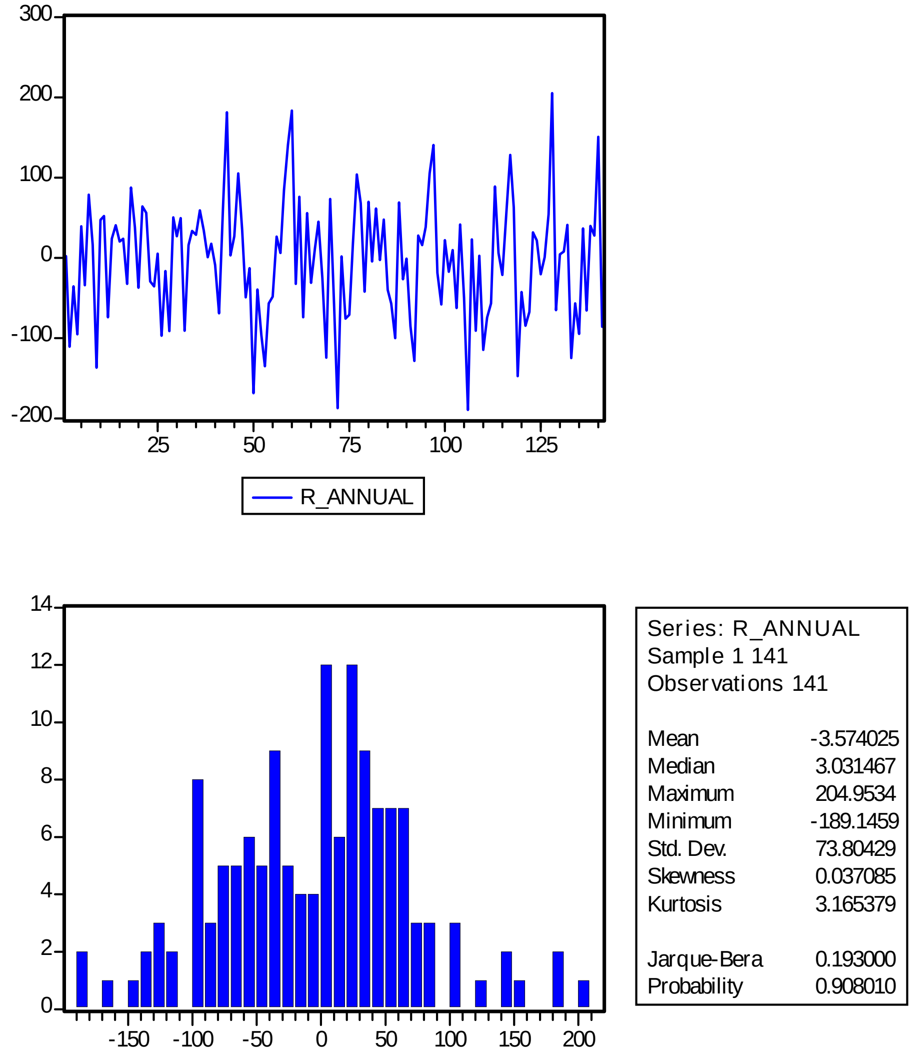


Figure 4.30: Time Plot and Histogram of Simulated GARCH Daily Returns, Annual Aggregation, $\alpha = 0.05, \beta = 0.92, \alpha + \beta = 0.97$

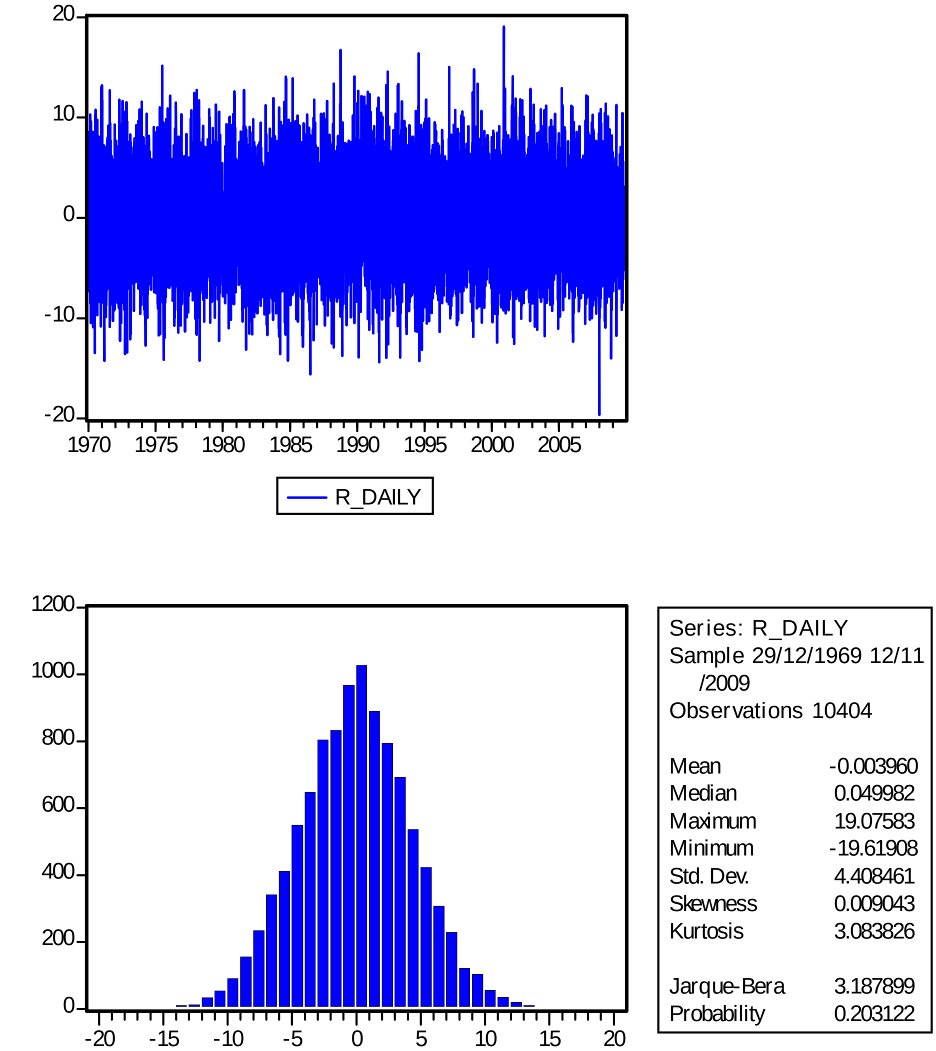


Figure 4.31: Time Plot and Histogram of Simulated GARCH Daily Returns, $\alpha = 0.05, \beta = 0.93, \alpha + \beta = 0.98$

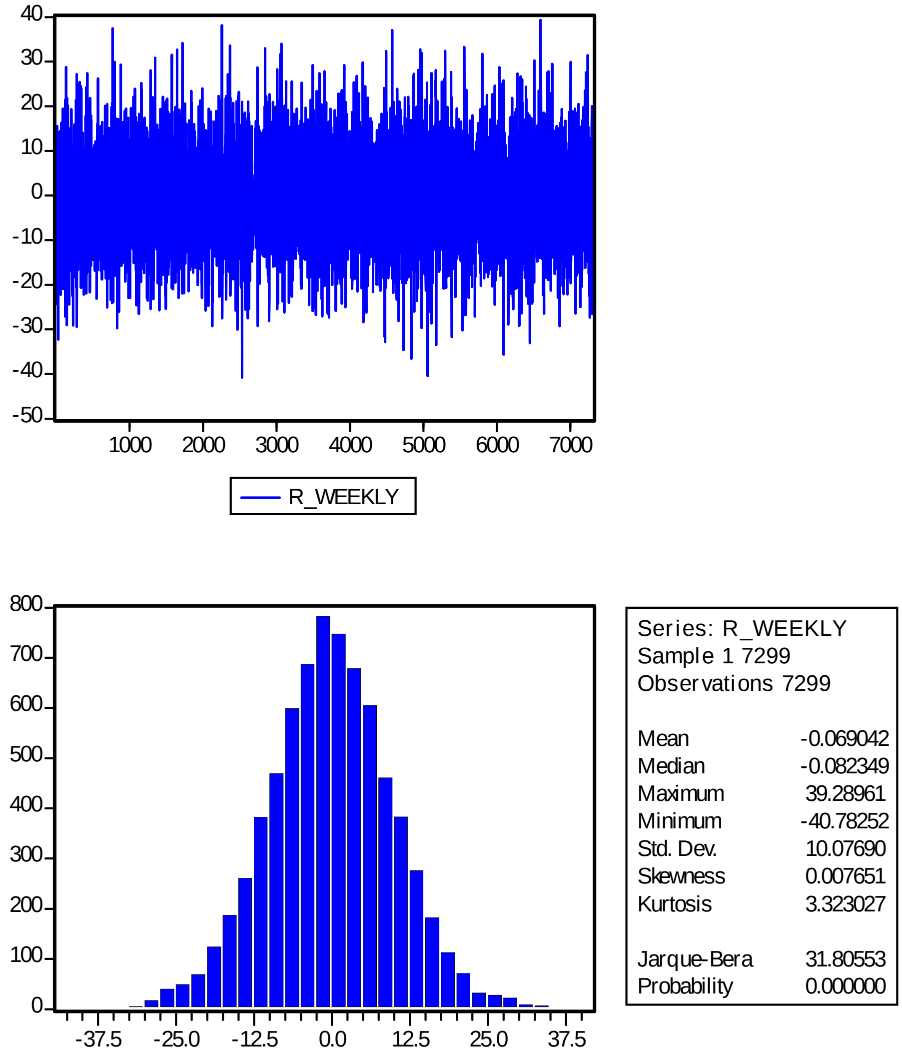


Figure 4.32: Time Plot and Histogram of Simulated GARCH Daily Returns, Weekly Aggregation, $\alpha = 0.05, \beta = 0.93, \alpha + \beta = 0.98$

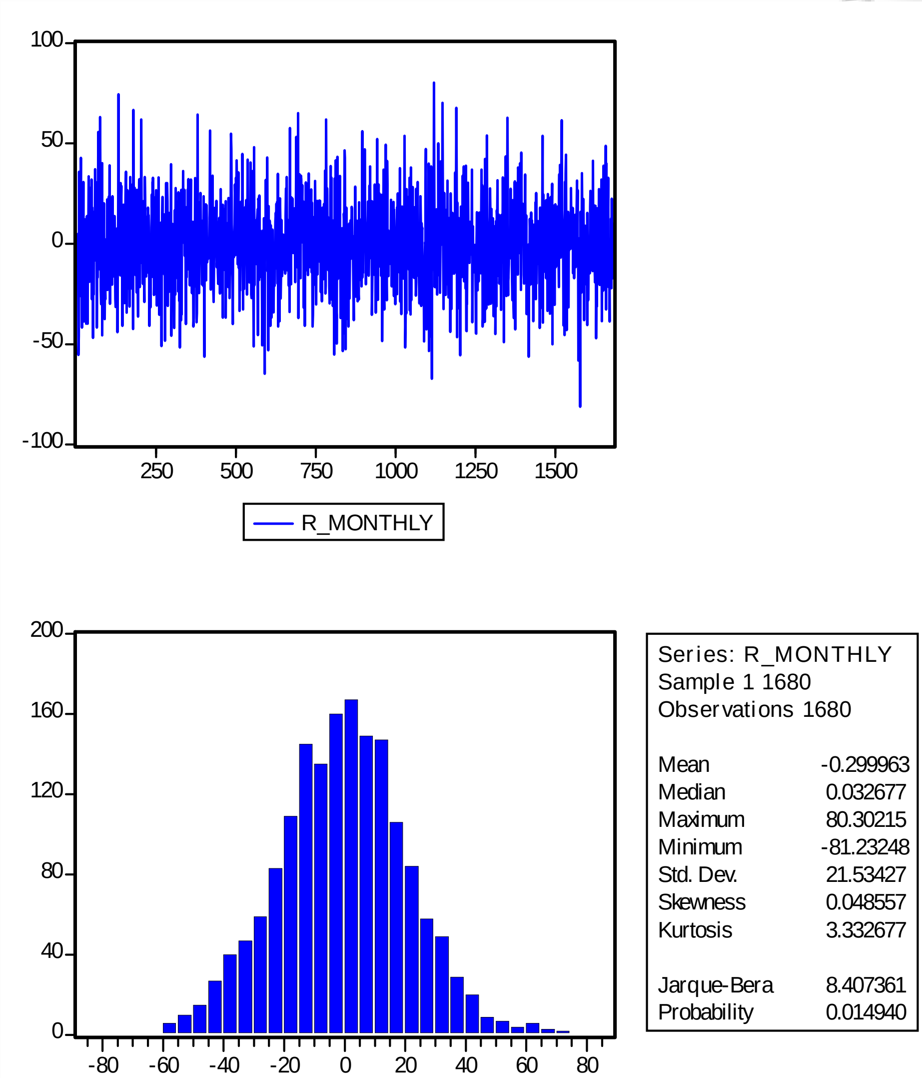


Figure 4.33: Time Plot and Histogram of Simulated GARCH Daily Returns, Monthly Aggregation,$\alpha = 0.05, \beta = 0.93, \alpha + \beta = 0.98$

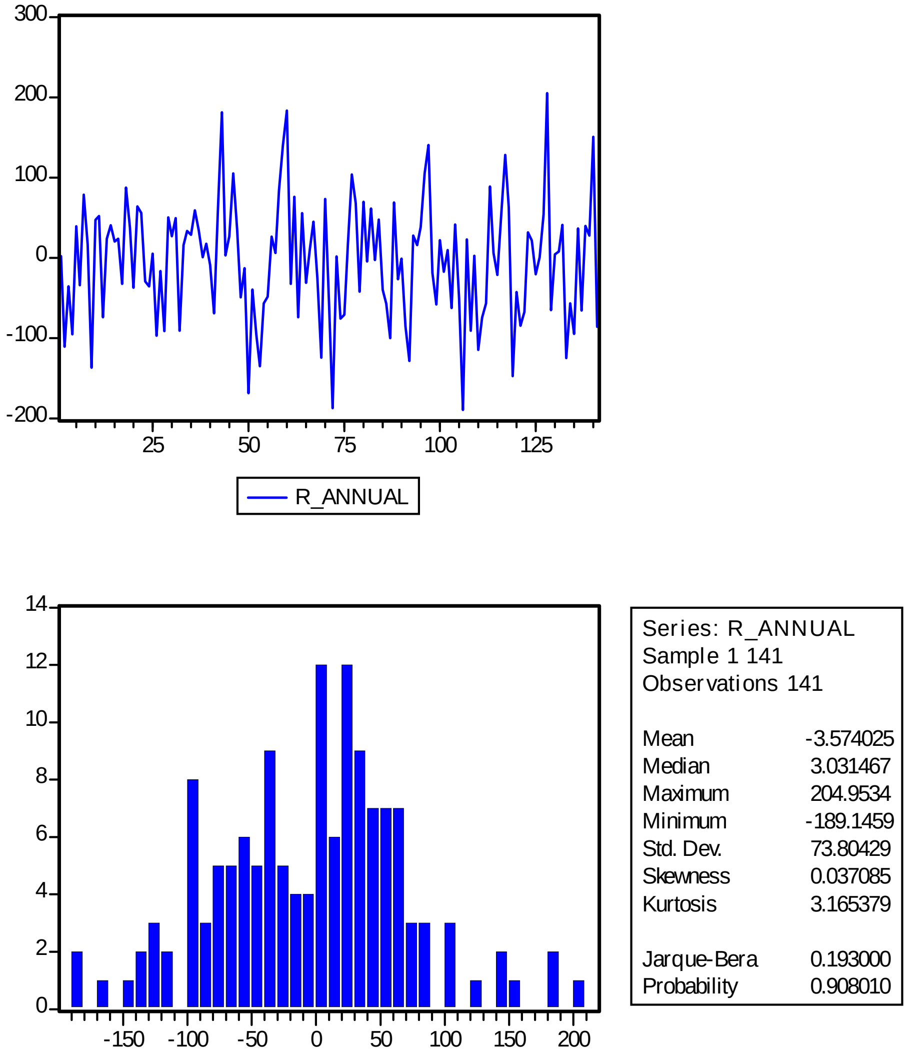


Figure 4.34: Time Plot and Histogram of Simulated GARCH Daily Returns, Annual Aggregation, $\alpha = 0.05, \beta = 0.93, \alpha + \beta = 0.98$

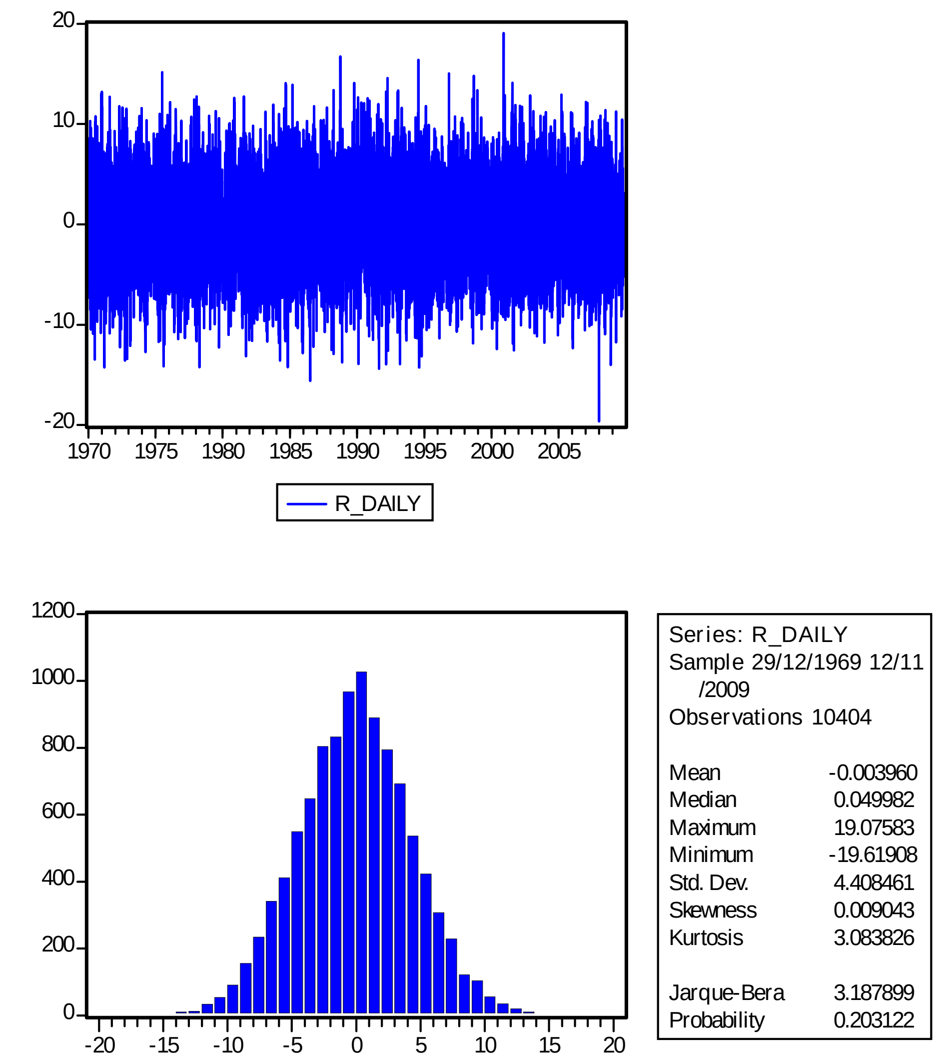


Figure 4.35: Time Plot and Histogram of Simulated GARCH Daily Returns,$\alpha = 0.05, \beta = 0.94, \alpha + \beta = 0.99$

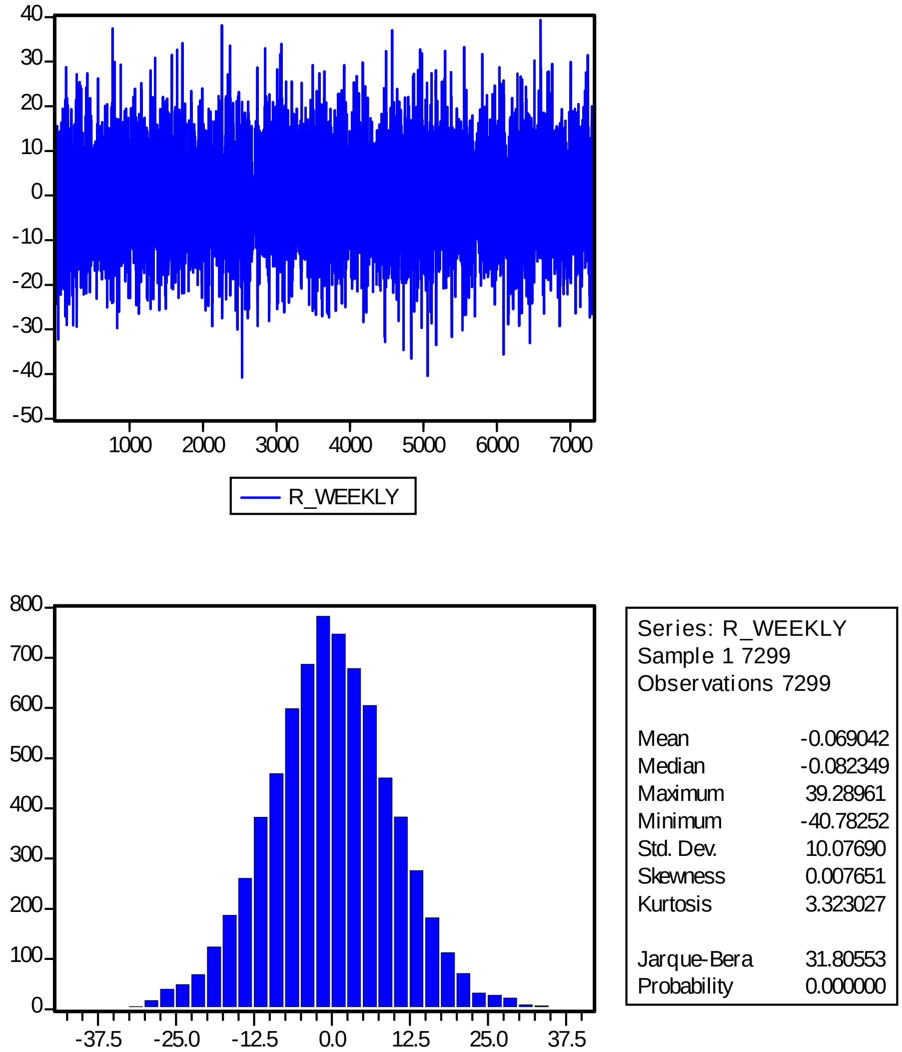


Figure 4.36: Time Plot and Histogram of Simulated GARCH Daily Returns, Weekly Aggregation, $\alpha = 0.05, \beta = 0.94, \alpha + \beta = 0.99$

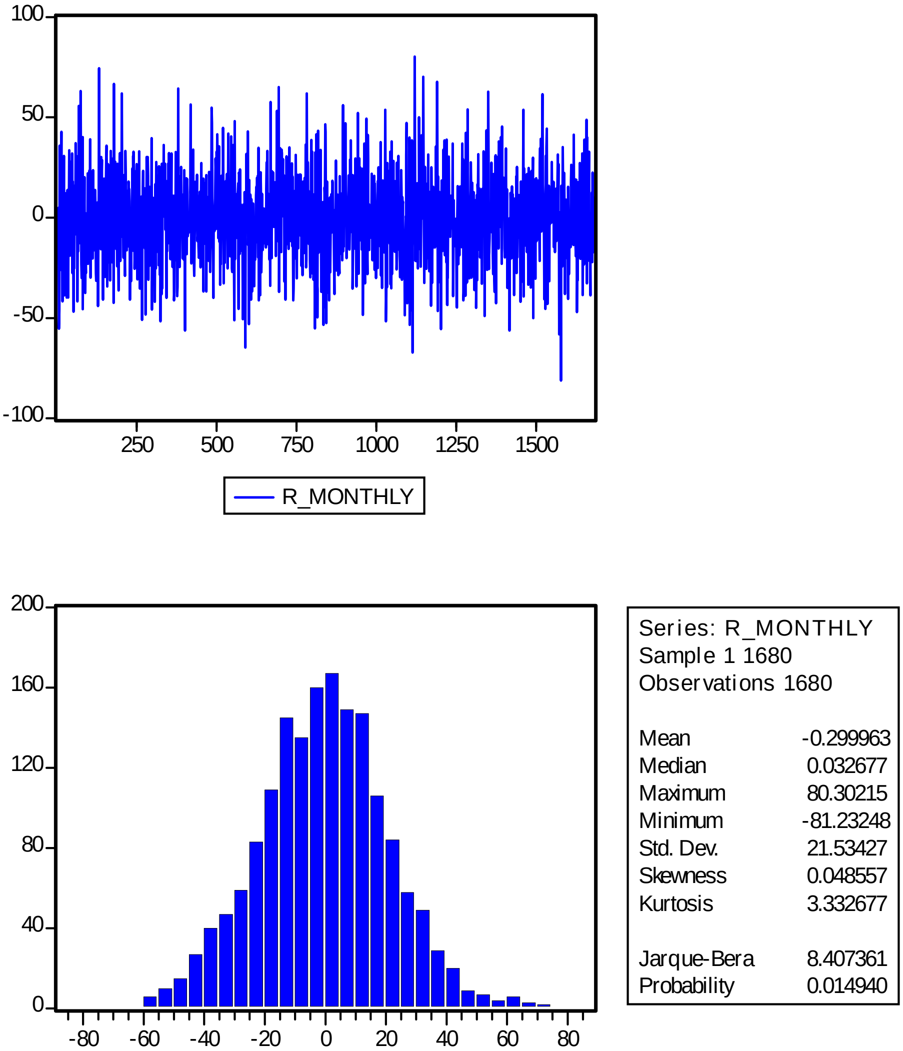


Figure 4.37: Time Plot and Histogram of Simulated GARCH Daily Returns, Monthly Aggregation,$\alpha = 0.05, \beta = 0.94, \alpha + \beta = 0.99$

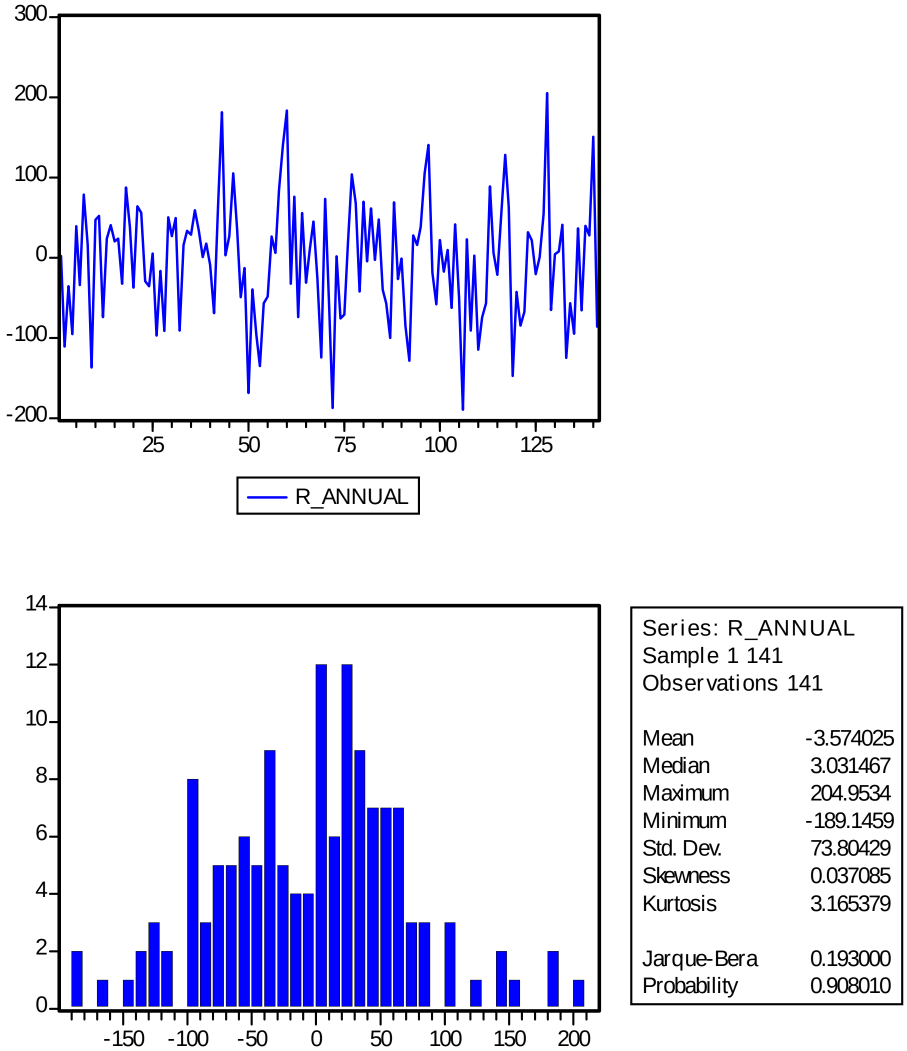


Figure 4.38: Time Plot and Histogram of Simulated GARCH Daily Returns, Annual Aggregation, $\alpha = 0.05, \beta = 0.94, \alpha + \beta = 0.99$

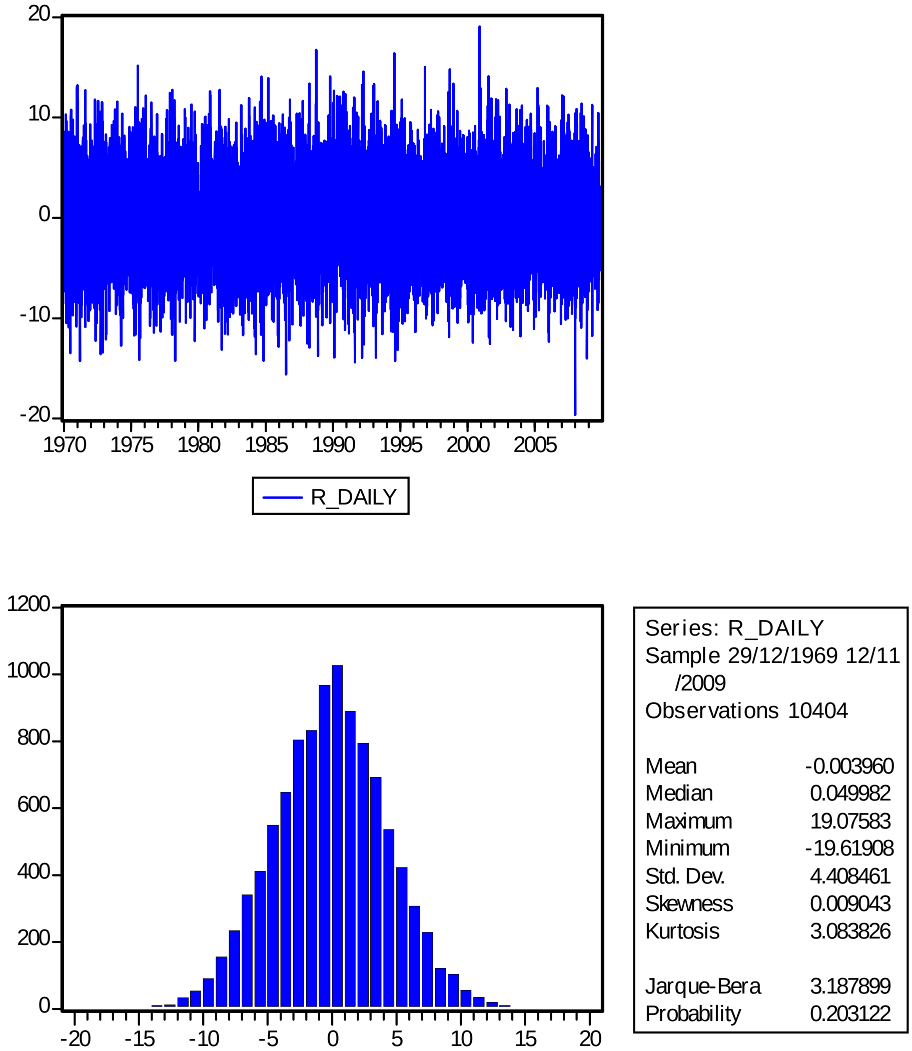


Figure 4.39: Time Plot and Histogram of Simulated GARCH Daily Returns,$\alpha = 0.05, \beta = 0.95, \alpha + \beta = 1$

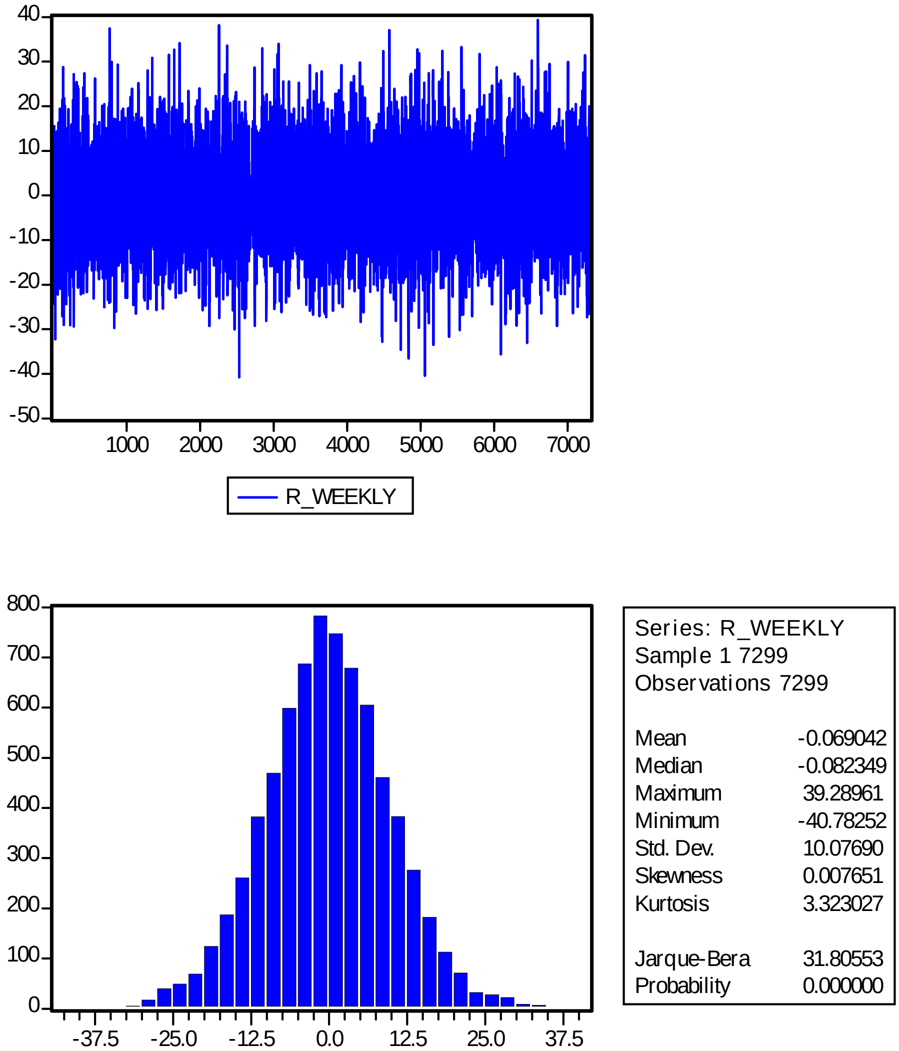


Figure 4.40: Time Plot and Histogram of Simulated GARCH Daily Returns, Weekly Aggregation, $\alpha = 0.05, \beta = 0.95, \alpha + \beta = 1$

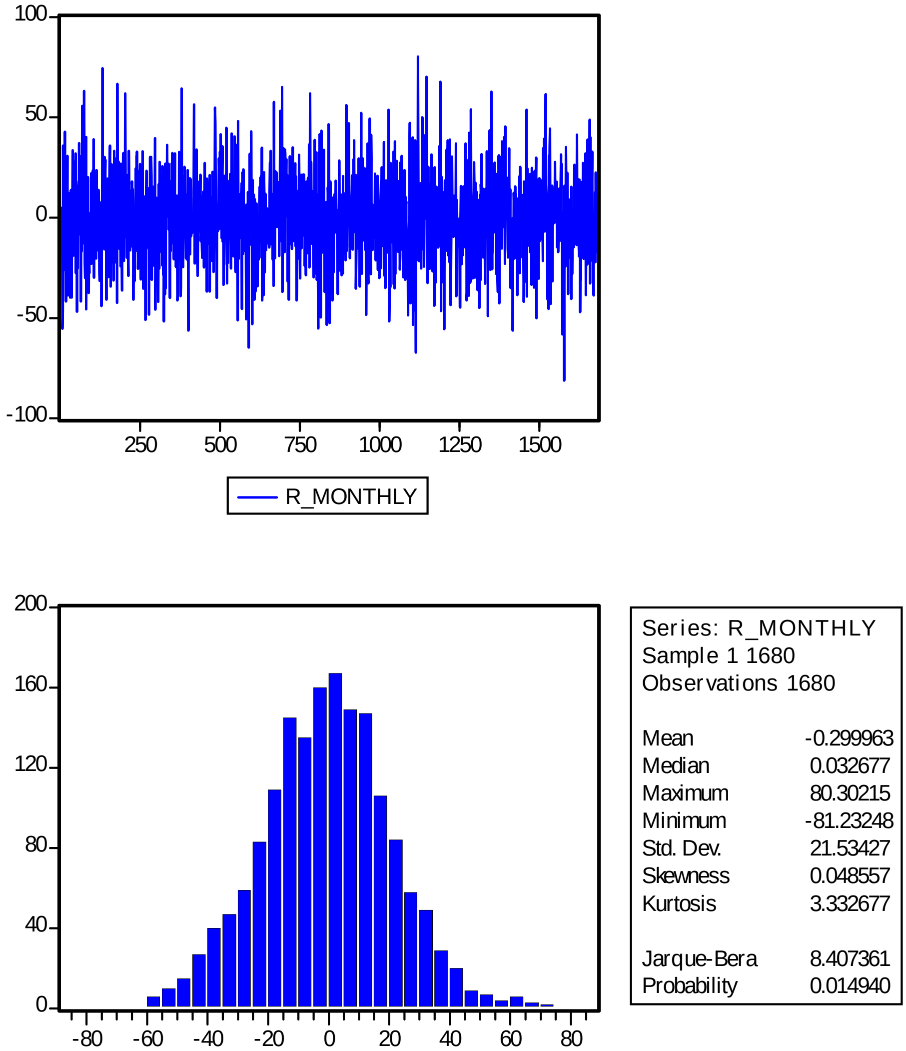


Figure 4.41: Time Plot and Histogram of Simulated GARCH Daily Returns, Monthly Aggregation,$\alpha = 0.05, \beta = 0.95, \alpha + \beta = 1$

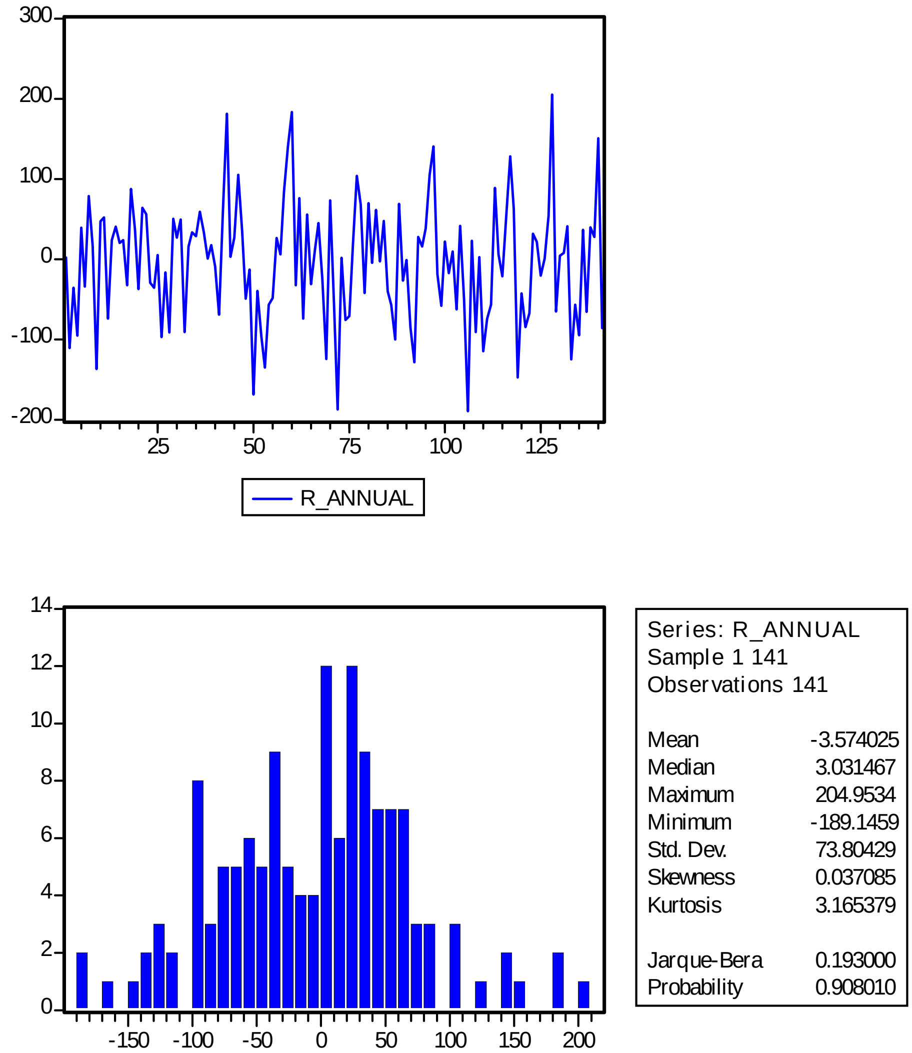


Figure 4.42: Time Plot and Histogram of Simulated GARCH Daily Returns, Annual Aggregation, $\alpha = 0.05, \beta = 0.95, \alpha + \beta = 1$

CHAPTER 5

Selectivity, Market Timing and the Morningstar Star-Rating System

5.1 Introduction

Designing appropriate methods to measure mutual fund (more generally, managed portfolio) performance is an unresolved issue despite the existence of a large body of literature on this topic (see, e.g., Hendricks et al., 1993, Goetzmann and Ibbotson, 1994, and Brown and Goetzmann, 1995), as this requires overcoming a number of difficulties. Traditional, unconditional approaches (see, e.g., Jensen, 1972) have been shown to be unreliable, in the sense that they are unable to distinguish between common time-varying risk (premia) and performance of individual portfolios. An alternative approach was suggested by Ferson and Schadt (1996), who put forward a conditional performance evaluation method. They introduce conditioning (public) information variables into the model and are able to estimate time-varying conditional betas. Their key point is that if it is possible to replicate a managed portfolio strategy using publicly available information then such a portfolio cannot be deemed to outperform the others - in other words, superior information and/or market timing ability cannot be invoked as an explanation in the presence of time-varying risk (premia) which cannot be distinguished from average performance. Applying their model to data for 67 mutual funds over the period 1968-1990, Ferson and Schadt (1996) find that the estimated alphas are close to zero (rather than negative as in the unconditional framework of Jensen, 1968 or Elton et al., 1992); also, they find no evidence of the negative market timing performance reported by previous studies (such as Treynor and Mazuy, 1966 and Merton and Henriksson, 1981), which had interpreted it as an indication of poor performance. Overall, a conditional model leads to the conclusion that funds perform much better than it would be inferred on the basis of a traditional, unconditional evaluation, which overlooks a possibly non-zero covariance between the betas and market returns.

The paper by Ferson and Schadt (1996) and most other studies focus on the performance of mutual funds themselves. By contrast, very little attention has been paid to the usefulness of the Morningstar star-rating system of mutual funds, which is increasingly used by investors to select mutual funds (and as a predictor of future performance, despite the emphasis put by Morningstar on "achievement"). Its importance was documented, for instance, by a survey reported by Damato (1996) in the Wall Street Journal. A few exceptions are the papers by Blume (1998), Sharpe (1998), Khorana and Nelling (1998), and finally Blake and Morey (2000). The latter is most interesting in that, rather than analysing persistence only, it examines the predictive ability of Morningstar ratings for mutual fund performance. The conclusion of this study is that low ratings are indeed associated with poor future performance, whilst it is not at all clear that very high ratings produce a better future performance than slightly lower or average ratings. All in all, Morningstar ratings by themselves appear to have only a slight advantage over alternative methods to predict future fund performance. An intriguing idea has more recently been put forward by Del Guercio and Tkac (2008), who apply an event-study methodology to analyse more than 10,000 Morningstar star-rating *changes* - their evidence suggests that it is the *change* in the star-rating (as opposed to the rating itself) which affects investment flows into or out of mutual funds.

The present chapter focuses on whether Morningstar ratings themselves enable investors to select funds that are likely to exhibit superior performance in the future. Instead of relying on out-of-sample performance measures as in Blake and Money (2000), we conduct a full unconditional as well as conditional performance evaluation of the Morningstar rating system using the framework advocated by Ferson and Schadt (1996). More in detail, we proceed as follows. We create five alternative portfolios of funds (funds-of-funds). The first portfolio, named STAR1, consists of all the funds that in each time period, t, are rated one-star by Morningstar. To be more specific, in period t=1 (the first period in our sample) we invest an amount A$ in a portfolio consisting of all the funds (equally weighted) that have been given one-star by Morningstar in period t=1. In period t=2, the amount $(1+\mathrm{R}_1^{1*})\times A$ (R_1^{1*} being the return of the portfolio between periods 1 and 2) is invested again in a portfolio consisting solely of funds that in period 2 were rated one-star by Morningstar. We continue this process until we reach period t=T, i.e. the last period of our sample. In this way, we obtain a series of returns R_1^{1*},R_2^{1*}, ...,R_T^{1*} generated by investing exclusively in one-star funds. These are interpreted as being a random vector from the process $\{R_t^{1*}\}$ generating one-star portfolio returns. We repeat the same procedure for two-, three-, four- and five-star funds, thus obtaining samples from the returns processes $\{R_t^{2*}\}$, $\{R_t^{3*}\}$, $\{R_t^{4*}\}$, $\{R_t^{5*}\}$, which are supposed to generate returns for the two-, three-, four- and five-star funds respectively. We are interested in examining whether the statistical properties of these five returns processes are different.

To put it differently, we wish to evaluate the following simple investment strategy: if in each time period we create a portfolio consisting only of five-star funds, are the risk-adjusted returns on this portfolio higher than the corresponding ones on a portfolio consisting solely of, say, two-star funds? The idea is that if the better performance of five-star fund is really due to superior management skills then these should be reflected in the returns on a fund including only five-star funds.

Our study does not attempt to evaluate the performance of individual mutual funds. Rather it aims to assess “an evaluation procedure”, namely that of Morningstar. In other words, if we create a portfolio consisting only of those funds that in each time period have received a five-star (or four-, three-, two-, one-) score by Morningstar, and then evaluate the risk-adjusted performance of this portfolio by more traditional portfolio evaluation procedures, what would the result be? Will these traditional evaluation procedures detect any differences in the risk-adjusted performance of the one-, two-, three-, four- and five-star portfolios?

The remainder of the chapter is organised as follows. Section 2 describes the data and analyses the statistical properties of the returns on the five funds-of-funds we construct. Section 3 outlines the standard and conditional CAPM models we adopt to address the issue of whether the better performance of higher-rated funds is in fact attributable to superior management skills. Section 4 discusses the empirical results from both unconditional and conditional portfolio performance evaluation. Section 5 summarises the main findings and offers some concluding remarks.

5.2 Statistical Properties of the Five Funds-of-Funds Returns

The data used in our study were taken from Morningstar Direct, which provides historical monthly returns of selected mutual funds and their star- based ranking calculated by Morningstar. We focus on the subset represented by equity mutual funds, that is funds that invest at least 90% of their Non-cash Adjusted Total Assets in equity securities. To avoid dealing with currency risk exposure we only consider funds quoted in US Dollars. At present, the Morningstar Direct database contains 21322 equity funds in US Dollars. In order to perform our evaluation we need sufficiently long series, and therefore we have restricted our sample to funds that have been star-rated for at least 10 years. This reduces the sample to 1511 equity funds. For these funds, historical returns and their Morningstar ranking are available since 01/1998.

We begin our analysis of the returns of the five funds-of-funds, STAR1, STAR2, STAR3, STAR4 and STAR5 defined in the previous section by examining the univariate properties of $\{R_t^{1*}\}$, $\{R_t^{2*}\}$, $\{R_t^{3*}\}$, $\{R_t^{4*}\}$, and $\{R_t^{5*}\}$. Table 1A reports descriptive statistics for each of them, together with their first-order autocorrelation. Table 1B presents estimates of the correlation matrix. The results can be summarised as follows:

(i) The sample mean of returns is an increasing function of the number of stars. The average monthly returns of STAR1 and STAR5 are 0.31% and 0.51% respectively.

(ii) The pattern is less clear for sample standard deviations: the smallest is exhibited by STAR3, followed by STAR4 and STAR2. The standard deviation of STAR5 is the second highest, after that of STAR1.

(iii) The distributions of all the five returns exhibit negative skewness, this being highest for STAR2 and lowest for STAR5.

(iv) All the five returns series are serially uncorrelated, i.e. their degree of persistence is zero.

(v) The estimated correlation coefficients are very large. They range from 0.87 (the coefficient between STAR1 and STAR5) and 0.98 (the coefficient between STAR1 and STAR2, STAR2 and STAR3 and also STAR3 and STAR4). With correlation coefficients so close to unity, it is rather unlikely that any differences in the risk-adjusted performance of these funds-of-funds will be detected whatever the definition of "risk".

The greatest difference in mean returns is between STAR5 and STAR1, for which the smallest correlation coefficient (0.87) is also obtained. To examine whether the mean return of STAR5 is statistically different from the mean return of the other four portfolios, we generate the return-differential series, $\Delta R_t^{5i*} = R_t^{5*} - R_t^{i*}$, $i = 1, 2, 3, 4$, and test, by means of a t-test, whether the means, μ_i, of ΔR_t^{5i*} are different from zero. This can be done by running a regression of ΔR_t^{5i*} on a constant term, c_i, that is $\Delta R_t^{5i*} = c + \nu_{it}$ and testing the significance of the coefficient c_i, $i = 1, 2, 3, 4$. The results are reported in Table 1C, together with a series of misspecification tests, in order to provide some information on the time series properties of the four return-differential series, and establish whether the conditions are met for the employed t-test to have good properties. Specifically, in addition to the usual tests of serial correlation in ν_{it}, we also report the results from testing for the presence of non-linear temporal dependence in ν_{it}, i.e. the first-order autocorrelation of the squared residuals ν_{it}^2 along with the Ljung-Box $Q(l)$-statistic for testing the hypothesis that the first l autocorrelations are equal to zero. Moreover, we report the well-known BDS test proposed by Brock, Dechert, Scheinkman and LeBaron (1996), which is designed to test the stronger assumption that the noise series, ν_{it}, is independent and identically distributed (i.i.d.). The results can be summarised as follows:

(i) The null hypothesis that the mean of return-differentials is equal to zero is not rejected for any of the four series. Nevertheless, it might be worth noting that the point estimates of the mean as well as the value of the t-statistic increase with the "star-difference" (that is, they are bigger for ΔR_t^{51*} followed by $\Delta R_t^{52*}, \Delta R_t^{53*}, \Delta R_t^{54*}$).

(ii) The noise series ν_{it}, and hence the series ΔR_t^{5i*}, are not serially correlated but they exhibit strong second-order temporal dependence. In particular, the BDS test strongly rejects the null hypothesis that the series ΔR_t^{5i*} are i.i.d.

On the basis of the above results, we examine whether the inability to reject the null hypothesis of zero return-differentials by the t-tests might be due to the presence of non-linear dependence in ν_{it}, which has not been taken into account. Specifically, we re-estimate the models $\Delta R_t^{5i*} = c + \nu_{it}$, assuming that the errors ν_{it} follow GARCH(1,1) processes, that is $\nu_{it} = h_{it}\varepsilon_{it}$, $h_{it}^2 = d_i + a_i h_{it-1}^2 + b_i \varepsilon_{it-1}^2$. The results, reported in Table 1C(ii), suggest the following:

(iii) When second-order dependence is taken into account, the statistical inference on the existence of differentials in the average returns among the star-rated funds changes drastically. The null hypothesis that the mean return-differential is zero is rejected for ΔR_t^{51*}, ΔR_t^{52*}, and ΔR_t^{53*} at the 5% level, and even for ΔR_t^{54*} at the 10% level. This means that the star-rating system of Morningstar does produce a classification of funds which exhibit some significant differences in terms of their average monthly returns.

(iv) The results reported above seem to be trustworthy since the hypothesis that the standardised noise series, $\varepsilon_{it} = \nu_{it}/h_{it}$ are i.i.d. is not rejected by the BDS test. Moreover, additional tests (not reported) for the presence of structural breaks within the sample seem to support the hypothesis that the standardised error process $\{\varepsilon_{it}\}$ is identically distributed.

5.3 Risk-Adjusted Returns: The Models

In this section we investigate whether the higher average returns of the five-star funds are the reward for the additional risk that the managers of these funds bear relative to the risk incurred by the managers of, say, three-star funds or reflect instead superior management skills of the managers of the five-star funds.

For this purpose, we consider an asset pricing model that describes the relationship between the expected return and risk of the various portfolios under consideration. Specifically, we adopt the conditional CAPM model of Ferson and Schadt (1996) and Shanken (1990) in which the level of the systematic portfolio risk is a function of the observed variables (see also, Lettau and Ludvingson 2001). This in turn implies that the relationship between the excess returns of the portfolio j and the excess returns of the market factor is given by the following relationships:

$$r_{j,t+1} = b_j(Z_t)r_{m,t+1} + \varepsilon_{j,t+1} \tag{5.1}$$

$$E(\varepsilon_{j,t+1} \mid Z_t) = 0 \tag{5.2}$$

$$E(\varepsilon_{j,t+1}r_{m,t+1} \mid Z_t) = 0 \tag{5.3}$$

where $r_{jt} = R_t^{j*} - R_{ft}$, $i = 1, 2, ..., 5$, R_{ft} is the return of a one-month Treasury bill, $Z_t = [Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t}]$ is an n-vector of state variables observable by the managers at time t, and $r_{mt} = R_t^m - R_{ft}$ with R_t^m stands for the returns of the market factor. This specification implies that the systematic risk of the portfolio j, as measured by $b_j(Z_t)$, changes with time. The time-varying nature of beta is due to the fact that the portfolio manager receives at time t an “information signal”, contained in the state variables Z_t, on the basis of which he changes the beta of his portfolio. If the signal is “correct” and the manager succeeds in “receiving” it, then the changes in the beta of the portfolio at time t will be consistent with the realized returns $r_{m,t+1}$ at time t+1. To put it differently, if $r_{m,t+1} > 0$ then the correct interpretation of the signal implies that the manager will shift the portfolio towards including stocks with high betas. The preceding discussion implies that the ability of the fund manager to "time" the market depends on the extent to which he/she can translate the information content of Z_t into predictions on the future behavior of $r_{m,t+1}$. This does not necessarily mean that "everybody" in the market can "read" the information contained in Z_t. In other words, although the variables Z_t are indeed publicly available, the information content of Z_t might be available only to a "skilful" fund manager.

The next question concerns the specification of the function $b_j(Z_t)$. Since the true functional form is unknown, we shall approximate it by using a first or second-order Taylor series expansion. We begin by applying a first order approximation of β_j, in which case equation (5.1) becomes

$$r_{j,t+1} = \beta_{j,0} + \sum_{i=1}^{n} \beta_{j,i} Z_{i,t} r_{m,t+1} + \varepsilon_{j,t+1} \tag{5.4}$$

It is quite natural to assume that the dependent variable, $r_{j,t+1}$, in (5.4) and $r_{m,t+1}$ are I(0) variables. However, the stationarity of $Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t}$ cannot be assumed a priori. In the absence of stationarity of those variables, we may face the problem that (5.4) is an "unbalanced" regression. As a result, before we proceed any further we must analyse in detail the alternative models (all based on (5.4)) that arise depending on the statistical properties of the variables $Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t}$. Specifically, we need to distinguish three cases:

(i) All the variables $Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t}$ are I(0). In this case, the new variables $Z_{i,t} r_{m,t+1}$ will also be I(0) (given that the market returns variable $r_{m,t+1}$ is quite naturally I(0)), and equation (5.4) is legitimate since $r_{j,t+1}$ is also quite naturally I(0).

(ii) Some (or all) of the variables $Z_{1,t}, Z_{2,t}, \ldots, Z_{n_0,t}$, $n_0 \leq n$ are I(1) and not cointegrated. In this case, the product variables $Z_{i,t} r_{m,t+1}$, $i = 1, 2, ..., n_0$ will have an asymptotically unbounded unconditional variance and will not be I(0). In such a case, we have the problem of an "unbalanced regression" since the dependent variable, $r_{j,t+1}$ is I(0).

(iii) Some (or all) of the variables $Z_{1,t}, Z_{2,t}, \ldots, Z_{n_0,t}$, $n_0 \leq n$ are I(1) and cointegrated.

In this case, we proceed as follows: First, we rewrite (5.4) as

$$r_{j,t+1} = \beta_{j,0} + r_{m,t+1} \sum_{i=1}^{n} \beta_{j,i} Z_{i,t} + u_{j,t+1}. \tag{5.5}$$

Equation (5.5) shows that under the assumption that the returns processes $\{r_{j,t}\}_{t\geq 1}$ and $\{r_{m,t}\}_{t\geq 1}$ are $I(0)$, the only case where some of the $Z_{i,t}$ can be $I(1)$ is when the corresponding coefficients are such that only multiples of the cointegration relations between those $Z_{i,t}$s that are cointegrated are left on the right-hand side of (5.5).

To see the difference between cases (ii) and (iii) more clearly, let us examine the following example:

$$r_{j,t+1} = \beta_{j,0} + \left(\beta_{j,1} Z_{1,t} + \beta_{j,2} Z_{2,t}\right) r_{m,t+1} + u_{j,t+1} \tag{5.6}$$

If $Z_{1,t}$ and $Z_{2,t}$ are not cointegrated and $\beta_{j,1}\beta_{j,2} \neq 0$, the unconditional variance of the right-hand side will grow to infinity as $t \to \infty$, violating our initial assumption that the unconditional variance of $r_{j,t+1}$ is bounded. Therefore the estimated values of $\beta_{j,1}$ and $\beta_{j,2}$ will be very close to 0 when the sample is large. If, on the other hand, $Z_{1,t}$ and $Z_{2,t}$ are cointegrated and satisfy

$$Z_{1,t} = a_0 + a_1 Z_{2,t} + w_t$$

where $\{w_t\}_{t\geq 1}$ is $I(0)$, the only way for the unconditional variance of $\beta_{j,1} Z_{1,t} + \beta_{j,2} Z_{2,t}$ to remain asymptotically bounded, with nonzero $\beta_{j,1}$ and $\beta_{j,2}$, is the case where

$$\beta_{j,1} Z_{1,t} + \beta_{j,2} Z_{2,t} = \lambda \left(Z_{1,t} - a_1 Z_{2,t}\right) = \lambda \left(a_0 + w_t\right), \, \lambda \in \mathbb{R}.$$

The last equation implies that

$$\beta_{j,1} = \lambda$$

and

$$\beta_{j,2} = -a_1 \beta_{j,1} \, .$$

If we identify the cointegrating relationship between $Z_{1,t}$ and $Z_{2,t}$, then we can rewrite

(5.6) as

$$r_{j,t+1} = \beta_{j,0} + \beta_{j,1} q_{1,t} r_{m,t+1} + u_{j,t+1} \tag{5.7}$$

where $q_{1,t} = Z_{1,t} - a_1 Z_{2,t} = a_0 + w_t$.

The previous example demonstrates that it is necessary for the sum $\sum_{i=1}^{n} \beta_{j,i} Z_{i,t}$ in (5.5) to be $I(0)$ in order for (5.5) to be a legitimate regression. It also suggests how to treat the initial set of candidate state variables, in order to obtain in (5.5) a well-balanced regression. Specifically, the following steps must be taken:

First, we identify all those state variables (elements of Z_t) that are $I(1)$. Assume that the number of such I(1) variables is n_0. If $n_0 > 0$, without loss of generality, reordering the variables if necessary, we can make sure that, for $i \leq n_0$, $\{Z_{i,t}\}_{t\geq 1}$ are $I(1)$ and, for $n_0 < i \leq n$, $\{Z_{i,t}\}_{t\geq 1}$ are $I(0)$. Second, we identify any cointegrating relationships between the processes $\{Z_{i,t}\}_{t\geq 1}$, $1 \leq i \leq n_0$. Let $k < n_0$ be the rank of the cointegrating system. This means that we can find a $(k \times n_0)$ matrix A of order k, such that

$$A \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{n_0,t} \end{bmatrix} = U_t, \tag{5.8}$$

where $\{U_t\}_{t\geq 0}$ is $I(0)$ with nontrivial coordinates $U_{i,t}$, $1 \leq i \leq k$. Again, without any loss of generality, we can reorder the variables $Z_{i,t}$, $1 \leq i \leq n_0$ in (5.5), so that the first k columns of A are linearly independent. Therefore, we can write $A = [A_1, A_2]$, where the $k \times k$ matrix A_1 is invertible. Then, left multiplication of (5.8) by A_1^{-1} yields

$$\left[I_k, A_1^{-1} A_2\right] Z_t = A_1^{-1} U_t \ ,$$

which in turn gives

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{k,t} \end{bmatrix} = -A_1^{-1} A_2 \begin{bmatrix} Z_{k+1,t} \\ Z_{k+2,t} \\ \vdots \\ Z_{n_0,t} \end{bmatrix} + A_1^{-1} U_t. \tag{5.9}$$

The last equation is the first part of Phillips's (1991) triangular system. The second part of this system is

$$\begin{bmatrix} \Delta Z_{k+1,t} \\ \Delta Z_{k+2,t} \\ \vdots \\ \Delta Z_{n_0,t} \end{bmatrix} = V_t\ , \tag{5.10}$$

where V_t is also $I(0)$. Equations (5.8) and (5.10) provide us with n_0 $I(0)$ processes that can be considered as state variables in a new regression, replacing the $Z_{i,t}$, $1 \leq i \leq n_0$. Third, having defined the appropriate set, W_t, of I(0) state variables,

$$W_t = (W_{i,t})_{1 \leq i \leq n} = [U_{1,t}, U_{2,t}, \ldots, U_{k,t}, \Delta Z_{k+1,t}, \ldots, \Delta Z_{n_0,t}, Z_{n_0+1,t}, \ldots, Z_{n,t}]'$$

we can define the following regression

$$r_{j,t+1} = \beta^*_{j,0} + \sum_{i=1}^{k} \beta^*_{j,i} U_{i,t} r_{m,t+1} + \sum_{i=k+1}^{n_0} \beta^*_{j,i} \Delta Z_{i,t} r_{m,t+1} + \sum_{i=n_0+1}^{n} \beta^*_{j,i} Z_{i,t} r_{m,t+1} + u_{j,t+1}. \tag{5.11}$$

This regression can be rewritten in a more compact form in terms of W_t as

$$r_{j,t+1} = \beta^*_{j,0} + \sum_{i=1}^{n} \beta^*_{j,i} W_{i,t} r_{m,t+1} + u_{j,t+1}, \tag{5.12}$$

or simply

$$r_{j,t} = \beta^*_{j,0} + \sum_{i=1}^{d} \beta^*_{j,i} X_{i,t} + \zeta_{j,t}, \tag{5.13}$$

where $X_{i,t} = W_{i,t-1} r_{m,t}$ for $1 \leq i \leq n$. The regression defined in (5.12) can be considered a first order approximation of the general model given by

$$r_{j,t+1} = \beta^*_{j,0} + \beta^*_j(W_t)r_{m,t+1} + \varepsilon^*_{i,t+1}, \tag{5.14}$$

which involves only $I(0)$ processes.

The preceding discussion is based on approximating the unknown function $b_j(Z_t)$, or equivalently $\beta^*_j(W_t)$, by using a first-order Taylor series expansion. Alternatively, we can approximate $\beta^*_j(W_t)$ by using a second-order Taylor expansion. In such a case we have

$$r_{j,t+1} = \beta^*_{j,0} + \sum_{i=1}^{n} \beta^*_{j,i} W_{i,t} r_{m,t+1} + \sum_{1 \le i \le k \le n} \beta^*_{j,i,k} W_{i,t} W_{k,t} r_{m,t+1} + \zeta_{j,t+1}. \tag{5.15}$$

The last equation involves $n + \begin{pmatrix} n \\ 2 \end{pmatrix} = \frac{n(n+1)}{2} \triangleq d$ explanatory variables of the form $W_{i,t}r_{m,t+1}$ or $W_{i,t}W_{k,t}r_{m,t+1}$, $1 \le i \le k \le n$, which can be denoted as $X_{l,t+1}$, $1 \le l \le d$.

We can rewrite (5.15) as:

$$r_{j,t} = b_{j,0} + \sum_{i=1}^{d} b_{j,i} X_{i,t} + \zeta_{j,t} \tag{5.16}$$

where $b_{j,0} = \beta^*_{j,0}$, $b_{j,i} = \beta^*_{j,i}$ and $X_{i,t} = W_{i,t-1}r_{m,t}$ for $1 \le i \le n$, $b_{j,i} = \beta^*_{j,g,h}$ and $X_{i,t}$ is of the form $W_{g,t-1}W_{h,t-1}r_{m,t}$, when $n+1 \le i \le d$, for some $1 \le g, h \le n$.

The use of polynomial approximations of (5.14) may simplify its treatment, but if the order of approximation is underspecified, the estimation residuals are likely to be affected by the missing part of β^*_j. This is true even in the simple case where β^*_j is a second-order polynomial. For example, let the true model be:

$$r_{j,t+1} = \beta^*_{j,0} + \beta^*_{j,1} W_{1,t} r_{m,t+1} + \beta^*_{j,2} W_{2,t} r_{m,t+1} + \beta^*_{1,2,k} W_{1,t} W_{2,t} r_{m,t+1} + \zeta_{j,t+1}$$

and assume that a first-order approximation is used:

$$r_{j,t+1} = \beta^*_{j,0} + \beta^*_{j,1} W_{1,t} r_{m,t+1} + \beta^*_{j,2} W_{2,t} r_{m,t+1} + u_{j,t+1}.$$

One can see that any estimation of $\beta^*_{j,0}$, $\beta^*_{j,1}$, and $\beta^*_{j,2}$ using the first-order approximation will have to compensate for the missing expected value of $\beta^*_{1,2,k}W_{1,t}W_{2,t}r_{m,t+1}$. A non-zero expectation of $W_{1,t}W_{2,t}r_{m,t+1}$ will result in biased estimates of some or all of $\beta^*_{j,0}$, $\beta^*_{j,1}$, and $\beta^*_{j,2}$. Another implication concerns the second-order properties of $u_{j,t+1}$: it is clear that the behavior of the estimated $u'_{j,t+1}$s will be related to the missing term $\beta^*_{1,2,k}W_{1,t}W_{2,t}r_{m,t+1}$, and therefore these will exhibit conditional heteroscedasticity.

Finally, note that all the above models can be augmented by the market timing term, $\gamma_j r^2_{m,t+1}$, proposed by Traynor and Mazui (1966). A positive (negative) timing coefficient γ_j is interpreted as evidence suggesting superior (inferior) market timing abilities of the corresponding fund manager.

5.3.1 Empirical Results

Unconditional Portfolio Performance Evaluation

We begin our empirical analysis by considering the so-called unconditional evaluation of the star-rated funds-of-funds under consideration, which is based on the standard version of CAPM. The latter assumes that $b_j(Z_t) = b_j \ \forall t$. Under this hypothesis, the fund managers do not attempt to "time" the market, so they do not actively change the betas of their portfolios. Moreover this assumption also implies that the betas of the assets forming the portfolios do not change over time, or if they change the changes in the beta of one asset are exactly offset by those in the beta of another asset. Table 2A reports the results from the OLS estimation of equation

$$r_{j,t} = b_j r_{m,t} + \gamma_j r^2_{m,t} + \varepsilon_{j,t+1} \tag{5.17}$$

Given that the error terms in all the five regressions exhibit conditional heteroscedasticity, we report the results based on heteroscedasticity-consistent (HC) standard errors in Table 2B, and those from explicitly assuming that the errors are GARCH(1,1) processes in Table 2C. The results can be summarised as follows:

(i) The OLS and HC estimates are quite similar, suggesting that the unconditional estimates of a for all the portfolios except STAR1 are positive and significantly different from zero. Interestingly, the highest a is achieved by STAR5. On the contrary, the estimates of the market timing coefficient γ appears to be insignificantly different from zero for all portfolios.

(ii) When conditional heteroscedasticity is taken into account, the results change significantly. In particular, the stock selection coefficient, a, now appears to be significantly positive only for STAR3, STAR4 and STAR5. Moreover, the GARCH-based estimates of a for all five portfolios appear to be smaller than the corresponding OLS-based estimates. For example, the OLS and GARCH estimates of a for STAR5 are 0.52 and 0.31, respectively. The GARCH estimates of γ (similarly to the corresponding OLS estimates) are insignificantly different from zero. The overall picture emerging from the GARCH estimates suggests that the five portfolios under consideration can be classified into two groups. The first one consists of STAR1 and STAR2, and is characterised by neither stock selection nor market timing abilities. The second one includes STAR3, STAR4 and STAR5, and exhibits some stock selection ability (which is almost identical among the three portfolios belonging to this group) but no market timing ability.

(iii) It is worth noting that, despite the constant beta assumption in this unconditional fund performance evaluation, the estimates of a are generally positive. This is in contrast with the early results of Jensen (1968) and the subsequent results of Elton et. al (1992) (among others) who report negative estimates of a, which may be caused by (unaccounted) time variation in the betas. This in turn implies that, despite the bias in the estimates of a, caused by the possible time variation of the betas, a positive stock selection ability can be inferred for STAR3, STAR4 and STAR5.

(iv) Related to (iii): Diagnostic tests (not reported) for parameter stability of equation (5.17) estimated by OLS indicate the presence of significant time variation in the parameters of this model. However, this instability may be the result of omitted conditional heteroscedasticity (i.e. it may come from time variation of the standard error of the regression). When the GARCH effects are taken into account the observed instability is reduced, though not entirely eliminated. The possible time variation of beta, in particular, which may come from a response of the fund manager to changing economic conditions, is the focus of the next subsection.

Conditional Portfolio Performance Evaluation

The next issue is the selection of the variables in Z_t. Ferson and Schadt (1996) suggest including the one-month Treasury bill yield, z_{1t}, the term spread, z_{2t}, defined as the difference between the constant-maturity 10-year Treasury bond yield and the 3-month Treasury bill, and the quality spread, z_{3t}, in the corporate bond market defined as the Moody's BAA-rated corporate bond yield minus the AAA-rated corporate bond yield. In addition, we include variables that are usually considered important indicators by the financial community such as the weighted average of the foreign exchange value of the US dollar against a subset of the broad index currencies, z_{4t}, the Consumer Sentiment Index of the University of Michigan, z_{5t}, the price of oil, z_{6t}, and the Chicago Board Options Exchange volatility index (VIX), z_{7t}.

As explained in the previous section, the choice of the appropriate model for conditional portfolio evaluation depends on the statistical properties of the state variables z_{1t}, $z_{2t}, ..., z_{7t}$. The results from a variety of unit root tests, reported in Table 3A, unambiguously indicate that the first six series are I(1) while the last one is I(0). However, the tests on the cointegration properties of z_{1t}, $z_{2t}, ..., z_{6t}$, reported in Table 3B, lead to less clear-cut conclusions. In particular, when the lag-length, l, of the Vector Autoregressive model, VAR(l), on which the two tests are based, is relatively large, both test statistics suggest a cointegration rank, k, of at least one, and occasionally two. On the contrary, when l is relatively small, both tests are unable to reject the null hypothesis of no cointegration. As a result, we run three alternative conditional regressions assuming $k = 0$, $k = 1$ and $k = 2$, with the results (assuming GARCH(1,1) errors and including the market timing term $\gamma_j r^2_{m,t+1}$) being reported in Table 4A, 4B and 4C, respectively. The results can be summarised as follows:

(i) The results are very robust across the three alternative cointegration rank assumptions. Indeed, the information criteria for $k = 0$ are very close to those for $k = 1$ or $k = 2$ for all five portfolios under consideration.

(ii) Despite the significance of (some of) the state variables, the results on stock selection (a) and market timing (γ) abilities from the conditional evaluation are similar to those obtained from the unconditional evaluation. In particular, for $k = 0$ the stock selection coefficient, a, appears to be significantly positive for STAR2, STAR3, STAR4 and STAR5, whereas for $k = 1$ and $k = 2$ it appears to be significantly positive for STAR3, STAR4 and STAR5. Concerning the latter portfolio, the highest estimate of (conditional) a is 0.36, obtained for $k = 2$, whereas the lowest is equal to 0.30, obtained for $k = 1$. Concerning market timing ability, no portfolio for any value of k produces a significantly positive estimate of γ_j. On the contrary the estimates of γ_j are negative and in many cases significantly so.

(iii) The results in (ii) suggest that the positive excess returns produced by STAR3, STAR4 and STAR5 should be thought of as the result of superior stock selection rather than market timing abilities.

5.4 Combining Morningstar Rating System with Asset Allocation Strategies

The information provided by the Morningstar Rating System, regarding the relative comparison of mutual funds, is only one component that an investor might utilize for her asset allocation decisions. Depending on the degree of efficiency, investors will always seek for valuable information, like the Morningstar Rating System, that will allow them to try to outperform the market. In this section, we compare the performance of common asset allocation strategies, namely momentum and contrarian, when applied either unconditionally, that is with no additional information, on the universe of our dataset or conditionally on the information provided by Morningstar, that is we apply these strategies on the subsets of mutual funds with different star rating. In addition, we considered the case where the investor follows a naive strategy of randomly selecting her assets. We bask-tested all suggested strategies, for the whole sample period, that is January 1998 untiil September 2008, assuming that each month the investo re-allocates her assets.

We constructed portfolios consisting of $n = \{5, 15, 30\}$ mutual funds.In the case of unconditional momentum (contrarian) strategy, we invested in the n funds that during the previous month, succeeded the highest (lower) monthly returns among all funds included in our dataset. In the case of conditional momentum (contrarian) strategy, we followed the same decision rule but this time, we invested in the n funds with the highest (lower) monthly return among the funds that during the previous month were rated the same. When we applied the naive strategy, either unconditional or conditional, n assets were randomly, using a random number generator of a uniform random variable, chosen from the corresponding subset of mutual funds. For brevity, we outline the main results.of this analysis, but in the appendix of this chapter, we included all the bask-testing strategies considered here.

Figure (5.1) demonstrates that momentum strategy works remarkably better in comparison with the two alternatives. In most cases, contrarian strategy seems not to be the most adequate strategy when investing in a fund of funds. In Figure () we can see that applying momentum strategy, using the information provided by Morningstar, results in increasing portfolio performance. In general we cannot observe a pattern that moving from 1 Star rated fund of funds to higher rates, will engage higher performance for the momentum strategy. Finally we can see that unconditional momentum strategy results in higher portfolio performance, but as we can see from Table (5.1) this is achieved with the cost of increasing risk, therefore resulting with a lower Sharpe ratio from the best conditional case which is a Momentum Strategy applied on 4 - star rated funds.The last column in Table (5.1), displays the p-value of the Null Hypothesis that Portfolio Mean Return is 0.

5.5 Conclusions

This chapter adds to the rather limited number of studies which to date have attempted to evaluate the Morningstar star-based system for ranking mutual funds (see, e.g., Blake and Morey, 2000). Its aim is to provide evidence on whether portfolios assembled using

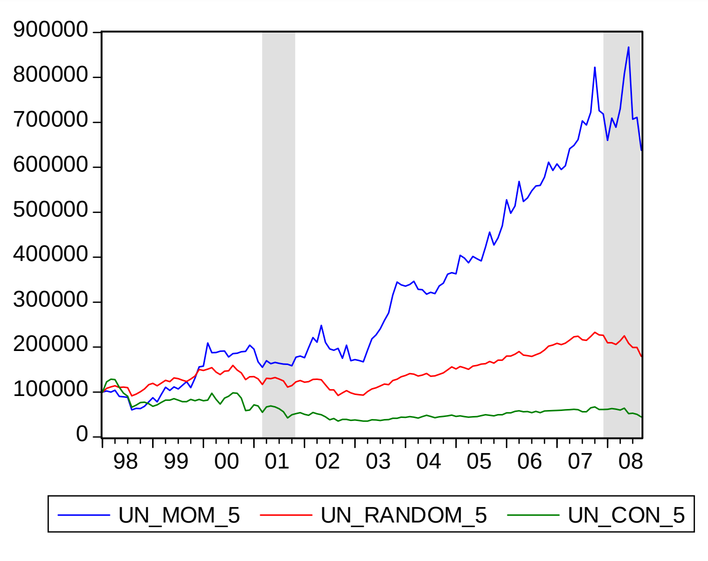


Figure 5.1: Momentum, Contrarian & Naive Strategy with no additional information

higher-rated mutual funds consistently outperform those made up of funds with lower Morningstar star-ratings. In particular, we are interested in examining whether a higher rating reflects superior management skills of the managers of those funds, and therefore a simple investment strategy could be adopted which would systematically result in higher average returns if the Morningstar ranking system is indeed informative about fund performance, this strategy consisting of always selecting the highest-rated funds when creating a portfolio. For this purpose, first we examine the statistical properties of the five funds-of-funds return series, i.e. the returns on the portfolios including respectively five-,

Strategy - Star - n	**Mean**	**St.Dev.**	**Sharpe**	**Prob**
Uncond. Momentum 5	1.45	8.69	0.17	0.06
Momentum 1 - 5	0.83	6.97	0.12	0.18
Momentum 2 - 5	0.73	8.11	0.09	0.31
Momentum 3 - 5	1.04	6.85	0.15	0.09
Momentum 4 - 5	1.16	6.40	0.18	0.04
Momentum 5 - 5	0.96	6.61	0.15	0.10

Table 5.1: Descriptive Characteristics of Portfolio Returns

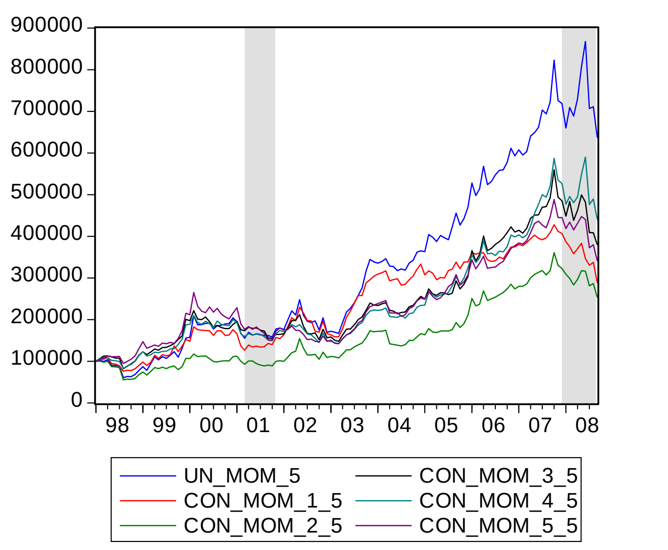


Figure 5.2: Unconditional Momentum Strategy with 5 Funds (Un_Mom_5) Vs Conditional Momentum Strategy with 5 Funds (Con_Mom_'Star'_5)

four-, three-, two- and one-star funds only (STAR5 to STAR1). We show that, provided second-order dependence is taken into account, statistically significant return differentials can indeed be found, the higher Morningstar rating being associated with higher returns. In order to establish whether this is in fact due to superior management skills, we estimate appropriate asset pricing models for risk-adjusted returns. Specifically, we consider both a standard version of the CAPM model for unconditional portfolio performance evaluation, and a conditional CAPM (see Ferson and Schadt, 1996, and Shanken, 1990) in which portfolio risk is a function of observed variables in order to carry out a conditional evaluation as well. The results based on the former specification (when allowing for conditional heteroscedasticity) indicate that only the three highest-rated categories of funds are characterised by some stock selection ability, whilst none of the five categories exhibit market timing ability. Similar results are obtained for the conditional portfolio evaluation, the evidence suggesting that the better performance of the STAR3, STAR4 and STAR5 categories reflects superior stock selection rather than market timing abilities. Overall, the implication for the Morningstar ranking system is that this is most effective in identifying the worst-performing funds (those to which one or two stars are assigned) rather than the best-performing ones: it can be used as a guide to avoid one- and two-star rated funds, but it is not really able to discriminate between three-, four- and five- star funds (although this does not rule out that mutual fund investors are more sensitive to *changes* in the ratings compared with the ratings themselves, as highlighted by Del Guercio and Tkac, 2008).

TABLES

Table 1

Time Series Properties of Monthly Returns of Star-Rated Funds-of-Funds

A. Univariate Characteristics

Fund-of-Funds	mean	s.d.	skewn.	kurt.	J-B*	min.	max.	$\widehat{\rho}_1$	Q(12)*
STAR1	0.31	5.35	-0.90	5.24	0.00	-23.26	12.60	0.053	0.93
STAR2	0.39	4.70	-0.97	4.92	0.00	-19.90	8.96	0.091	0.97
STAR3	0.42	4.32	-0.95	4.47	0.00	-17.01	7.42	0.095	0.99
STAR4	0.45	4.35	-0.87	4.06	0.00	-16.15	8.71	0.094	0.95
STAR5	0.51	4.87	-0.70	4.08	0.00	-16.58	12.98	0.082	0.78

*:p-val

B. Correlation Matrix

	STAR1	STAR2	STAR3	STAR4	STAR5
STAR1	1	0.98	0.95	0.93	0.87
STAR2	0.98	1	0.98	0.97	0.90
STAR3	0.95	0.98	1	0.98	0.92
STAR4	0.93	0.97	0.98	1	0.96
STAR5	0.87	0.90	0.92	0.96	1

C. Testing for Zero Mean in Return-Differentials

(i) OLS: $\Delta R_t^{5i*} = c_i + \nu_{it}$.

$H_0 : c_i = 0$

				Serial Correlation		Non-Linear Dependence		
Return-Differential	$\widehat{c}_i$	$s.e.(\widehat{c}_i)$	t-stat.	$\widehat{\rho}_1(\nu_{it})$	Q(12)*	$\widehat{\rho}_1(\nu_{it}^2)$	Q(12)*	BDS*
STAR5-1	0.19	0.23	0.82	0.051	0.38	0.47	0.00	0.00
STAR5-2	0.11	0.18	0.61	0.069	0.09	0.55	0.00	0.00
STAR5-3	0.09	0.16	0.53	0.029	0.08	0.45	0.00	0.00
STAR5-4	0.05	0.11	0.42	-0.009	0.11	0.52	0.00	0.00

*:p-val

(ii) GARCH(1,1): $\Delta R_t^{5i*} = c_i + \nu_{it}$, $\nu_{it} = h_{it}\varepsilon_{it}$, $h_{it}^2 = d_i + a_i h_{it-1}^2 + b_i \varepsilon_{it-1}^2$

$H_0 : c_i = 0$

				i.i.d. for standard. residuals
Return-Differential	$\widehat{c}_i$	$s.e.(\widehat{c}_i)$	t-stat.	BDS
STAR5-1	0.27	0.09	2.91	0.43
STAR5-2	0.14	0.06	2.13	0.70
STAR5-3	0.11	0.05	2.08	0.12
STAR5-4	0.05	0.03	1.80	0.52

Table 2

Unconditional Evaluation of Star-Rated Funds-of-Funds

A. OLS

Fund-of-Funds	a	$t(a)$	b	$t(b)$	γ	$t(\gamma)$	$\overline{R}^2$
STAR1	0.42	1.63	1.05	18.20	-0.01	-1.84	0.79
STAR2	0.40	2.28	0.98	25.07	-0.01	-1.78	0.87
STAR3	0.37	3.20	0.95	36.47	-0.01	-1.91	0.93
STAR4	0.39	3.31	0.96	35.85	-0.00	-1.58	0.93
STAR5	0.52	2.47	0.99	20.86	-0.01	-1.56	083

			Serial Cor.	Non-Lin. Depend.	
	AIC	SIC	Q(12)	Q(12)	BDS
STAR1	4.653	4.720	0.59	0.00	0.00
STAR2	3.884	3.951	0.12	0.00	0.00
STAR3	3.049	3.116	0.15	0.02	0.04
STAR4	3.107	3.173	0.11	0.00	0.01
STAR5	4.268	4.334	0.13	0.00	0.00

B. Heteroscedasticity-Consistent s.e.'s

Fund-of-Funds	a	$t(a)$	b	$t(b)$	γ	$t(\gamma)$
STAR1	0.42	1.52	1.05	18.94	-0.01	-1.32
STAR2	0.40	2.10	0.98	25.52	-0.01	-1.13
STAR3	0.37	3.02	0.95	37.09	-0.01	-1.36
STAR4	0.39	3.19	0.96	35.03	-0.00	-1.31
STAR5	0.52	2.55	0.99	18.14	-0.01	-1.47

C. GARCH(1,1)

Fund-of-Funds	a	$t(a)$	b	$t(b)$	γ	$t(\gamma)$	AIC	SIC
STAR1	-0.05	-0.36	1.05	24.39	0.00	0.07	4185	4.319
STAR2	0.15	1.22	0.99	32.82	0.00	0.25	3.547	3.681
STAR3	0.29	2.47	0.95	39.22	-0.00	-0.62	3.044	3.177
STAR4	0.31	2.88	0.96	38.21	-0.00	-0.86	3.068	3.202
STAR5	0.31	2.66	0.99	38.76	-0.00	-0.81	3.625	3.759

Table 3

Statistical Properties of the State Variables

A. Unit Root Tests

Variable	ADF	PP	DF-GLS	MZ_a	MZ_t
z_{1t}	-1.05	-1.25	-1.22	-4.71	-1.42
z_{2t}	-1.54	-2.06	-1.16	-2.66	-1.14
z_{3t}	-0.91	-0.65	-0.63	-3.59	-0.79
z_{4t}	-0.89	-0.60	-0.88	-2.32	-0.87
z_{5t}	-1.97	-1.75	-1.96	-6.35	-1.76
z_{6t}	0.76	1.40	1.47	2.89	1.50
z_{7t}	-3.67	-3.38	-2.21	-12.15	-2.12
5% c.v.'s	-2.87	-2.87	-1.94	-8.10	-1.98

B. Testing for Cointegration Among z_{1t}, z_{2t}, ..., z_{6t}

	$l = 1$		$l = 6$		5% c.v.'s	
	$\lambda - \max$	TR	$\lambda - \max$	TR	$\lambda - \max$	TR
$k = 0$	38.14	89.49	43.11	116.27	40.07	95.75
$k = 1$	26.81	51.35	32.49	73.15	33.87	69.81
$k = 2$	10.87	24.53	17.23	40.66	27.58	47.85
$k = 3$	9.21	13.65	14.14	23.42	21.13	29.79
$k = 4$	3.35	4.44	5.37	9.28	14.26	15.49
$k = 5$	1.08	1.08	3.91	3.91	3.84	3.84

Notes:

1) z_{1t} =one-month treasury bill yield, z_{2t} =constant-maturity 10-year Treasury bond yield minus 3-month Treasury bill, z_{3t} = Moody's BAA-rated corporate bond yield minus AAA-rated corporate bond yield, z_{4t} =the exchange rate of the dollar, z_{5t} =the consumer confidence index, z_{6t} =the price of oil, z_{7t} =the CBOE's VIX volatility index.

2) ADF and PP refer to the standard Augmented Dickey-Fuller (1979) and Phillips-Perron (1988) tests respectively for the null hypothesis of a unit root. The lag-length and the bandwidth parameter in ADF and PP respectively, were selected by the Schwarz information criterion and the Newey and West (1994) procedure respectively. DF-GLS refers to the unit root test proposed by Elliot, Rothenberg and Stock (1996) based on GLS detrended series. MZ_a and MZ_t are two of the four tests proposed by Ng and Perron (2001).

3) l denotes the lag length of the unrestricted Vector Autoregressive Model (VAR) based on which the Johansen (1991) maximum eigenvalue (λ – max) and trace (TR) statistics were calculated.

Table 4

Conditional Evaluation of Star-Rated Funds-of-Funds

(GARCH(1,1) Error Specification)

A: Cointegration Rank among z_{1t}, z_{2t}, ..., z_{6t} is Equal to Zero.

Δz_{1t}, Δz_{2t}, ..., Δz_{6t} and z_{7t} are Employed as State Variables.

Fund-of-Funds	a	$t(a)$	γ	$t(\gamma)$	Significant State Variables	AIC	SIC
STAR1	-0.01	-0.02	-0.008	-1.12	Δz_{4t}, Δz_{6t}	4.175	4.353
STAR2	0.26	2.32	-0.008	-1.92	Δz_{1t}, Δz_{2t}, Δz_{4t}, Δz_{6t}	3.505	3.728
STAR3	0.33	3.42	-0.007	-2.71	Δz_{4t}, Δz_{6t}	3.004	3.183
STAR4	0.26	2.58	-0.005	-1.98	Δz_{6t}, z_{7t}	3.016	3.194
STAR5	0.34	2.73	-0.010	-3.36	Δz_{3t}, Δz_{6t}, z_{7t}	3.505	3.706

B: Cointegration Rank among z_{1t}, z_{2t}, ..., z_{6t} is Equal to One.

Cointegration Relation, u_{1t}, Together with z_{7t} are Employed as State Variables.

Fund-of-Funds	a	$t(a)$	γ	$t(\gamma)$	AIC	SIC
STAR1	-0.07	-0.50	-0.001	-0.23	4.179	4.357
STAR2	0.14	1.24	0.001	0.04	3.545	3.723
STAR3	0.23	2.23	-0.003	-0.85	3.043	3.222
STAR4	0.28	2.66	-0.004	-1.17	3.024	3.202
STAR5	0.30	2.47	-0.005	-1.57	3.531	3.709

C: Cointegration Rank among z_{1t}, z_{2t}, ..., z_{6t} is Equal to Two.

Cointegration Relations, u_{1t}, and u_{2t} Together with z_{7t} are Employed as State Variables.

Fund-of-Funds	a	$t(a)$	γ	$t(\gamma)$	AIC	SIC
STAR1	-0.09	-0.61	-0.002	-0.326	4.193	4.394
STAR2	0.14	1.16	0.000	0.051	3.562	3.762
STAR3	0.23	2.07	-0.003	-0.961	3.056	3.256
STAR4	0.29	2.68	-0.003	-0.949	3.033	3.233
STAR5	0.36	3.10	-0.010	-3.01	3.544	3.745

5.6 Appendix

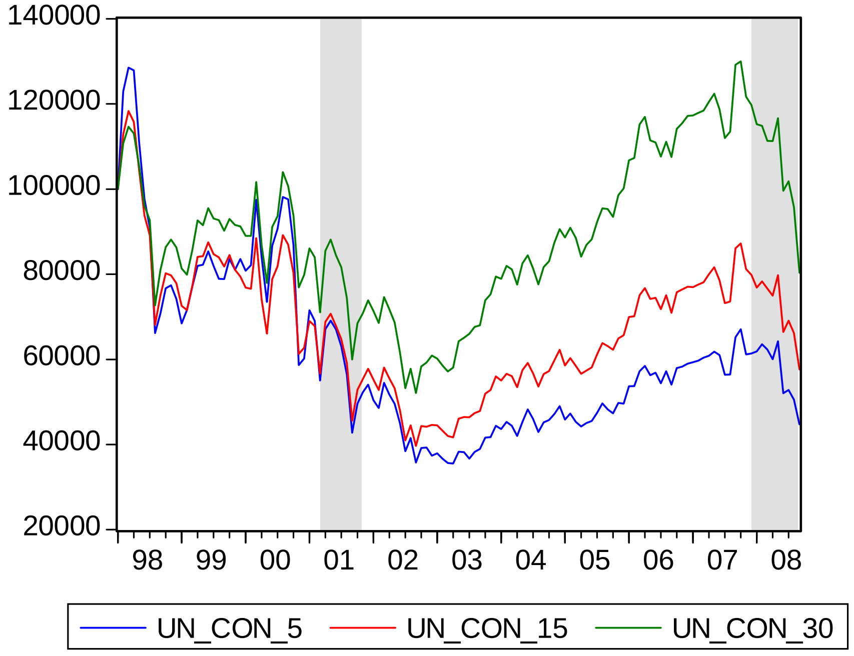


Figure 5.3: Unconditional Momentum Strategy $n = \{5, 15, 30\}$

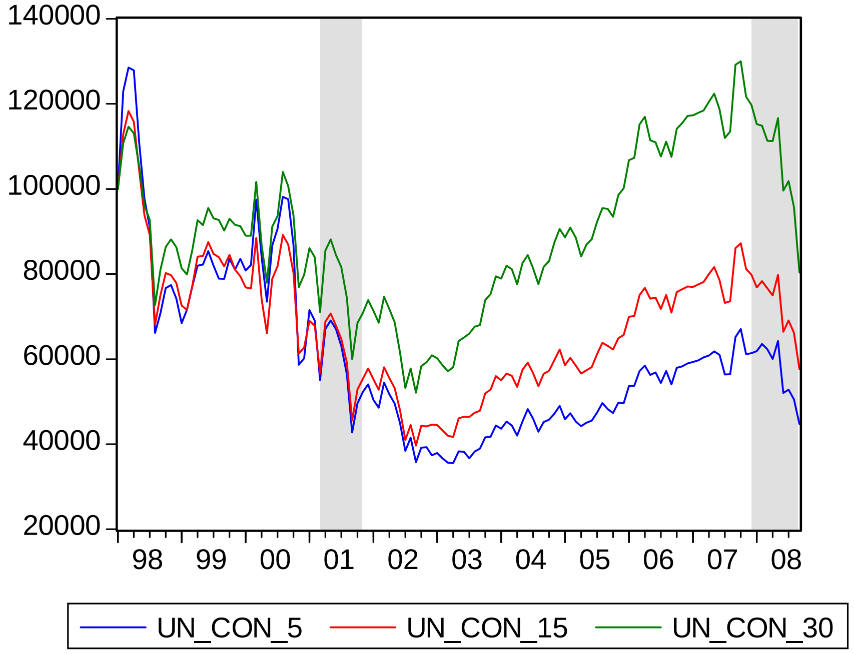


Figure 5.4: Unconditional Contrarian Strategy $n = \{5, 15, 30\}$

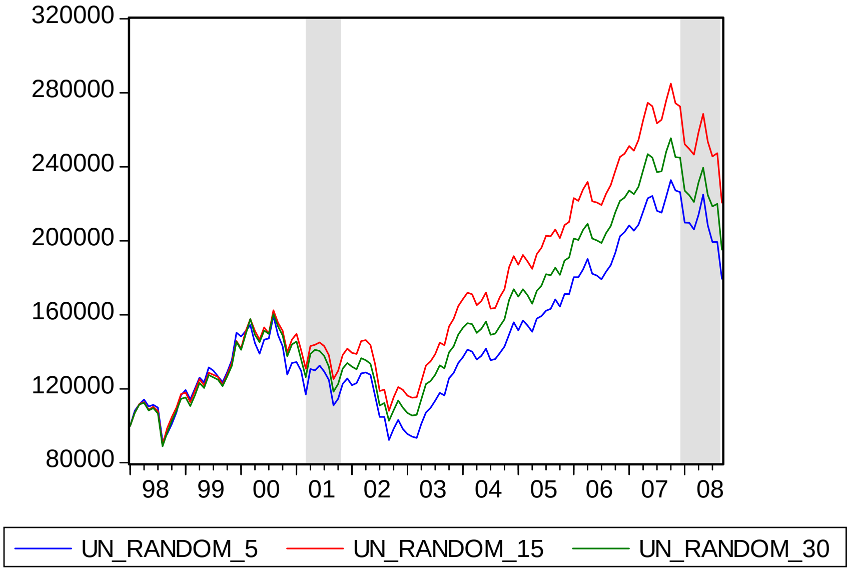


Figure 5.5: Unconditional Naive Strategy

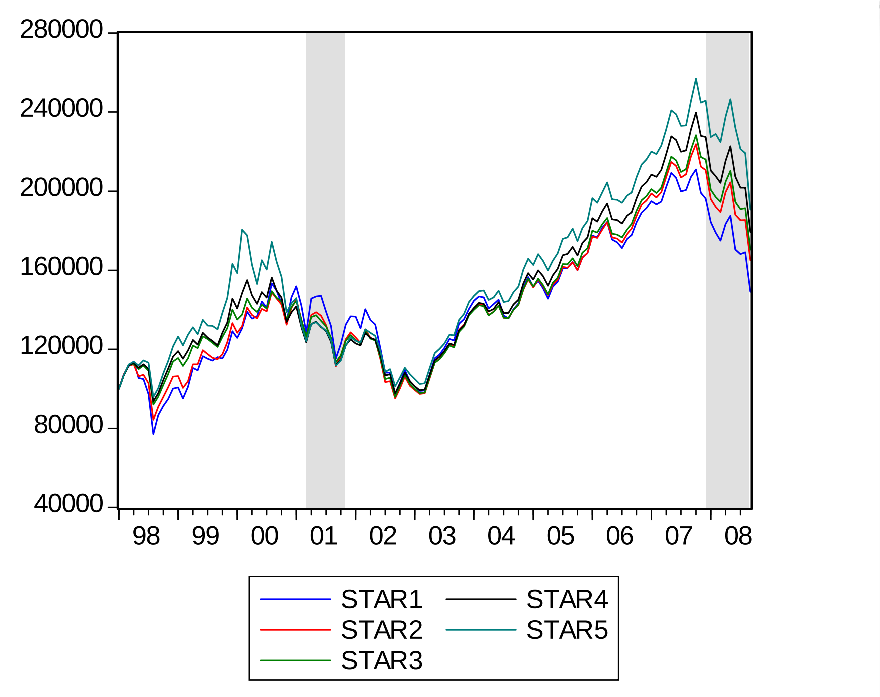


Figure 5.6: Morningstar Star Rates Portfolios

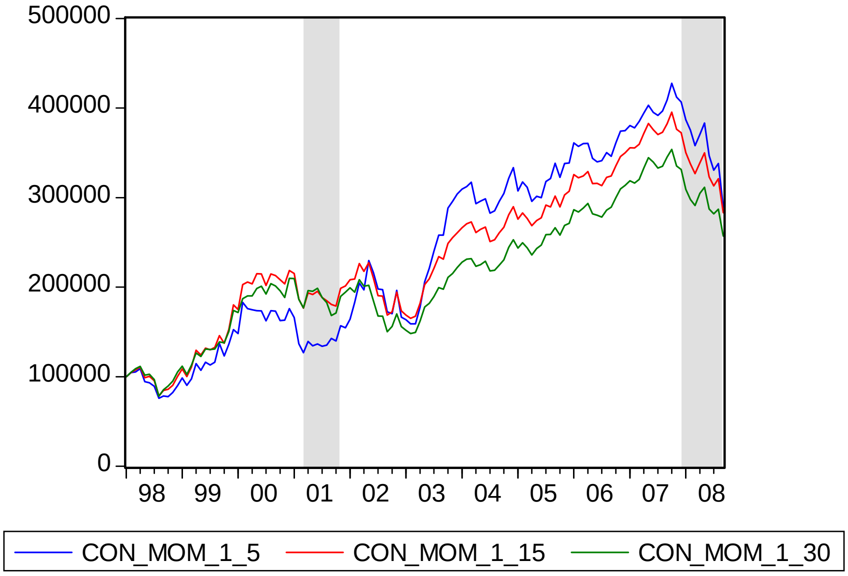


Figure 5.7: Momentum & Morningstar Rating - 1 Star Strategy $n = \{5, 15, 30\}$

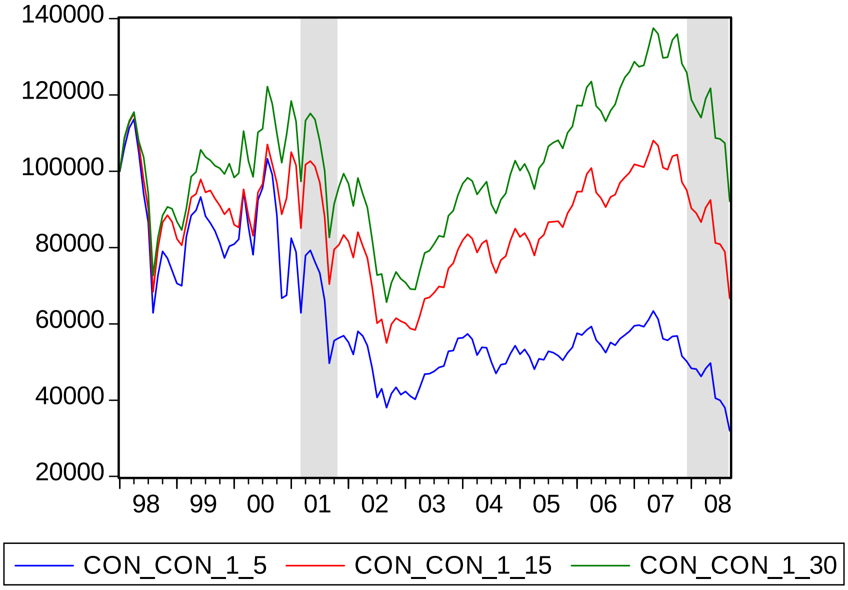


Figure 5.8: Contrarian & Morningstar Rating - 1 Star Strategy $n = \{5, 15, 30\}$

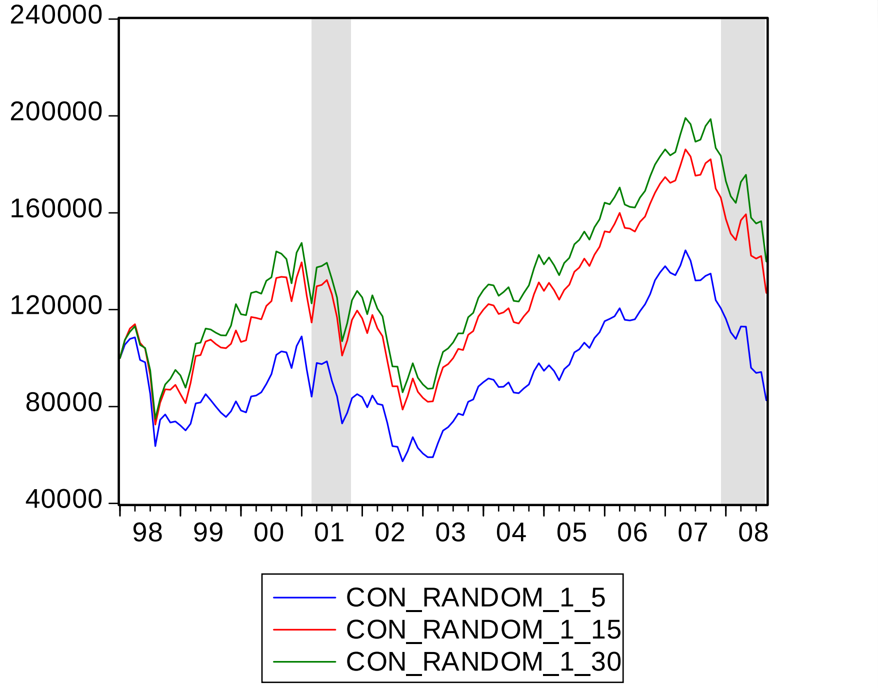


Figure 5.9: Naive & Morningstar Rating - 1 Star Strategy $n = \{5, 15, 30\}$

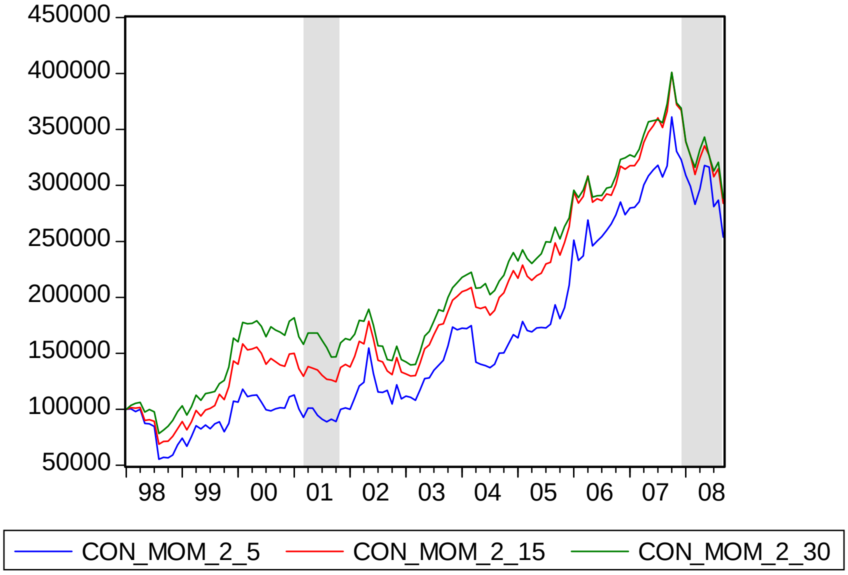


Figure 5.10: Momentum & Morningstar Rating - 2 Star Strategy $n = \{5, 15, 30\}$

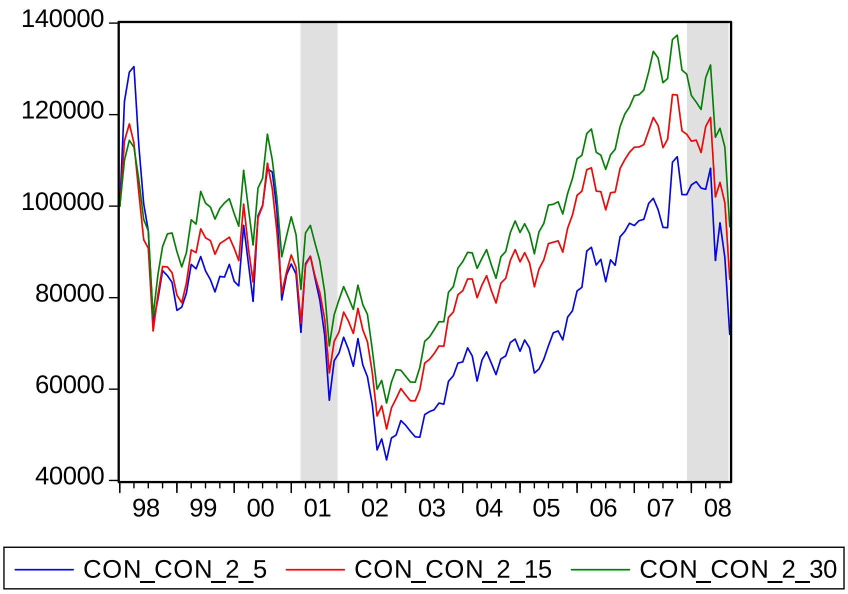


Figure 5.11: Contrarian & Morningstar Rating - 2 Star Strategy $n = \{5, 15, 30\}$

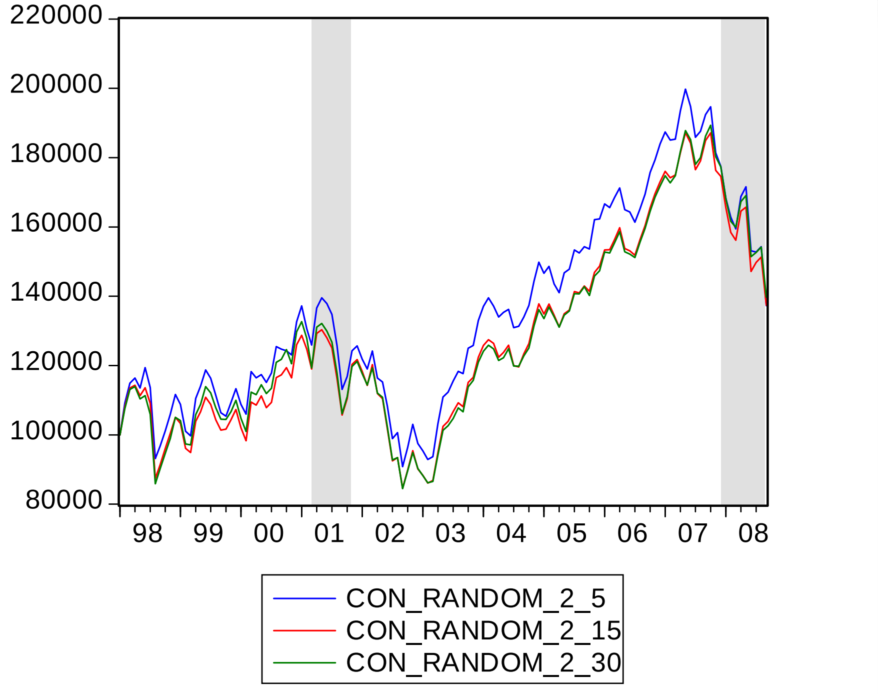


Figure 5.12: Naive & Morningstar Rating - 2 Star Strategy $n = \{5, 15, 30\}$

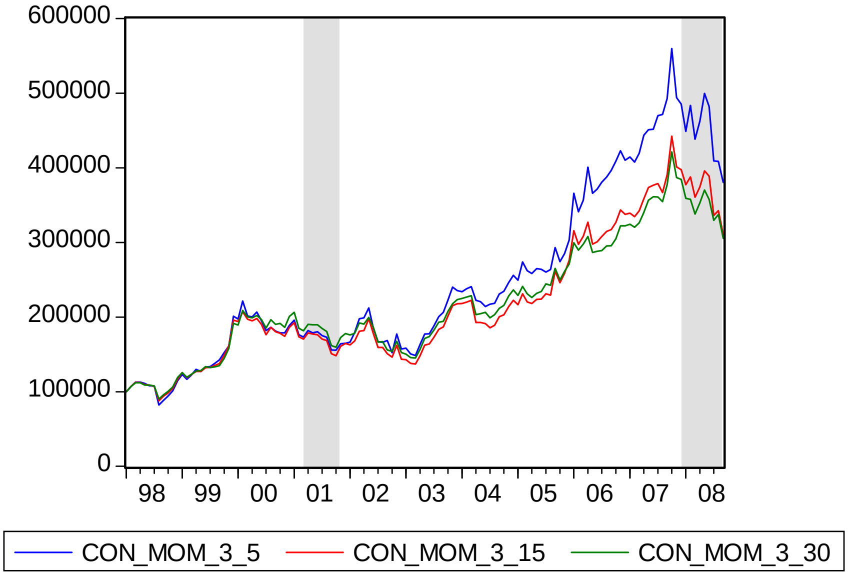


Figure 5.13: Momentum & Morningstar Rating - 3 Star Strategy $n = \{5, 15, 30\}$

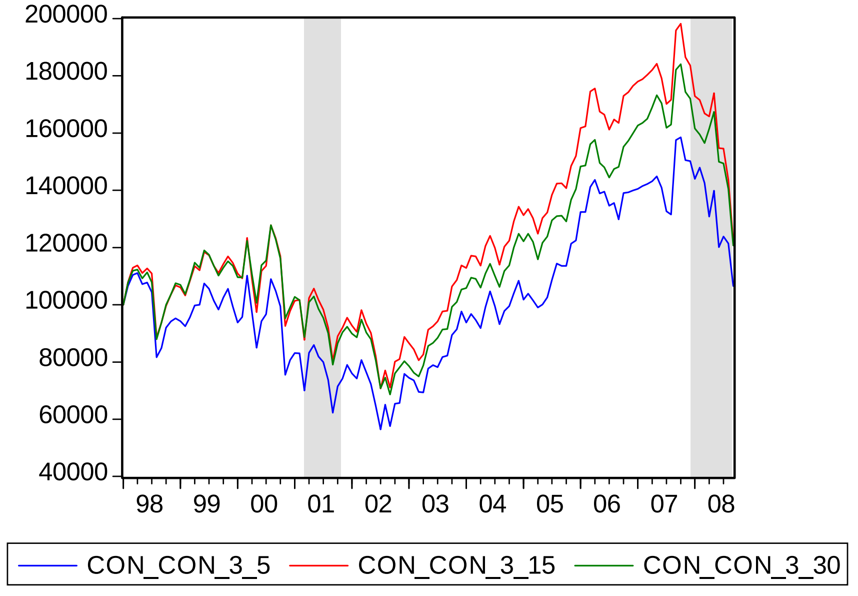


Figure 5.14: Contrarian & Morningstar Rating - 3 Star Strategy $n = \{5, 15, 30\}$

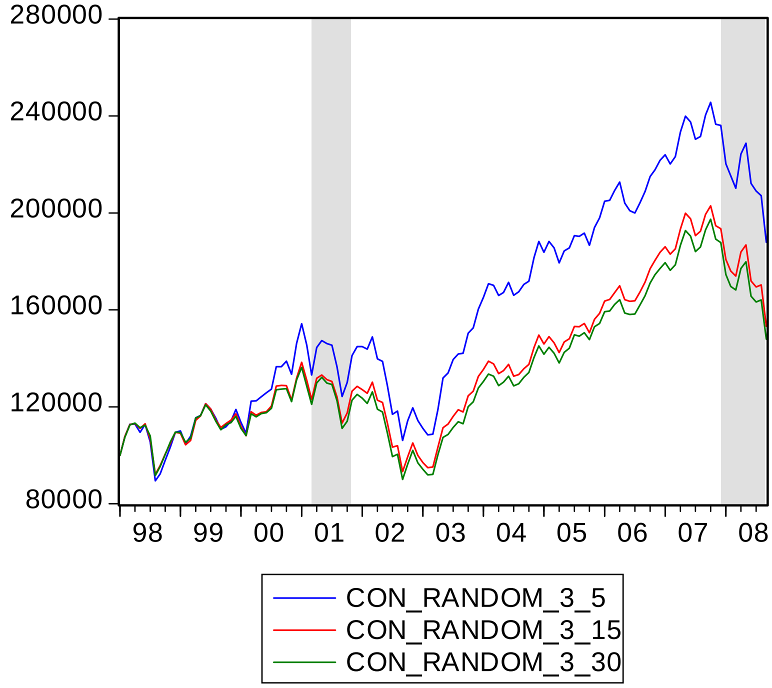


Figure 5.15: Naive & Morningstar Rating - 3 Star Strategy $n = \{5, 15, 30\}$

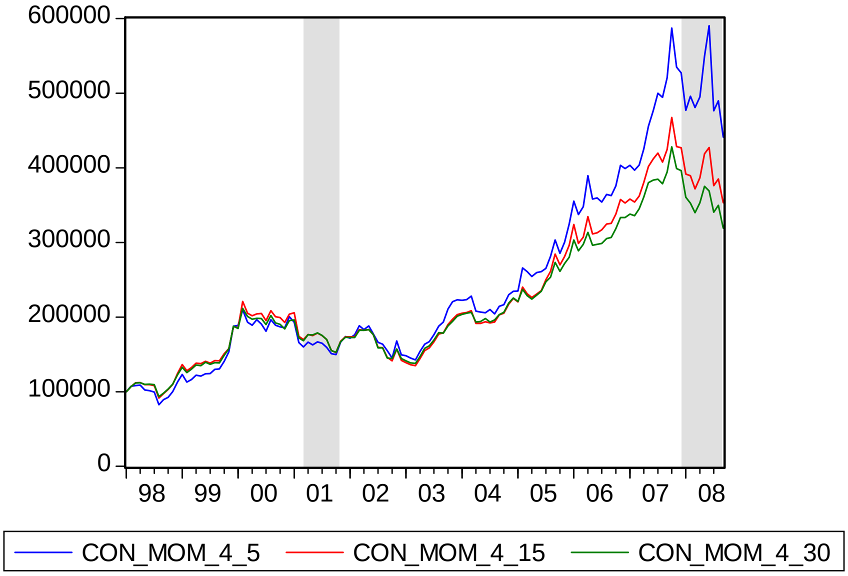


Figure 5.16: Momentum & Morningstar Rating - 4 Star Strategy $n = \{5, 15, 30\}$

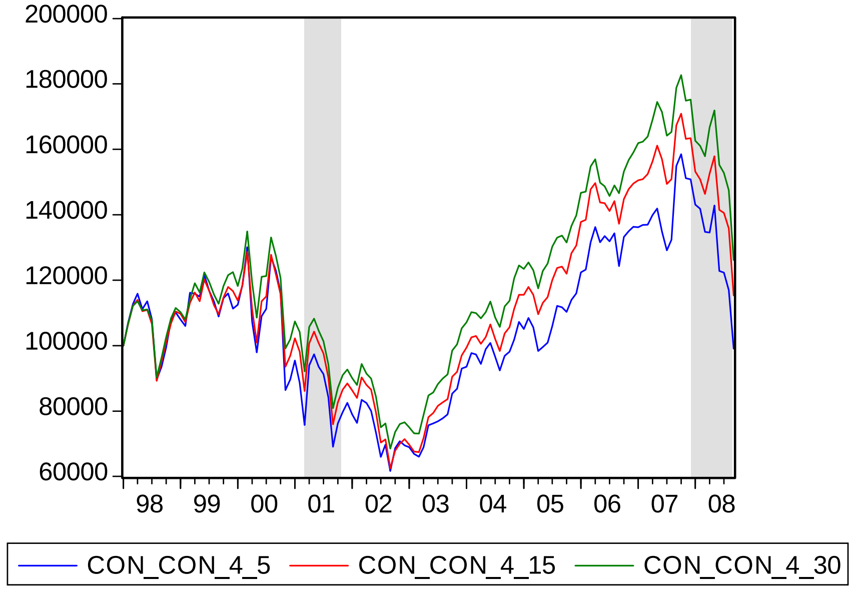


Figure 5.17: Contrarian & Morningstar Rating - 4 Star Strategy $n = \{5, 15, 30\}$

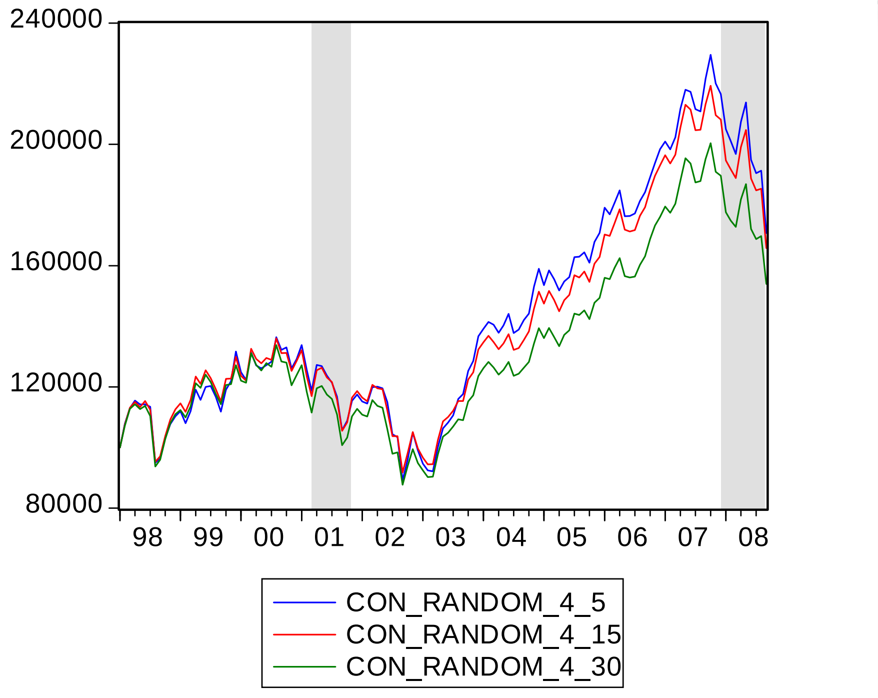


Figure 5.18: Naive & Morningstar Rating - 4 Star Strategy $n = \{5, 15, 30\}$

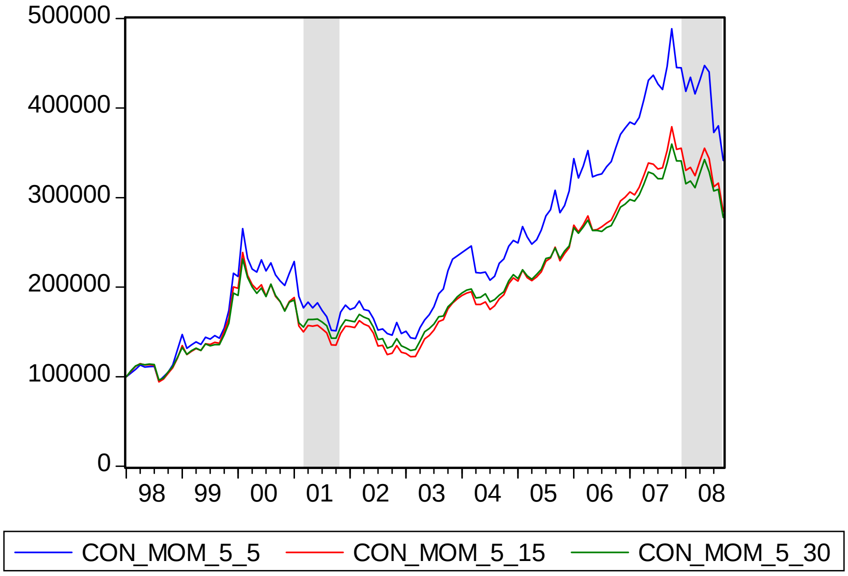


Figure 5.19: Momentum & Morningstar Rating - 5 Star Strategy $n = \{5, 15, 30\}$

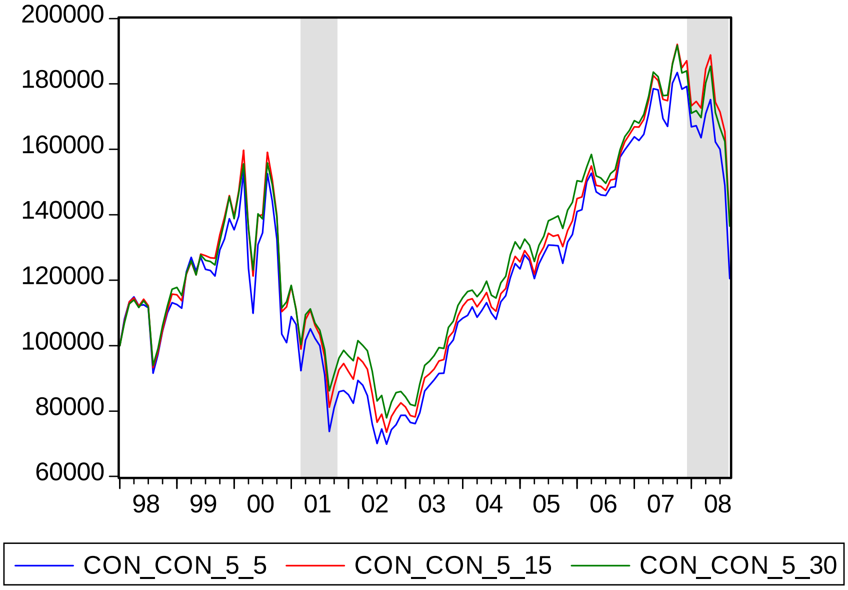


Figure 5.20: Contrarian & Morningstar Rating - 5 Star Strategy $n = \{5, 15, 30\}$

Table 5.2: Descriptive Statistics: Portfolio Returns of Unconditional Strategies

Strategy - n	**Mean**	**St.Dev.**	**Sharpe**	**Prob**
Contrarian 5	-0.63	9.01	-0.07	0.43
Contrarian 15	-0.43	7.88	-0.05	0.54
Contrarian 30	-0.17	7.15	-0.02	0.79
Momentum 5	1.45	8.69	0.17	0.06
Momentum 15	0.97	7.45	0.13	0.14
Momentum 30	0.85	6.68	0.13	0.15
Naive 5	0.46	4.87	0.09	0.29
Naive 15	0.62	4.63	0.13	0.13
Naive 30	0.52	4.60	0.11	0.20

Table 5.3: Descriptive Statistics: Portfolio Returns of a Naive Investor Conditional on the Morningstar Rating System

Strategy - Star - n	**Mean**	**St.Dev.**	**Sharpe**	**Prob**
Naive 1 -5	-0.15	6.24	-0.02	0.79
Naive 1 -15	0.19	5.61	0.03	0.71
Naive 1 -30	0.26	5.43	0.05	0.59
Naive 2 -5	0.25	4.74	0.05	0.56
Naive 2 -15	0.25	4.74	0.05	0.55
Naive 2 -30	0.26	4.66	0.06	0.53
Naive 3 -5	0.49	4.57	0.11	0.22
Naive 3 -15	0.33	4.21	0.08	0.37
Naive 3 -30	0.31	4.22	0.07	0.41
Naive 4 -5	0.42	4.50	0.09	0.30
Naive 4 -15	0.39	4.26	0.09	0.30
Naive 4 -30	0.34	4.14	0.08	0.36
Naive 5 -5	0.50	4.60	0.11	0.22
Naive 5 -15	0.49	4.26	0.11	0.20
Naive 5 -30	0.47	4.23	0.11	0.21

Table 5.4: Descriptive Statistics: Portfolio Returns of a Contrarian Strategy, Conditional on the Morningstar Rating System

Strategy - Star - n	**Mean**	**St.Dev.**	**Sharpe**	**Prob**
Contrarian 1 -5	-0.89	8.38	-0.11	0.23
Contrarian 1 -15	-0.32	6.67	-0.05	0.59
Contrarian 1 -30	-0.06	6.05	-0.01	0.90
Contrarian 2 -5	-0.26	7.91	-0.03	0.71
Contrarian 2 -15	-0.14	6.58	-0.02	0.82
Contrarian 2 -30	-0.04	5.98	-0.01	0.95
Contrarian 3 -5	0.05	7.39	0.01	0.94
Contrarian 3 -15	0.16	6.48	0.02	0.78
Contrarian 3 -30	0.15	5.81	0.03	0.78
Contrarian 4 -5	-0.01	7.08	0.00	0.99
Contrarian 4 -15	0.11	6.05	0.02	0.84
Contrarian 4 -30	0.18	5.62	0.03	0.72
Contrarian 5 -5	0.15	6.55	0.02	0.80
Contrarian 5 -15	0.25	5.94	0.04	0.63
Contrarian 5 -30	0.24	5.47	0.04	0.62

Table 5.5: Descriptive Statistics: Portfolio Returns of a Momentum Strategy, Conditional on the Morningstar Rating System

Strategy - Star - n	**Mean**	**St.Dev.**	**Sharpe**	**Prob**
Momentum 1 -5	0.83	6.97	0.12	0.18
Momentum 1 -15	0.81	5.85	0.14	0.12
Momentum 1 -30	0.74	5.44	0.14	0.13
Momentum 2 -5	0.73	8.11	0.09	0.31
Momentum 2 -15	0.82	5.99	0.14	0.13
Momentum 2 -30	0.83	5.28	0.16	0.08
Momentum 3 -5	1.04	6.85	0.15	0.09
Momentum 3 -15	0.89	6.06	0.15	0.10
Momentum 3 -30	0.87	5.42	0.16	0.07
Momentum 4 -5	1.16	6.40	0.18	0.04
Momentum 4 -15	0.99	5.73	0.17	0.05
Momentum 4 -30	0.91	5.12	0.18	0.05
Momentum 5 -5	0.96	6.61	0.15	0.10
Momentum 5 -15	0.82	5.72	0.14	0.11
Momentum 5 -30	0.80	5.29	0.15	0.09

Bibliography

[1] Akharif, A., Hallin, M. (2003). Efficient detection of random coefficients in autoregressive models. Annals of Statistics: 31, 675-704.

[2] Bera, A.K., Higgins, M.L. and S. Lee (1992). Interaction between Autocorrelation and Conditional Heteroscedasticity: A Random-Coefficient Approach. Journal of Business & Economic Statistics, 10(2): 133-42.

[3] Belsley D., Kuti E.,1973, Time-Varying Parameter Structures: An Overview, Annals of Economic and Social Measurament,2, 375-379

[4] Blake, C.R and M.R. Morey (2000), "Morningstar Ratings and Mutual Fund Performance", Journal of Financial and Quantitative Analysis, 35, 3, 451-483.

[5] Blattberg RC, Gonedes NJ. 1974. A comparison of the stable and student distributions as statistical models for stock price. Journal of Business 47: 244-280.

[6] Bollerslev T., 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31,. 307-327

[7] Blume, M.(1998), "An anatomy of Morningstar ratings", Financial Analysts journal, March/April, 19-27.

[8] Bradley RC. 1988. A central limit theorem for stationary -mixing sequences with infinite variance. The Annals of Probability 16: 313-332.

[9] Brown, S.J. and W. N. Goetzmann (1995), "Performance persistence", Journal of Finance, 50, 679-698.

[10] Campbell, J.Y., Chan, Y.L., Viceira L.M., 2001. A Multivariate Model of Strategic Asset Allocation. NBER Working Papers 8566, National Bureau of Economic Research.

[11] Carrasco M, Chen X. 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18: 17-39.

[12] Clark PK. 1973. A subordinate stochastic process model with finite variance for speculative prices. Econometrica 41: 135-155.

[13] Conlisk, J. (1974). Stability in a Random Coefficient Model. International Economic Review: 15, 529-533.

[14] Conlisk, J., (1976). A Further Note on Stability in a Random Coefficient Model. International Economic Review: 17, 759-764.

[15] Damato, K. (1996), "Morningstar edges toward one-year ratings", Wall Street Journal, April 5th, 1996.

[16] Del Guercio, D. and P.A. Tkac (2008), "The effect of Morningstar ratings on mutual fund flow", Journal of Financial and Quantitative Analysis, 43, 4, 907-936.

[17] Diebold FX, Lopez JA. 1995. Modeling Volatility Dynamics. In Macroeconometrics: developments, tensions, and prospects, edt. Kevin D. Hoover Kluwer Massachusetts.

[18] Diebold FX. 1988 Empirical Modeling of Exchange Rate Dynamics. (Lecture Notes in Economics and Mathematical systems, No. 303). Springer-Verlag. New York, Heidelberg and Tokyo

[19] Diebold FX.1986. Modeling the Persistence of Conditional Variances: A comment. Econometric Reviews 5: 51-56.

[20] Drost F, Nijman T. 1993. Temporal aggregation of GARCH processes. Econometrica 61: 909-27.

[21] Elton, E., Gruber, M., Das, S. and M. Hlavka (1992), "Efficiency with costly information: a reinterpretation of evidence for managed portfolio", Review of Financial studies, 6, 1-22.

[22] Engle, Robert F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, 50, 987-1007

[23] Feigin, P.D. and R.L. Tweedie, (1985). Random Coefficient Autoregressive Processes: A Markov Chain Analysis of Stationarity and Finiteness of Moments. Journal of Time Series Analysis: 6, 1-14.

[24] Ferson, W.E. and R.W. Schadt (1996), "Measuring fund strategy and performance in changing economic conditions", Journal of Finance, 51, 2, 425-461.

[25] Fong WM, See KH. 2001. Modeling the Conditional Volatility of Commodity Index Futures as a Regime Switching Process. Journal of Applied Econometrics 16: 133-163.

[26] Fornari F, Mele A. 1997. Sign- and Volatility-Switching ARCH Models: Theory and Applications to International Stock Markets. Journal of Applied Econometrics 12: 49-65.

[27] Francq C, Zakoian JM. 2006. Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. Econometric Theory 22: 815-834.

[28] Goetzmann, W.N. and R.G. Ibbotson (1994), "Do winners repeat?", Journal of Portfolio Management, Winter, 9-18.

[29] Granger, C.W.J. and A. P. Andersen (1978). An Introduction to Bilinear Time Series Models. Vandenhoeck and Ruprecht: Gottingen.

[30] Hendricks, D., Patel, J. and R. Zeckhauser (1993), "Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988", Journal of Finance, 48, 93-130.

[31] Hwang, S.Y. and I.V. Basawa (1993). Asymptotic Optimal Inference for a Class of Nonlinear Time Series Models. Stochastic Processes and Their Applications: 46, 91-113.

[32] Hwang, S.Y. and I.V. Basawa (1998). Parameter Estimation for Generalized Random Coefficient Autoregressive Processes. Journal of Statistical Planning and Inference: 68, 323-337.

[33] Ibragimov IA, Linnik YV. 1971. Independent and Stationary Sequences of Random Variables. Wolters Noordhof, Groningen.

[34] Ibragimov IA. 1962. Some limit theorems for stationary processes. Theory of Probability and Its Applications 7: 349-382.

[35] Jensen, M.C. (1968), "The performance of mutual funds in the period 1945-1964", Journal of Finance, 23, 389-416.

[36] Jensen, M.C. (1972), "Optimal utilization of market forecasts and the evaluation of investment performance", in G.P. Szego and K. Shell, eds., Mathematical Methods in Investment and Finance, North-Holland, Amsterdam.

[37] Kapetanios, G. (2002). A Note on Covariance Stationarity Conditions for Dynamic Random Coefficient Models, University of London, Queen Mary College, Department of Economics, WP no 475.

[38] Kendall MG. 1953. The analysis of economic series-Part I: Prices. Journal of the Royal Statistical. Society, Series A (General) 116: 11-25.

[39] Khorana, A. and E. Nelling (1998), "The determinants and predictive ability of mutual fund ratings", The Journal of Investing, November, 61-66.

[40] Koul, H.L and A. Schick (1996). Adaptive Estimation in a Random Coefficient Autoregressive Model. The Annals of Statistics: 24, 1025-1052.

[41] Kourogenis N, Pittis N. 2008. Testing for a unit root under errors with just barely infinite variance. Journal of Time Series Analysis 29: 1066-1087.

[42] Kourogenis N, Pittis N. 2010. Mixing conditions, central limit theorems and invariance principles: A survey of the literature with some new results on heteroscedastic sequences. Econometric Reviews.(forthcoming).

[43] Kourogenis, N., Pittis, N., 2008. Estimation and Inference in Autoregressive Models with Trending Innovation Variance. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1333125 , unpublished manuscript.

[44] Lamoureux CG, Lastrapes WD. 1990. Persistence in Variance, Structural Change, and the GARCH Model. Journal of Business & Economic Statistics 8: 225-234.

[45] Lamoureux, C. G., Lastrapes, W. D., 1990. Persistence in Variance, Structural Change, and the GARCH Model. Journal of Business & Economic Statistics 8, 225-234.

[46] Lee, S. & B. Hansen (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, Econometric Theory 10, 29-53.

[47] Lee, S. (1998). Coefficient Constancy Test in a Random Coefficient Autoregressive Model. Journal of Statistical Planning and Inference: 74, 93-101.

[48] Lettau, M. and S. Ludvingston (2001), "Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time varying", Journal of Political Economy, 109, 1238-1287.

[49] Levy P. 1935. Proprietes asymptotique des sommes de variables aleatoires independantes ou enchainees. Journal de Math matiques Pures et appliqu es 14: 347-402.

[50] Liu J-C. 2009. Integrated Markov-switching GARCH processes. Econometric Theory 25: 1277-1288.

[51] Mandelbrot B. 1963. The variation of certain speculative prices. Journal of Business 36: 394-419.

[52] Merton, R.C. and R.D. Henriksson (1981), "On market timing and investment performance II: statistical procedures for evaluating forecasting skills", Journal of Business, 54, 513-534.

[53] Mey n, S. P. and Guo, L. (1993). Geometric ergodicity of a bilinear time series model. Journal of Time Series Analysis: 14, 93-108.

[54] Nelson D. 1990. Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6: 318--334.

[55] Nicholls, D.F. and B.G. Quinn (1980). The Estimation of Random Coefficient Autoregressive Models. I. Journal of Time Series Analysis: 1, 37-46.

[56] Nicholls, D.F. and B.G. Quinn (1981). The Estimation of Random Coefficient Autoregressive Models. II. Journal of Time Series Analysis: 2, 185-203.

[57] Nicholls, D.F. and B.G. Quinn (1982). Random Coefficient Autoregressive Models: An Introduction. Lecture Notes in Statistics: 11. Springer, New York.

[58] Pagan A. (1980). Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients, Journal of Econometrics: 13, 341-363.

[59] Peligrad M. 1990. On Ibragimov-Iosifescu conjecture for -mixing sequences. Stochastic processes and their applications 35: 293-308.

[60] Phillips, P.C.B. (1991), "Optimal inference in cointegrated systems", Econoemtrica, 59, 283-306.

[61] Pourahmadi M., 1986, On Stationarity of the Solution of a Doubly Stochastic Model, Journal of Time Series Analysis, 2, 123-131.

[62] Quinn, B.G. and D.F. Nicholls (1981). The Stability of Random Coefficient Autoregressive Models. International Economic Review: 22, 741-744.

[63] Ramanathan, T.V. and M.B. Rajarshi (1994). Rank Tests for Testing the Randomness of Autoregressive Coefficients. Statistics and Probability Letters: 21, 115-120.

[64] Robinson, P.M. (1978). Statistical Inference for a Random Coefficient Autoregressive Model. Scandinavian Journal of Statistics: 5, 163-168.

[65] S&P GSCI® Index Methodology, October 2009.

[66] Schick, A. (1996). $\sqrt{n}$-Consistent Estimation in a Random Coefficient Autoregressive Model. Australian Journal of Statistics: 38, 155-160.

[67] Shanken, J. (1990), "Intertemporal asset pricing: an empirical investigation", Journal of Econometrics, 45, 99-120.

[68] Sharpe, W.F. (1998), "Morningstar performance measures", Financial Analysts Journal, July/August, 21-33.

[69] Starica, C., Granger, C., 2005. Nonstationarities in Stock Returns, Review of Economics and Statistics 87, 503-522.

[70] Tjostheim D. (1986). Some Doubly Stochastic Time Series Models, Journal of Time Series Analysis: 7, 51-72.

[71] Treynor, J. and K. Mazui (1966), "Can mutual funds outguess the market?", Harvard Business Review, 44, 131-136.

[72] Tsay, R.S. (1987). Conditional Heteroskedastic Time Series Models, Journal of the American Statistical Association, 82: 590-604.

[73] Weiss, A.A. (1985). The Stability of the AR(1) Process with an AR(1) Coefficient. Journal of Time Series Analysis, 6: 181-186.

[74] White, H. and I. Domowitz (1984). Nonlinear Regression with Dependent Observations. Econometrica, 52: 143-162.