Studies in nonlinear and long memory time series econometrics

By

Rehim Kılıç

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2002

ABSTRACT

Studies in nonlinear and long memory time series econometrics

By

Rehim Kılıç

This dissertation explores long memory and nonlinear dynamics in foreign exchange, commodity and stock markets. The first two chapters of this dissertation explore nonlinearity and long memory in econometrics. In particular, chapter one provides a concise overview of Smooth Transition Autoregressive (STAR) models. The discussion is cast in terms of specification procedures for smooth transition models. This chapter provides simulation evidence on the power and size properties of nonlinearity tests designed in the literature against STAR type of nonlinear behavior in a univariate time series. The chapter also studies the small sample properties of nonlinear least squares method in estimating STAR models. Long memory Autoregressive Fractionally Integrated Moving Average (ARFIMA) models for the conditional mean of a process, Generalized Autoregressive Heteroscedastic (GARCH) and Fractionally Integrated GARCH models for the conditional volatility of a process are discussed in terms of specification, estimation and inference in chapter two.

Chapter three of the dissertation investigates a well known puzzle in international finance literature. The purchasing power parity puzzle relates to the slow adjustment of real exchange rates. We investigate the transactions cost-nonlinearity explanation

of the puzzle by utilizing STAR models. The findings in the chapter point out the difficulty in explaining the puzzle by by the transactions cost theory alone. The estimated models and further analysis reveal the extreme persistence in real exchange rates over the floating period.

The fourth chapter of this dissertation investigates long memory dynamics in commodity markets. Both cash and future prices of several commodities, (coffee, corn, gold, silver, soybean and unleaded gasoline) are analyzed. The findings indicate that commodity cash and future prices are approximately martingale with long term dependence in the higher moments. The volatility proxies, for example, squared returns, absolute returns, and intraday range are found to exhibit long memory component. The finding of the long memory has important implications for optimal hedge ratios.

Chapter five of the dissertation analyzes the long memory dynamics in an emerging capital market, the Istanbul Stock Exchange (ISE) National 100 daily and weekly dollar index returns and its absolute and squared returns. Both parametric FIGARCH models and nonparametric methods are employed. Results indicate the presence of long memory dynamics in the conditional variance which can be modelled adequately by a $FIGARCH$ model.

The last chapter revisits the persistence and nonlinearity of deviations from PPP. It develops new unit root test that is specifically designed to test random walk without drift and random walk with drift against stationary exponential smooth transition autoregressive models. The asymptotic distributions of the tests are derived and shown to be nonstandard. The power and size of the tests in finite samples studied by simulations. The fitted exponential STAR models and further analysis reveal the nonlinear nature of real exchange rates as well as the persistence of the deviations.

For Ada, Gülen and in loving memory of my mother, and my brothers.

ACKNOWLEDGMENTS

"We act as though comfort and luxury were the chief requirements of life, while all that we really need is something to be enthusiastic about."

Albert Einstein

The completion of this dissertation would not have been possible without the input, advise, assistance and encouragement of many individuals throughout the entire process. I am happy to have the opportunity to express my gratitude to those individuals, although I realize that it may be impossible to express, in these pages, the depth of my appreciation to the many who are deserving.

When I started my PhD research, I wasn't quite sure whether this was something to be enthusiastic about. When I look back, I can surely say that it was. A decisive factor in bringing about this change of mind has been the opportunity to know the distinguished economists, Richard T Baillie, Robert de Jong, Peter Schmidt, Rowena Pecchenino and Jeffrey Wooldridge. Having the opportunity to know them personally and working with them, it is impossible not to become enthusiastic about economic and econometric research. I can only hope to continue to exchange ideas and cooperate with them in the future.

I thank Professor Richard T Baillie not only as being my supervisor, but also for his great support and inspiration from the very first until the very last day of my years at Michigan Sate University (MSU). Despite his busy schedule, he has been

incredible in directing and stimulating my research. I deeply appreciate his generous financial support throughout research assistantship which not only helped me and my family financially but also lead to completion of a chapter of this dissertation. I thank the members of my committee, Professor Peter Schmidt, Professor Rowena Pecchenino, and Professor Jeffrey Wooldridge for their insightful comments on this thesis. I thank them being there whenever I needed their advise and help. They have been great support and inspiration throughout my years at MSU. I special word of thanks to Professor Robert de Jong, for his constant help on questions that I didn't have a clue about. He has been not only a great inspiration but also a very good friend to talk about several issues. I am also grateful to Peter Schmidt and Rowena Pecchenino for all of the support and advisement that they offered me throughout my PhD years at MSU. I would like to thank to Professor S. Tamer Cavusgil for reading my dissertation, his financial support throughout a research assistantship and his moral support. I would also like to thank to Professors Ana Maria Herrera and Steven J. Matusz for their useful comments and suggestions about certain chapters of this dissertation.

I thank to my colleague PhD students at Michigan State University, especially to Scott Adams, Ali M. Berker, Chirok Han, Vinit Jagdish, Daiji Kawaguchi, Alina Luca, Pınar Özbay, Yuri Soares, and Chien-Ho Wang for their advise and help on certain stages of this dissertation and for their friendship. I am grateful to the staff in the Department of Economics at MSU, in particular to Ms. Margaret Linch, Ms. Amy Fekete, Ms. Wendy Tate and Ms. Linda Wirick for their support. I would like to thank also to the Graduate School at MSU and the Dean of the College of Social Sciences at MSU for providing financial support.

Of course there are things in life to be more enthusiastic about than writing a dissertation- although at times I have found it necessary that other people pointed this out to me. My special thanks are to my best friend and my wife Gülen Kılıç, and my daughter Ada Birge, for reminding and arousing my enthusiasm for life. Without Gülen's never failing moral support and love and the joy, Ada brought to my life, this this dissertation wouldn't exist.

I would like to thank to my best friends, Gülten and Sait Akgün, and Gün Ayhan Utkan being always there when I needed them both morally and financially. I would like to thank my friends, Leyla Parvizi-Yılan for proof reading my writings and Gültekin Yılan for his moral support. Finally, I would like to thank to my father, my sisters, and my mother in law and father in law for their moral support throughout these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Smooth Transition Autoregressive Model: specification, estimation, and inference

1.1 Introduction

The aim of this chapter is to review the smooth transition model and discuss aspects of the model that are relevant to the subsequent chapters. The presentation is framed in terms of empirical specification and estimation of the smooth transition autoregressive models, the basics of which are discussed in Granger and Teräsvirta (1993), Teräsvirta (1994), and Eitrheim and Teräsvirta (1996). A review of the STAR similar in spirit to this chapter is given by Teräsvirta (1998), and by van Dijk, et al. (2000). This chapter contains three Monte Carlo simulation experiments. The first experiment suggests that standard lag selection criteria (i.e. AIC, BIC) may not always select the correct lag order in STAR models. The second experiment examines the properties of standard and heteroscedasticity consistent (HCC) variants of non-linearity tests. The results suggest that both variants have comparable power, (i.e. the ability to reject linearity when false). However, the size of the standard tests becomes worse when compared to that of HCC variants. The third experiment examines the finite sample properties of nonlinear least squares (NLS) estimates of STAR models. The results indicate that in sample sizes of 100 (which is approximately the available sample size for several macroeconomic variables) the estimation performs poorly in terms of mean square errors. When the sample size is doubled the NLS method performs better.

1.2 The STAR Model: Representation, Specification, and Inference

The smooth transition model for a univariate time series y_t, which is observed at times $t = 1-p, -p, \ldots, -1, 0, 1, \ldots, T-1, T$, is given by

$$y_t = \pi_1' x_t (1 - F(z_t; \gamma, c)) + \pi_2' x_t F(z_t; \gamma, c) + u_t \qquad t = 1, \ldots, T, \tag{1.1}$$

where x_t is a vector consisting of lagged endogenous and exogenous variables, $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p}, w_{1t}, \ldots, w_{kt})'$ and $\pi_i = (\pi_{i,0}, \ldots, \pi_{i,m})'$, $i = 1, 2$, with $m = p + k$. The STAR is obtained if one considers $\tilde{x} = (y_{t-1}, \ldots, y_{t-p})'$. The presentation in this chapter is restricted to the STAR model as it is the model that is used in the applications in this dissertation. The disturbances, (u_t) are assumed to be a martingale difference sequence with respect to the history of the time series up to time $t-1$, which is denoted by $\Omega_{t-1} = y_{t-1}, \ldots, y_{1-p}$. This means that, $E[u_t|\Omega_{t-1}] = 0$. For simplicity, we also assume that the conditional variance of u_t is constant, that is, $E[u_t^2|\Omega_{t-1}] = \sigma^2$. The transition function $F(z_t; \gamma, c)$ is a continuous function that is bounded between 0 and 1. The transition variable z_t can be a lagged endogenous variable, $z_t = y_{t-d}$ for a certain integer $d > 0$, as assumed most of the time in empirical applications. It can also be an exogenous variable, or a function of both lagged exogenous and endogenous variables, say $z_t = z(\tilde{x})$. This function, in principle, can be either, linear or nonlinear and it can be parametric or non-parametric. In most of the applications it is taken to be a linear function of lagged endogenous variables. Another possibility is to let z_t to be a function of a linear time trend $z_t = t$, which is simply the STAR model with smoothly changing parameters, see Lin and Teräsvirta (1994). In order to keep the generality, we do not assume any particular form for the transition function throughout this chapter. One can write out the STAR model given in equation (1.1) in more detail as follows;

$$\begin{aligned} y_t &= (\pi_{1,0} + \pi_{1,1} y_{t-1} + \dots + \pi_{1,p} y_{t-p})(1 - F(z_t; \gamma, c)) \\ &\quad + (\pi_{2,0} + \pi_{2,1} y_{t-1} + \dots + \pi_{2,p} y_{t-p}) F(z_t; \gamma, c) + u_t \end{aligned} \tag{1.2}$$

There are two possible ways of interpreting the STAR model. The STAR model can be thought of as a regime switching model that allows for two regimes, associated with the extreme values of the transition function, $F(.) = 0$ and $F(.) = 1$, where the transition from one regime to the other is gradual. Alternatively, it can also be thought that the STAR model involves a ćontinum of regimes, each associated with a different value of the transition function between 0 and 1. The regime that prevails at time t is determined by the observable variable, z_t and the associated value of $F(.)$. Different choices for the transition function, $F(.)$, leads to different types of state-dependency and/or regime-switching behavior. In most of the applications in econometrics, either logistic,

$$F(z_t; \gamma, c) = \frac{1}{1 + \exp[-\gamma(z_t - c)]}, \gamma > 0, \tag{1.3}$$

or exponential function,

$$F(z_t; \gamma, c) = 1 - \exp[-\gamma(z_t - c)^2], \gamma > 0, \tag{1.4}$$

are the most popular choices. The choice of the logistic function leads to the logistic STAR (LSTAR) model, while the choice of the exponential function results in so called exponential STAR (ESTAR) model. The parameter, c in the LSTAR model is interpreted as the threshold between the two regimes corresponding to $F(.) = 0$ and $F(.) = 1$, in the sense that the logistic function changes from 0 to 1 as z_t increases and $F(c, \gamma, c) = 0.5$. The parameter γ determines the smoothness of the change in the value of the logistic function and thus smoothness of the transition from one regime to the other. Figure 1.1 shows graphs of the logistic and the exponential functions for different parameter specifications. From the figure it is obvious that as γ becomes larger and larger the logistic function approaches to the indicator function $I[z_t > c]$, defined as $I(.) = 1$ if argument is true and $I(.) = 0$, otherwise. As a result the transition from one regime to the other happens almost instantaneously at $z_t = c$. This implies that the LSTAR model nests a two-regime threshold autoregressive (TAR) model as a special case. When $z_t = y_{t-d}$ the model is called the self-exciting TAR model. TAR models are discussed extensively in Tong(1990). When γ is close to zero the logistic function is equal to the constant 0.5 and when $\gamma = 0$, the LSTAR model reduces to a linear model.

The type of regime switching implied by the LSTAR model may be useful for modelling certain economic time series that exhibit asymmetries in terms of expansions and recessions. This is because in the LSTAR model the two regimes correspond to the small and large values of the transition variable z_t relative to the threshold c. Hence it allows one to distinguish expansions and recessions in a given time series. That is the reason why the LSTAR model has been used in the empirical business cycle literature for modelling asymmetric behavior of macroeconomic variables, such as output and unemployment, over a business cycle. For example, if y_t is the rate of unemployment, and if the transition variable is the unemployment rate at a predetermined date, say, the unemployment rate of previous period, $z_t = y_{t-1}$, then the model is capable of distinguishing high and low unemployment relative to a threshold rate, say the natural rate of unemployment, assuming such a rate exists, over the business cycle. Similarly, if y_t is the growth rate of an output variable, and if the transition variable is taken to be the growth rate in the previous period, if $c \approx 0$, then the LSTAR model can distinguish periods of positive and negative growth, namely periods of expansions and contractions over the business cycle. The LSTAR model has been applied by Teräsvirta and Anderson (1992) and Teräsvirta, Tjøstheim and Granger (1994) to study the the different dynamics of industrial production in a number of OECD countries.

It is quite plausible to come up with empirical problems in economics where different types of regime-switching behavior may be much more appropriate than the one implied under the LSTAR model. A major example would be the behavior of real exchange rates. The dynamic behavior of real exchange rates could possibly depend on the magnitude of the deviations from purchasing power parity [PPP]. For instance, the presence of transaction costs may lead to the notion of different regimes in real exchange rates. In particular, the profits from commodity arbitrage, which is generally thought to be the ultimate force behind maintaining PPP, do not make up for the costs involved in the necessary transactions for small deviations from the equilibrium value. This means that there may exist a band around the equilibrium rate in which there is no tendency for the real exchange rate to revert to its equilibrium value. Whenever the rate is outside the band that is specified by the relevant costs, arbitrage becomes profitable. This in turn forces the real exchange rate back towards the band. Dumas (1992), for instance, builds a general equilibrium model that implies the type of behavior outlined above.

If we want to model the type of behavior that is described in the above example by a STAR model, with y_t being the real exchange rate and $z_t = y_{t-d}$, it appears much more appropriate to choose the transition function such that the regimes are associated with small and large absolute values of z_t. A specification along these lines for the transition function would be, for example, the exponential function given in

(1.4) as it may allow one to model symmetric adjustment towards the equilibrium value of real exchange rates. The ESTAR model has been applied to real exchange rates by Michael, Nobay, and Peel(1997), Taylor, Peel, and Sarno (2001) among others.

Note the fact that the exponential function in (1.4) has the property that whenever $\gamma \to 0$ or $\gamma \to \infty$, it becomes a constant, see figure 1. Thus the ESTAR model becomes linear in both cases and it does not nest a self exciting threshold autoregressive (SETAR) model as a special case. To remedy this drawback use of the quadratic logistic function;

$$F(z_t; \gamma, c) = \frac{1}{1 + \exp[-\gamma(z_t - c_1)(z_t - c_2)]}, \; c_1 \leq c_2, \; \gamma > 0 \qquad (1.5)$$

has been suggested in some literature, see for instance, Jansen and Teräsvirta (1996). With the quadratic transition function, if $\gamma \to 0$, the model becomes linear. While when $\gamma \to \infty$, and $c_1 \neq c_2$, the transition function is equal to 1 for $z_t < c_1$ and $z_t > c_2$ and equal to 0 in between. Thus the specification for the transition function in (1.5) nests a three regime SETAR model.

1.3 Properties of the STAR Model

In this section we briefly discuss some properties of the STAR family models. The discussion here is rather informal and intuitive. A much more formal discussion of STAR models is given in Tong (1990) and Teräsvirta (1994). Throughout this section we concentrate on those models with autoregressive lag equal to 1 as it is easier to present the important characteristics of the models without exposing their complex details.

One of the first things to note about STAR models is the relatively large variety of dynamic patterns that can be obtained from choosing the parameters appropriately.

To get an impression of the potential dynamic patterns that can be generated from STAR models, panels of figure 1.2 show realizations of $T = 250$ observations from an ESTAR model with $p = 1$ and $z_t = y_{t-1}$. The realizations are obtained by setting $\pi_{1,1} = -0.3$, $\pi_{2,1} = 0.7$ and the parameters in the exponential function, (γ, c) are set equal to 3 and 0 respectively. The disturbances $u_t, t = 1, \ldots T$ are drawn independently from a standard normal distribution, i.e. $u_t \sim i.i.d.\aleph(0,1)$. All series are started with $y_0 = 0$, and the same values for the disturbances are used to generate subsequent observations. The intercepts $\pi_{1,0}$ and $\pi_{2,0}$ are varied to generate different behavior. One thing that is observed in the panels of figure 1.2 is that by just changing the intercepts over the regimes one can obtain quite rich dynamic patterns in STAR models. In other words by keeping the autoregressive parameters in the two extreme regimes the same, but varying the intercepts generates series with quite different behavior. This also illustrates how the constant terms can play an important roles in nonlinear models. To get some idea about the dynamics of STAR models with different parameter specifications in the autoregressive parameters, realization from the ESTAR model with $\pi_{1,1} = 1, \pi_{2,1} = -0.3$ where all other parameter specifications are the same as above except $\pi_{1,0} = \pi_{2,0} = 0$ is given in panels of figure 1.2 as well. The panel f of figure 1.2 gives a sample realization from an LSTAR model with quadratic logistic function given in (1.5), with $c_1 = 0, c_2 = 0.5$, $\pi_{1,0} = \pi_{2,0} = 0$, and $\pi_{1,1} = 1$, $\pi_{2,1} = -0.3$. In these latter panels of figure 1.2, the autoregressive parameter in the inner/middle regime is unity. This implies that the process acts like a unit root process in the inner/middle regime and becomes a stationary process in the outer regime. Thus as the deviation of the transition variable (in these examples, y_{t-1}) from the threshold level becomes larger and larger, the process becomes increasingly mean reverting in the sense that it tends to move back to the inner/middle regime. Therefore, the generated processes although locally behave as a random walk, globally they are stationary. In this sense the time series realizations are *globally* stationary.

Conditions that need to hold for the stationarity of STAR models is relatively less explored. The required conditions for the stationarity in STAR models have only been established for the first-ordered SETAR model which is obtained from (1.2) with $p = 1$ and (1.3) by allowing $\gamma \to \infty$. Chan, Petrucelli, Tong, and Woolford (1985) show the conditions for the stationarity of the first order SETAR model. They show that the SETAR model is stationary if and only if one of the following conditions is satisfied:

$$\begin{array}{rc} i. & \pi_{1,1} < 1,\ \pi_{2,1} < 1,\ \pi_{1,1},\ \pi_{2,1} < 1; \\ ii. & \pi_{1,1} = \pi_{2,1} < 1,\ \pi_{1,0} > 0; \\ iii. & \pi_{1,1} < 1,\ \pi_{2,1} = 1,\ \pi_{2,0} < 0; \\ iv. & \pi_{1,1} = 1,\ \pi_{2,1} = 1,\ \pi_{2,0} < 0 < \pi_{1,0}; \\ v. & \pi_{1,1}\pi_{2,1} = 1,\ \pi_{1,1} < 0,\ \pi_{2,0} + \pi_{2,1}\pi_{1,0} > 0. \end{array}$$

Condition (i) allows one of the autoregressive (AR) parameters to become smaller than -1. Note also that the conditions $(ii - iv)$ allow unit root behavior in one or both of the regimes. In these cases, the time series is locally nonstationary. Local stationarity is obtained because of the conditions on the intercept terms in two regimes. The conditions $(ii - iii)$ on the intercepts $\pi_{1,0}$ and $\pi_{2,0}$ are such that the time series has a tendency to revert to the stationary regime and hence, the time series is globally stationary. The condition in (iv) also allows the two AR parameters to be unity and hence the time series to be nonstationary in both regimes globally but the conditions on the intercepts guarantees the global stationarity of the series. The testing problem for unit roots in SETAR models is discussed in Caner and Hansen (2001), Enders and Granger (1998) and Berben and van Dijk (1999) and in Chapter 6 of this dissertation.

1.4 Empirical Specification of STAR models

Issues relating to the empirical specification of STAR models have been discussed extensively in Granger(1993), Granger and Teräsvirta (1993), and Teräsvirta(1994). The empirical specification procedure advocated by these authors involve a specification strategy that starts with a simple or restricted model and proceeds to a more general one only if diagnostic tests indicate that the maintained model is inadequate. The procedure efficiently put forward in Teräsvirta (1994) consists of the following steps.

1. Specify an appropriate linear AR model of order $p\,[AR(p)]$ for the time series under study;

2. Test the null hypothesis of linearity against the alternative of STAR–type non-linearity. If linearity is rejected, select the appropriate transition variable z_t and the form of the transition function $F(z_t; \gamma, c)$;

3. Estimate the parameters in the selected STAR model;

4. Evaluate the model using diagnostic tests;

5. Modify the model if necessary;

6. Use the model for descriptive or forecasting purposes.

The following sections discuss each of these steps in detail.

1.4.1 Specifying an appropriate linear AR model

The important issue involved in specifying an $AR(p)$ for the time series under consideration is the selection of the lag order p. The residuals from the $AR(p)$ model need to be approximately white noise as the tests for nonlinearity that are used in the second step are sensitive to residual autocorrelation. There are several conventional methods that can be used for lag selection purposes. The most commonly used criteria in the linear models are the Akaike Information Criterion [AIC], $AIC = T \ln \hat{\sigma}^2 + 2k$, Schwartz Information Criterion [BIC], $BIC = T \ln \hat{\sigma^2} + k(\ln(T))$, Hannan and Quinn Criterion (HQ), $HQ = T \ln \hat{\sigma}^2 + k \ln(\ln(T))$ and the Ljung-Box (LB) statistic. The LB statistic is used to test directly for the residual autocorrelations. The LB statistic is $LB(j) = T(T+2) \sum_{k=1}^{m} \frac{1}{(T-k)} r_k^2(\hat{u})$ where $r_k(\hat{u})$ is the $k-th$ autocorrelation of the residuals. Under the null hypothesis of no residual autocorrelation at lags 1 through m the LB test has an asymptotic χ^2di stribution with $m-p$ degrees of freedom.

These methods are mostly developed for linear time series models. The use of these information criteria and (partial) autocorrelation based methods may not be quite appropriate in case of non-linear time series. One reason is the autocorrelations of non-linear time series processes may have quite different properties. For instance, Granger and Teräsvirta (1999) and Diebold and Inonue (2001) discuss certain regime switching models that have autocorrelations that resemble long memory properties. Especially in finite samples, estimated autocorrelations may be quite substantial and they may decline very slowly. Therefore, when an $AR(p)$ model is considered for these series the selected lag order may become large.

In order to better asses the appropriateness of the methods discussed above within the context of STAR models, the following simulation experiment was conducted. Time series are generated from the ESTAR model given in (1.2) with (1.4) and with $p = 1, z_t = y_{t-1}$. The parameters in the two regimes were specified to be $\pi_{1,1} = 0.6$, $\pi_{2,1} = 0.3$, the smoothness parameter was chosen to be $\gamma = 3$ and the threshold parameter was kept at $c = 0.5$ during simulations. The sample was taken to be $T = 250$ and $T = 500$ observations. The series were generated from $u_t \sim iid\, N(0,1)$. The constant terms in both regimes were kept at zero during simulations. An $AR(p)$ model is specified for the generated ESTAR series where p is set equal to the lag length that minimizes AIC, BIC, HQ, with maximum order $\bar{p} = 6$, or to the minimum lag length for which the LB statistic with $m = 15$ is not statistically significant at the 5% level. Table (1.1) shows the frequencies out of 1000 replications, for which different values of p are selected as the appropriate lag order. The results in (1.1) indicate that in some cases standard lag selection criteria over estimate the autoregressive lag order. This may mean that straightforward application of these criteria may not always be appropriate. Hence, one needs to pay particular attention when using these selection criteria in STAR type modelling.

1.4.2 Testing linearity against STAR

Once an $AR(p)$ model is specified, one can proceed with testing linearity against the alternative of STAR-type nonlinearity. This step is crucial as the failure of rejecting the null hypothesis of linearity will invalidate the STAR modelling for the time series under investigation.

In order to facilitate the discussion in this section re–write the STAR model given in (1.1)

$$y_t = \pi_1' x_t (1 - F(z_t; \gamma, c)) + \pi_2' x_t F(z_t; \gamma, c) + u_t,\ t = 1, \cdots, T, \tag{1.6}$$

where $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, \ldots y_{t-p})'$. The null hypothesis of linearity can be formulated in different ways. A straightforward formulation involves setting the autoregressive parameters in the two regimes to be equal, that is, $H_0 = \pi_1' = \pi_2'$ against the alternative hypothesis $H_1 = \pi_{1,j} \neq \pi_{2,j}$ for at least one $j \in 0, \ldots p$. The testing for linearity against STAR–type nonlinearity is complicated because of the *nuisance* parameters problem. More explicitly, the testing for linearity becomes complicated as there exist unidentified nuisance parameters under the null hypothesis. This is because the STAR model contains parameters which are not restricted by the null hypothesis, but they are present when the null hypothesis holds true. For instance, the null hypothesis given above does not restrict the parameters in the transition function, namely, γ and c. However observe the fact that whenever the null hypothesis holds true the transition function, $F(z_t, \gamma, c)$,and hence, γ and c drop out of the model.

The presence of unidentified nuisance parameters problem can also be seen when expressing the null hypothesis of linearity in several different ways. In addition to the equality of the AR parameters in two regimes, $H_0 = \pi_1' = \pi_2'$, one can formulate the null hypothesis $H_0' = \gamma = 0$. This alternative formulation of the null hypothesis also gives rise to a linear model. For example, if $\gamma = 0$ the logistic function in (1.3) is equal to 0.5 for all values of z_t, and the STAR model in (1.6) reduces to an AR model with parameter $\frac{(\pi_1+\pi_2)}{2}$. Similarly under H_0' the exponential function in (1.4) becomes zero and hence the ESTAR model reduces to a linear AR model with parameter π_1. Under this alternative null hypothesis, $\pi_1$and $\pi_2$and the threshold parameter c can take any values.

A recent account of the problem of unidentified nuisance parameters under the null hypothesis is given in Hansen (1996). The main consequence of the presence of unidentified parameters under the null hypothesis is that the conventional statistical theory can not be applied to obtain the asymptotic distribution of the test statistics. The relevant test statistics in general tend to have non–standard distributions for which an analytic expression is not available. Hence the critical values need to be determined by means of simulation methods which in turn can be quite prohibitive depending on the statistic.

To avoid the nuisance parameters problems in testing for linearity against the STAR type nonlinearity, Luukkonen, Saikkonen and Teräsvirta (1988) proposed to replace the transition function $F(.)$ by a suitable Taylor series approximation. The benefit of such a solution is that the problem is re-parameterized so that the identification problem is no longer present. The linearity is then tested by means of a

Lagrange Multiplier [LM] statistic which has a standard asymptotic χ^2–distribution under the null hypothesis. This procedure is quite appealing as it does not require the estimation of the model under the alternative hypothesis. It also avoids the use of simulation methods to assess the significance of test statistics. One shortcoming of this method is that the LM tests can potentially have power against any other form of misspecification or nonlinearity that may be approximated by the transition function used. In other words, rejection of the null may not always indicate that the correct specification is a STAR model. Thus, diagnostic tests need to be used in evaluating the fit of the models before concluding on the STAR type nonlinearity.

As noted in Granger and Teräsvirta (1993), in testing linearity against the alternative of a STAR model, based on an AR(p) model under the null hypothesis, one needs to distinguish three situations depending on the nature of the transition variable z_t:

1. z_t is a lagged endogenous variable y_{t-d}, with $1 \leq d \leq p$;

2. z_t is a lagged endogenous variable y_{t-d} with $d > p$, or an exogenous variable w_t;

3. z_t is a linear combination of $y_{t-1}, \ldots, y_{t_p}$, that is $\alpha'\tilde{x}$, with α unknown.

The first two situations test linearity against STAR with a specified transition variable, which is most often encountered in applications of STAR modelling in economics and finance. The test statistic differs slightly in the first situation compared to the second as z_t is contained as a regressor in the model under the null hypothesis whenever $d \leq p$. The test statistics that result in situation three are usually interpreted as general tests against STAR-type of nonlinearity, see for instance Teräsvirta (1998). In the rest of this section we first present derivations of the test statistics that are used in the first situation and then give some remarks on the differences that arise in the second and third cases.

Testing against LSTAR

In order to facilitate the presentation we first discuss the tests against the LSTAR model and then the ESTAR model. Given the LSTAR model as in (1.6) with the transition function (1.3) and with $z_t = y_{t-d}$ for certain $1 \leq d \leq p$, re-write (1.6) as

$$y_t = \pi_1' x_t + (\pi_2 - \pi_1)' x_t F(y_{t-d}, \gamma, c) + u_t \tag{1.7}$$

Following the suggestion of Luukkonen *et al.* (1988) approximating the transition function with a first order Taylor approximation around $\gamma = 0$, we have

$$\begin{aligned} F_1(y_{t-d}, \gamma, c) &= F(y_{t-d}, 0, c) + \gamma \frac{\partial F(y_{t-d}, \gamma, c)}{\partial \gamma}|_{\gamma=0} + R_1(y_{t-d}, \gamma, c) \\ &= \frac{1}{2} + \frac{1}{4}\gamma(y_{t-d} - c) + R_1(y_{t-d}, \gamma, c) \end{aligned} \tag{1.8}$$

where $R_1(.)$ is the remainder term. Substituting $F_1(.)$ for $F(.)$ in (1.7) and rearranging terms gives the auxiliary model

$$y_t = \phi_{0,0} + \phi_0' \tilde{x}_t + \phi_1' \tilde{x}_t y_{t-d} + \eta_t \tag{1.9}$$

where $\eta_t = u_t + (\pi_2 - \pi_1)' x_t + R_1(y_{t-d}, \gamma, c)$. Note that under the null hypothesis, the remainder term is equal to 0 and $\eta_t = u_t$. Thus the remainder term does not affect the properties of residuals under the null hypothesis. This in turn implies that the distribution of the test statistics will not be affected by the remainder term. The relationship between the parameters $\phi_i = (\phi_{i,1}, \cdots, \phi_{i,p}), i = 0, 1$, in the auxiliary regression model in (1.9) and the parameters in the LSTAR model in (1.7) are given by

$$\phi_{0,0} = \frac{1}{2}(\pi_{1,0} + \pi_{2,0}) - \frac{1}{4}\gamma c(\pi_{2,0} - \pi_{1,0}) \tag{1.10}$$

$$\phi_{0,d} = \frac{1}{2}(\pi_{1,d} + \pi_{2,d}) - \frac{1}{4}\gamma c(\pi_{2,d} - \pi_{1,d}) - (\pi_{2,0} - \pi_{1,0}) \tag{1.11}$$

$$\phi_{0,j} = \frac{1}{2}(\pi_{1,j} + \pi_{2,j}) - \frac{1}{4}\gamma c(\pi_{2,j} - \pi_{1,j}),\ j = 1, \cdots p,\ j \neq d, \tag{1.12}$$

$$\phi_{1,j} = \frac{1}{4}\gamma c(\pi_{2,j} - \pi_{1,j}),\ j = 1, \cdots, p. \tag{1.13}$$

These relationships show that the restrictions $\pi_1 = \pi_2$ or $\gamma = 0$ imply $\phi_{1,j} = 0$ for $j = 1, \cdots, p$. Therefore testing the null hypothesis $H_0 : \pi_1 = \pi_2$ or $H_0' : \gamma = 0$ in (1.7) is equivalent to testing the null hypothesis $H_0'' : \phi_1 = 0$ in (1.9). This hypothesis can be tested by a standard variable addition test. The test statistic is the standard Lagrange Multiplier test for parameter restriction and denoted by LM_1. This statistic is χ^2 distributed with p degrees of freedom under the null hypothesis of linearity under certain regularity conditions which are given in Saikkonen and Luukonen (1988). This test is usually referred to LM–*type* statistic because the LM_1 statistic does not test the original null hypothesis $H_0' : \gamma = 0$ but rather the auxiliary null hypothesis $H_0'' : \phi_1 = 0$.

The above test statistic does not have power in cases where only the intercept is different across regimes, that is when $\pi_{1,0} \neq \pi_{2,0}$ but $\pi_{1,j} = \pi_{2,j}\, j = 1, \cdots, p$. This can easily be seen from (10–13) which shows that $\phi_{1,j} = 0$, $j = 1, \cdots, p$. Luukonen *et al.* (1988) suggest use of a third order Taylor approximation of the transition function to solve this problem. This is because the second order Taylor approximation of the Logistic function around $\gamma = 0$ is zero. The third order Taylor approximation of the transition function is;

$$\begin{aligned} F_3(y_{t-d}, \gamma, c) = F(y_{t-d}, 0, c) + \gamma \frac{\partial^3 F(y_{t-d}, \gamma, c)}{\partial \gamma^3}|_{\gamma=0} + \\ \frac{1}{6}\gamma^3 \frac{\partial^3 F(y_{t-d}, \gamma, c)}{\partial \gamma^3}|_{\gamma=0} + R_3(y_{t-d}, \gamma, c) \\ = \frac{1}{2} + \frac{1}{4}\gamma\,(y_{t-d} - c) + \frac{1}{48}\gamma^3\,(y_{t-d} - c)^3 + R_3(y_{t-d}, \gamma, c), \end{aligned} \tag{1.14}$$

Now replacing the transition function $F(.)$ with its third order approximation results in the auxiliary model

$$y_t = \phi_{0,0} + \phi_0' \tilde{x}_t + \phi_1' \tilde{x}_t y_{t-d} + \phi_2' \tilde{x}_t y_{t-d}^2 + \phi_3' \tilde{x}_t y_{t-d}^3 + \eta_t \tag{1.15}$$

where $\eta_t = u_t + (\pi_2 - \pi_1)' x_t R_3(y_{t-d}, \gamma, c)$, and $\pi_{0,0}$ and the ϕ_i, $i = 1, 2, 3$, are functions of the parameters π_1, ϕ_2, γ, and c. The null hypothesis of linearity H_0' becomes H_0'' :

$\phi_1 = \phi_2 = \phi_3 = 0$. This hypothesis can also be tested by a standard LM-type test. Under the null hypothesis of linearity, the test statistic denoted by LM_3, has an asymptotic χ^2 distribution with $3p$ degrees of freedom. A parsimonious version of LM_3 statistic can be obtained by first observing that the only parameters that depend on the constants $\pi_{1,0}$ and $\pi_{2,0}$ are $\phi_{2,d}$ and $\phi_{3,d}$ and hence, augmenting the auxiliary equation (1.9) with regressors y^3_{t-d} and y^4_{t-d}, that is,

$$y_t = \phi_{0,0} + \phi_0'\tilde{x}_t + \phi_1'\tilde{x}_t y_{t-d} + \phi_{2,d}' y^3_{t-d} + \phi_{3,d} y^4_{t-d} + \eta_t \tag{1.16}$$

The null hypothesis of linearity can be tested by testing the hypothesis $H_0 : \phi_1 = 0$ and $\phi_{2,d} = \phi_{3,d} = 0$. The resulting test statistic denoted by LM_3^E, has an asymptotic χ^2 distribution with $p + 2$ degrees of freedom.

Testing against ESTAR

Granger and Teräsvirta (1993)and Teräsvirta (1994) show that linearity can be tested against an ESTAR alternative, given by (1.7) with (1.4), by replacing the exponential transition function with a first order Taylor approximation around $\gamma = 0$. Approximating the exponential function around $\gamma = 0$ gives

$$\begin{aligned} F_1(y_{t-d}, \gamma, c) &= F(y_{t-d}, 0, c) + \gamma \frac{\partial F(y_{t-d}, \gamma, c)}{\partial \gamma} |_{\gamma=0} + R_1(y_{t-d}, \gamma, c) \\ &= \gamma(y_{t-d} - c)^2 + R_1(y_{t-d}, \gamma, c), \end{aligned} \tag{1.17}$$

which leads to the auxiliary model,

$$y_t = \phi_{0,0} + \phi_0'\tilde{x}_t + \phi_1'\tilde{x}_t y_{t-d} + \phi_2'\tilde{x}_t y^2_{t-d} + \eta_t \tag{1.18}$$

where $\eta_t = u_t + (\pi_2 - \pi_1)' x_t R_1(y_{t-d}, \gamma, c)$. Granger and Teräsvirta (1993) and Teräsvirta (1994) show that the restriction $\gamma = 0$ corresponds with $\phi_1 = \phi_2 = 0$ in (1.18). The LM_2 statistic which tests this null hypothesis has an asymptotic χ^2 distribution with $2p$ degrees of freedom.

Recently Escribano and Jordå(1999) argue that a first order approximation for the exponential function is not sufficient to capture certain characteristics of the exponential function, especially, the two inflection points of the function. They suggest a second order Taylor approximation,

$$\begin{aligned}F_2(y_{t-d},\gamma,c) &= F(y_{t-d},0,c)+\gamma\frac{\partial F(y_{t-d},\gamma,c)}{\partial\gamma}\mid_{\gamma=0}\\&+\frac{1}{2}\gamma^2\frac{\partial^2 F(y_{t-d},\gamma,c)}{\partial\gamma^2\mid_{\gamma=0}}+R_2(y_{t-d},\gamma,c)\\&=\gamma(y_{t-d}-c)^2-\frac{1}{2}(y_{t-d}-c)^4+R_2(y_{t-d},\gamma,c).\end{aligned} \tag{1.19}$$

Substituting back to (1.7) yields the auxiliary regression,

$$y_t=\phi_{0,0}+\phi_0'\tilde{x}_t+\ \phi_1'\tilde{x}_t y_{t-d}+\phi_2'\tilde{x}_t y_{t-d}^2+\phi_3'\tilde{x}_t y_{t-d}^3+\phi_4'\tilde{x}_t y_{t-d}^4+\eta_t \tag{1.20}$$

The null hypothesis to be tested is $H_0':\phi_1=\phi_2=\phi_3=\phi_4=0$. The resulting LM type test is denoted by LM_4. It has an asymptotic χ^2 distribution with $4p$ degrees of freedom under the null hypothesis. Escribano and Jordå(1999) show by simulation that the LM_4 test have higher power compared to the LM_2 test statistic. When z_t is a lagged endogenous variable y_{t-d} with $d>p$ or an exogenous variable, w_t, the resulting test statistics are very similar to the ones derived above. The only difference is the additional regressors, z_t^i, $i=1,2,\cdots$, that enter the auxiliary model. For example, the auxiliary model (1.18) based on the first Taylor approximation of the exponential function now becomes

$$y_t=\phi_{0,0}+\phi_0'\tilde{x}_t+\phi_{1,0}z_t+\phi_1'\tilde{x}_t z_t+\eta_t$$

while the auxiliary model (1.15)based on the third-order Taylor approximation of the logistic function becomes;

$$y_t=\phi_{0,0}+\phi_0'\tilde{x}_t+\ \phi_{1,0}z_t+\phi_1'\tilde{x}_t z_t+\ \phi_{2,0}z_t^2+\phi_2'\tilde{x}_t z_t^2+\phi_{3,0}z_t^3+\phi_3'\tilde{x}_t z_t^3+\eta_t.$$

In the case linearity is tested against an alternative with $z_t=\alpha'\tilde{x}_t$, the number of auxiliary regressors in the re-parameterized model increases very rapidly when the parameter vector α, which defines the linear combination of $y_{t-1}, \cdots, y_{t-p}$, that is used as transition variable, is left completely unspecified. In order to compute the test in practice, p needs to be set fairly small or the length of the time series has to be sufficiently large. Discussion of this issue can be found in Granger and Teräsvirta (1993).

In the small samples, the usual suggestion is to use $F-$versions of the LM test statistic because these have better size and power properties than the χ^2 versions. The $F-$versions of the LM tests can be computed as follows;

1. Estimate the model under the null hypothesis of linearity by regressing y_t on x_t. Compute the residuals, $\hat{u}_t$ and the sum of squared residuals $SSR_0 = \sum_{t=1}^{T} \hat{u}_t^2$.

2. Estimate the relevant auxiliary regression of $\hat{u}_t$ on x_t and $\tilde{x}_t y_{t-d}^i$, where i will be based on the LM statistic considered. For instance, in the case of LM_3 statistic based on (1.15) i runs from 1 to 3. After estimating the relevant auxiliary model compute the sum of squared residuals and label it by SSR_1.

3. The LM_i statistic is computed as

$$LM_i = \frac{(SSR_0 - SSR_1)/df0}{SSR_1/df1}$$

where $df0$ and $df1$ refers to the relevant degrees of freedoms for the numerator and the denominator which will depend on the LM statistic considered. For example, in the case of LM_3based on (1.15), the $F-$ version is

$$LM_3 = \frac{(SSR_0 - SSR_1)/3p}{SSR_1/(T-4p-1)},$$

which under the null hypothesis is approximately F distributed with $3p$ and $T-4p-1$ degrees of freedom.

Selection of transition variable and function

The selection of an appropriate transition variable in the STAR model and choice of a suitable transition function are usually done during the linearity testing step of the specification. As illustrated in Teräsvirta (1994) the LM_3 statistic, although developed for testing linearity against LSTAR alternative, should have power against ESTAR alternative as well. Intuitively this can be seen by comparing the auxiliary models (1.15) and (1.18) which are used for computing LM_2 and LM_3 statistics respectively. It is easy to see all auxiliary regressors in (1.15) are included in (1.18). Hence it is intuitive to think that LM_3 test might have power against ESTAR alternatives. Observing this Teräsvirta (1994) suggests that the appropriate transition variable in the STAR model can be determined by first, without specifying the form of the transition function, by computing the LM_3 statistics for several candidate transition variables $z_{1t}, \cdots, z_{mt}$, say, and selecting the one for which the $p-$value of the test is smallest. The rationale behind this procedure is that the test should have the highest power when the alternative model is correctly specified, that is, if the correct transition variable is used. In other words if the auxiliary regression model that is used in calculating the LM_3 statistic is considered to approximate the (L)STAR model to a certain degree of accuracy, then selecting z_t as the choice which minimizes the residual variance of the auxiliary model is equivalent to selecting z_t as the variable that maximizes the $LM-$type statistic. This is because $LM-$type statistic is a monotonic transformation of the residual variance. Simulation results in Teräsvirta (1994) indicates that this procedure works quite well in a univariate setting.

If linearity tests indicate presence of STAR type nonlinearity in the time series and an appropriate transition variable has been selected then one usually proceeds with selection of the transition function that appropriately models the STAR type of nonlinear dynamics. In general, the logistic, the exponential, or the quadratic logistic function given in equations, (1.3), (1.4) and (1.5), are used. Teräsvirta (1994) suggests using a decision rule based upon a sequence of tests nested within the null hypothesis corresponding to LM_3. In particular, he proposes to test the hypotheses

$$H_{03} : \phi_3 = 0,$$

$$H_{02} : \phi_2 = 0 \,|\phi_3 = 0,$$

$$H_{01} : \phi_1 = 0 \,|\phi_3 = \phi_2 = 0,$$

in (1.15) by means of LM-type tests. Under the assumption that a first order Taylor approximation of the exponential function is sufficient, it can be observed by inspecting the expressions for the auxiliary parameters, ϕ_1, ϕ_2 and ϕ_3 in terms of parameters of the original STAR model that ϕ_3 is nonzero only if the model is an LSTAR model, that ϕ_2 is zero if the model is an LSTAR model with $\pi_{1,0} = \pi_{2,0}$ and $c = 0$ but is always nonzero if the model is an ESTAR model, and that ϕ_1 is zero if the model is ESTAR model with $\pi_{1,0} = \pi_{2,0}$ and $c = 0$ but is always nonzero if the model is an LSTAR model.These observations indicate the following decision rule; if the $p-$values corresponding to H_{02} is the smallest, an ESTAR model should be selected, while in all other cases an LSTAR model should be the preferred choice.

An alternative method proposed by Escirbano and Jordå(1999) involves use of LM_4 as a test for general STAR-type nonlinearity. The proposed decision rule for choosing between the LSTAR and ESTAR alternatives is based on the observation that, assuming $\pi_{1,0} = \pi_{2,0}$ and $c = 0$ in (1.7), the properties of ϕ_1 and ϕ_2 given above also apply to ϕ_3 and ϕ_4 in (1.20), respectively. Hence, they suggest using the following hypotheses

$$H_0^E : \phi_2 = \phi_4 = 0,$$

$$H_0^L : \phi_1 = \phi_3 = 0,$$

in (1.20). The selection rule is choose LSTAR (ESTAR) model if the minimum $p-$value is obtained for H_0^L (H_0^E). Their simulation results indicate that in case the true data generating process (DGP) is an LSTAR model, the power of the LM_3test is in general higher than the power of the LM_4 test, while reverse holds if DGP is an ESTAR model. This finding is intuitive as the p additional auxiliary regressors $\phi_4' \tilde{x}_t y_{t-d}^4$ in (1.20)are redundant in case of an LSTAR model, and the use of p extra degrees of freedom by the LM_4 statistic causes a loss in power. In case of an ESTAR model however, these extra terms contain vital information which more than compensates the use of additional degrees of freedom. They also find that their procedure in deciding between LSTAR and ESTAR models performs better than that of Teräsvirta (1994). Recent increases in computational power have made the above discussed decision rules about the transition function less important. It is now possible to estimate a number of STAR models with different transition functions and to choose among them at the evaluation stage by using misspecification tests. Given the results in Teräsvirta (1994) that the above mentioned procedure may not select the correct model always, it seems that rather than using these decision rules, one may prefer to estimate several STAR models and choose the one that best describes the data at hand by using certain misspecification tests that will be discussed in section 1.6.

Effects of Heteroscedasticity on tests of STAR type nonlinearity

If there is neglected heteroscedasticity it will have effects similar to residual autocorrelation, in that it may lead to spurious rejection of the null hypothesis of linearity. Wooldridge (1990, 1991) have developed specification tests which can be used in the presence of heteroscedasticity of unknown form. Wooldridge's (1990, 1991) procedure can be applied in the present context to robustify the tests against STAR-type nonlinearity, see also Granger and Teräsvirta (1993, pp.69-70). For an illustration consider the LM_3 test discussed above. The heteroscedasticity-consistent (HCC) variant of the LM_3 statistic based upon (1.15) can be computed as follows;

- Regress y_t on x_t and obtain the residuals $\hat{u}_t$;

- Regress the auxiliary regressors $\tilde{x}_t y^i_{t-d}$, $i = 1, 2, 3$, on x_t and compute the residuals $\hat{e}_t$;

- Weight the residuals $\hat{e}_t$ from the regression in step 2 with the residuals $\hat{u}_t$ obtained in step 1 and regress 1 on $\hat{u}_t\hat{e}_t$. The explained sum of squares from this regression is the $LM-$type statistic.

One issue raised by the simulation results in Lundebrgh and Teräsvirta (1998) on robustifying the linearity tests for the presence of unknown heteroscedasticity is that in some cases the robustification removes most of the power of the linearity tests, so that existing non-linearity may not be detected. In order to better understand the power and size properties of LM-type tests a simulation study is conducted. To see how the two versions of the linearity tests behave under a true DGP of linearity and nonlinearity in the conditional mean data from AR and LSTAR models generated with GARCH and without GARCH effects in the conditional variances. The parameter specifications for different models and conditional variances are given in (1.2), where a missing value denotes the corresponding parameter value in the respective model is equal to zero.

The number of replications in the simulations study is set to 2000. The length of the generated time series is 100, 300, 500, and 1000 observations after removing the first 100 observations from the beginning of the series to eliminate the effects of the initial values which are set to zero. For each replicate two versions of LM_2, LM_3 and LM_4 tests against STAR-type of nonlinearity and corresponding p-values are computed. Namely, standard least squares based version and heteroscedasticity consistent version based on Wooldridge (1990, 1991) are computed.

To see how the two versions of the tests behave when nonlinearity is present in the conditional mean data is generated from LSTAR models with autoregressive lag orders set equal to 1 and 2. For convenience, these DGPs are denoted by LSTAR(1) and LSTAR(2). The conditional variances are generated to be either constant or follow a GARCH(1,1) process. The results from this experiment are given in table (1.3). One clear result from the table is that as the sample size increases the empirical power of the LM-type tests increases substantially for all of the tests considered. The power of the tests is better when LSTAR(2) is the alternative model against linearity. Both versions of the tests have better power when there is GARCH effects. There is a slight difference in power of two versions for moderate sample sizes in that LS versions of the tests have a slightly better power than the HCC version. But this difference disappears as the sample size increases. When there is nonlinearity and GARCH effects both versions have comparable power, the LS variants have marginally better performance, but this may be due to the fact that LS variants do not take GARCH effects into consideration and they may have some power against GARCH effects and thus they most often reject the null of linearity compared to HCC variants. In other words standard versions of the tests may spuriously suggest nonlinearity when there is heteroscedasticity in the conditional variance. This is also evident from table (1.4) which gives the empirical size of the tests. As is evident from table (1.4) the empirical size of the LS versions of all of the tests is higher than that of HCC variants. For most of the cases considered empirical sizes of the LS variants of the tests were found to be higher than the HCC variants and sometimes exceeds the nominal size of the test. Thus for some of the cases especially when there are GARCH effects standard tests suggest nonlinearity erroneously. The results from this simulation experiment indicates that both versions of the tests have good size and power properties in terms of detecting STAR-type of nonlinearity in the conditional mean of a given time series and the HCC version have better size properties than the LS version in the presence of heteroscedasticity of GARCH form.

Presence of outliers and their effects on nonlinearity tests

As might have been observed above STAR models can be parameterized to generate very asymmetric realizations, in the sense that its realizations resemble linear time series with a few outliers. A relevant question in this context is how the LM-type tests discussed above perform when the DGP is a linear model but the observations are contaminated by occasional outliers. This question is studied by van Dijk, Franses and Lucas (1999). Their findings show that in the presence of additive outliers these tests tend to reject the correct null hypothesis too often, even asymptotically. As a solution they suggest to use outlier-robust estimation techniques. An additive outlier can be viewed as an observation which is the genuine data point plus or minus some value. This later value can be nonzero because of a recording error or because of a cause outside the intrinsic economic environment that generates the time series data. For instance, in the case of stock market or exchange rate data a misinterpretation of sudden news flashes, which in turn can cause stock returns or exchange rate returns to take unexpectedly large absolute values. In this sense the data point is aberrant. An additive outlier for the time series y_t formally can be defined by $y_t = x_t + \varphi I[t = \tau]$, $t = 1, \cdots, T$, where $I[t = \tau]$ is an indicator variable, taking a value of 1 when $t = \tau$ and a value of zero otherwise. The time series x_t is the uncontaminated but unobserved time series, while y_t is the observed variable. The size of the outlier is given by φ, and in practice, the value of τ is unknown.

Robust estimators are developed to obtain better parameter estimates in the presence of contamination, by assigning less weight to influential observations such as outliers, see for instance Huber (1981). For example, a robust estimator for the $AR(p)$ model $y_t = \beta' x_t + u_t$ can be obtained as the solution to the first order conditions

$$\sum_{t=1}^{T} \omega_r(r_t) x_t (y_t - \beta' x_t) = 0 \tag{1.21}$$

where r_t denotes the standardized residual, $r_t \equiv (y_t - \beta' x_t)/(\sigma_u \omega_x(x_t)$, with σ_u a measure of scale of the residuals $u_t \equiv y_t - \beta' x_t$ and $\omega_x(.)$ and $\omega_r(.)$ are weight functions that are bounded between 0 and 1. From (1.21) it can be seen that the robust estimator is a type of weighted least squares estimator, with the weight for the $t-$th observation given by the value of $\omega_r(.)$. The functions $\omega_x(.)$ and $\omega_r(.)$ is chosen such that the $t-$th observation receives a relatively small weight if either the regressor x_t or the standardized residual r_t becomes unusually large. The weight function $\omega_r(r_t)$ usually specified in terms of a function $\psi(r_t)$ as $\omega_r(r_t) = \psi(r_t)/r_t$ for $r_t \neq 0$ and $\omega_r(0) = 1$. Common choices for the $\psi(.)$ function are the Huber and Tukey bisquare functions. The Huber $\psi(.)$ function is given by

$$\psi(r_t) = \begin{cases} -\kappa & \text{if } r_t \leq -\kappa, \\ r_t & \text{if } -\kappa < r_t \leq \kappa \\ \kappa & \text{if } r_t > \kappa, \end{cases} \tag{1.22}$$

or $\psi(r) = med(-\kappa, \kappa, r)$, where med denotes the median and $\kappa > 0$. The tuning constant κ determines the robustness and efficiency of the resulting estimator. Since robustness and efficiency properties of the estimator are decreasing and increasing functions of κ, the tuning constant should be chosen such that the two are balanced. Usually κ is taken to be 1.345 to produce an estimator that has an efficiency of 95 percent compared to ordinary least squares,(OLS) estimator if u_t is normally distributed. The weights implied by the Huber function have the attractive property that $\omega_r(r_t) = 1$, if $-\kappa \leq r_t < \kappa$. Only observations outside this region receive less weight. A noted disadvantage of the Huber function is that weights decline to zero very slowly, hence subjective judgement is required to decide whether a weight is small or not. The Tukey's bisquare function is given by

$$\psi(r_t) = \begin{cases} r_t(1 - (r_t/\kappa)^2)^2 & \text{if } \mid r_t \mid \leq \kappa, \\ 0 & \text{if } \mid r_t \mid > \kappa. \end{cases} \tag{1.23}$$

The tuning constant κ again determines the robustness and efficiency of the resultant estimator. Usually κ is set equal to 4.685 to achieve 95 percent efficiency for normally distributed u_t. In this function downweighting occurs for all nonzero values of r_t. Different from the Huber function the resulting weights decline to zero quite rapidly. There are several possibilities for the weighting function proposed in the literature, for a discussion of possible specifications for $\psi(.)$ see van Dijk et al. (1999).

The weight function $\omega_x(x_t)$ for the regressor is usually specified as

$$\omega_x(x_t) = \psi(d(x_t)^\alpha)/d(x_t)^\alpha, \tag{1.24}$$

where $\psi(.)$ is any appropriate function, $d(x_t)$ is the distance given by $d(x_t) = |x_t - m_x|/\sigma_x$, with m_x and σ_x measures of location and scale of x_t, respectively. These measures can be estimated robustly by the median $m_x = med(x_t)$ and median absolute deviation (MAD) $\sigma_x = 1.483.med|x_t - m_x|,$. The constant 1.483 is used to make the MAD estimator a consistent estimator of the standard deviation where x_t is normally distributed. It is usually the practice to set $\alpha = 2$ in order to obtain robust standard errors.

Since weights $\omega_r(.)$ depend on the unknown parameters β they need to be determined endogenously. This in turn implies that the first order condition given in (1.21) is nonlinear in β and σ_u, and estimation of these parameters requires an iterative procedure. Recognizing that $\omega_r(.)$ is a function of $(\beta, \sigma_u), \omega_r(\beta, \sigma_u)$, and denoting the estimates from the nth iteration by $\hat{\beta}^{(n)}$and $\hat{\sigma}_u^{(n)}$ respectively, it follows from (1.21) that $\beta^{(\hat{n+1})}$ can be obtained as the weighted least squares estimate

$$\beta^{(\hat{n+1})} = \frac{\sum_{t=1}^{n} \omega_r(\beta^{(\hat{n})}, \sigma_u^{(\hat{n})}) x_t y_t}{\sum_{t=1}^{n} \omega_r(\beta^{(\hat{n})}, \sigma_u^{(\hat{n})}) x_t^2}$$

where the estimate of σ_u can be updated at each iteration using a robust estimation of scale, such as MAD given above.

The above method gives robust estimators under the null hypothesis of linearity. Robust estimation of STAR models has not been developed yet. The robust estimation procedures allow one to construct test statistics that are robust to outliers. As illustrated in van Dijsk et al. (1999) outlier robust variants of LM type tests discussed above can be obtained as TR^2, using the R^2 from the regression of the weighted residuals $\hat{\psi}(\hat{r}_t) = \hat{\omega}_r(\hat{r}_t)$ on the weighted regressors $\hat{\omega}_x(x_t).*\nu'$ where $.*$ denotes element-by-element multiplication, ν' is the vector that includes the auxiliary regressors. For instance in the case of LM_3 statistics $\nu = (x_t', x_t'z_t, x_t'z_t^2, x_t'z_t^3)$. The weights are obtained from the robust estimation of the $AR(p)$ under the null. The $F-$versions of the tests can be computed as well. The simulation results in van Dijk et al. (1999) suggest that the robustified $LM-$type tests have good size properties in small samples, also in the presence of outliers. In the case of no outliers the power of the tests are lower than that of their non-robust counterparts. The power of standard tests decreases drastically in the presence of outliers while power of the robustifed tests is hardly affected.

1.5 Estimation of STAR Models

If the linearity tests indicate presence of STAR type of nonlinearity then one needs to determine the transition variable z_t and the transition function $F(z_t, \gamma, c)$ as above. The next step involves estimation of the relevant STAR model. The estimation of the STAR model carried out by nonlinear least squares (NLS). The parameter vector $\pi = (\pi_1', \pi_2', \gamma, c)'$ can be estimated as

$$\hat{\pi} = argmin_\pi\, Q_T(\pi) = argmin_\pi \sum_{t=1}^{T} (y_t - S(x_t; \pi))^2, \tag{1.25}$$

where $S(x_t; \pi)$ is the skeleton of the model, that is,

$$S(x_t; \pi) = \pi_1' x_t (1 - F(z_t, \gamma, c)) + \pi_2' x_t F(z_t, \gamma, c). \tag{1.26}$$

Under the normality assumption on disturbances NLS is equivalent to maximum likelihood estimates. Under certain regularity conditions, which are discussed in Gallant (1987) Pötcher and Prucha (1997) among others, the NLS estimates are consistent and asymptotically normal. In other words, under certain conditions

$$\sqrt{T}(\hat{\pi} - \pi_0) \rightarrow N(0, \Sigma), \tag{1.27}$$

where π_0 denotes the true parameter vector, and Σ denotes the asymptotic covariance matrix of the NLS estimates, $\hat{\pi}$. Σ can be estimated consistently by $\hat{H}_T^{-1}\hat{J}_T\hat{H}_T^{-1}$, where $\hat{H}_T$ is the Hessian evaluated at $\hat{\pi}$,namely;

$$\hat{H}_T = -\frac{1}{T}\sum_{t=1}^{T}\nabla^2 q_t(\hat{\pi} = \frac{1}{T}\sum_{t=1}^{T}[\nabla S(x_t;\hat{\pi})\nabla S(x_t;\hat{\pi})' - \nabla^2 S(x_t;\hat{\pi})\hat{u}_t], \tag{1.28}$$

with $q_t(\hat{\pi}) = (y_t - S(x_t;\hat{\pi}))^2$, $\nabla S(x_t;\hat{\pi}) = \partial S(x_t;\hat{\pi})/\partial\pi$, and $\hat{J}_T$ is the outer product of the gradient

$$\hat{J}_T = \frac{1}{T}\sum_{t=1}^{T}\nabla q_t(\hat{\pi})\nabla q_t(\hat{\pi})' = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\nabla S(x_t;\hat{\pi})\nabla S(x_t;\hat{\pi})'. \tag{1.29}$$

The estimation can be performed by using any standard nonlinear optimization procedure, see Hamilton (1994, sec. 5.7) for a brief survey. The following are the important issues that deserve attention when carrying out the estimation procedure.

Use of good starting values will help optimization procedure to work smoothly. In order to get good starting values, note that for fixed values of the parameters in the transition function, γ and c, the STAR model is linear in the autoregressive parameters π_1 and π_2. Thus conditional upon γ and c, estimates of $\pi = (\pi_1', \pi_2')'$ can be obtained by ordinary least squares (OLS)as

$$\hat{\pi}(\gamma, c) = (\sum_{t=1}^{T} x_t(\gamma, c)')^{-1}(\sum_{t=1}^{T} x_t(\gamma, c)y_t), \tag{1.30}$$

where $x_t(\gamma, c) = (x_t'(1 - F(z_t, \gamma, c)), x_t'F(z_t, \gamma, c))'$ and the notation $\pi(\gamma, c)$ indicates that the estimate of π is conditional upon γ and c. The OLS residuals and the corresponding variance can be computed as $\hat{u}_t = y_t - \hat{\pi}(\gamma, c)'x_t(\gamma, c)$ and $\hat{\sigma}^2(\gamma, c) = T^{-1}\sum_{t=1}^{T}\hat{u}_2^2(\gamma, c)$. An appropriate method proposed in the literature (see for instance Teräsvirta (1998)) for obtaining sensible starting values for the nonlinear optimization algorithm involves a two-dimensional grid search over γ and c and selects those parameter estimates which gives the smallest estimate for the residual variance $\hat{\sigma}(\gamma, c)$.

Another method suggested by Leybourne, Newbold and Vougas (1998) to simplify the estimation problem involves concentrating the sum of squares function. Since the STAR model is linear in the autoregressive parameters for fixed values of γ and c, the sum of squares function $Q_T(\pi)$ can be concentrated with respect to π_1 and $\pi_2$as

$$Q_T(\gamma, c) = \sum_{t=1}^{T}(y_t - \pi(\gamma, c)' x_t(\gamma, c))^2. \tag{1.31}$$

The estimates of $\pi(\gamma, c)$ is obtained from minimization of (1.31) for different values of γ and c and the one that gives the lowest residual variance is chosen for γ and c as the final estimates. This reduces the dimensionality of the NLS estimation problem considerably, as the sum of squares function given in (1.31) is minimized with respect to the two parameters γ and c only.

One difficulty reported on the estimation of STAR models is obtaining a precise estimate of the smoothness parameter γ. A reason why it is difficult to obtain a precise estimate of γ is that for large values of γ, the shape of the transition function changes only little. Thus in order to get an accurate estimate of γ one needs many observations in the immediate neighborhood of the threshold c. As this is not typically the case, the estimate of γ is usually imprecise and often insignificant when judged by its t-statistic. Granger and Teräsvirta (1993) and Teräsvirta (1994) argue that insignificance of the estimate of γ should not be taken as evidence against the presence of STAR-type nonlinearity. This should be assessed by means of different diagnostics, some of which will be discussed in the next section.

To better understand the finite sample properties of the NLS estimates, the following simulation experiment is performed. Time series are generated from an ESTAR model, with $\pi_1 = 1, 0.8, 0.5$, $\pi_1^* = 0.9, 0.4, -0.5$, $\gamma = 1, 5, 15$, $c = 0, 0.5$ and $u_t \sim i.i.d.N(0, 1)$. The sample size is taken to be $T = 100$, 300, and 500 observations.

In each replication the first 100 observations are deleted in order to minimize the initialization problem. The parameters in the STAR model, with the lag orders set at their true values and the correct transition function and variable, is estimated by the NLS. Tables 1.5 through 1.10 show the mean parameter estimates, mean standard errors, and root mean squared errors, skewness and kurtosis. The simulation results are based on 2000 replications. The findings of the simulation experiment indicate that as the sample size grows from 100 to 500 the parameter estimates improve in terms of having smaller biases, root mean square errors and smaller standard errors. It seems that for most of the designs the estimate of autoregressive and threshold parameters are very precise especially for samples sizes of 300 and 500. On the other hand, the estimate of the smoothness parameter has relatively higher biases, root mean square errors, skewness and kurtosis. Although the precision of the smoothness parameter increases with sample size, for small and large parameter specifications the estimates are relatively less precise. The skewness and kurtosis values indicate that the distribution of parameter estimates are far from being normal for especially small sample sizes. As the sample size increases estimated skewness and kurtosis statistics get closer to values that are more in line with a normally distributed random variable. The kurtosis for π and γ is mostly above 3 indicating that larger estimates are obtained for these parameters than one would expect under a normally distributed random variable. On the other hand kurtosis estimates for π^* and c are mostly piled up around values less than 3. In all experimental designs, the parameter estimates have positive skewness except in one of the designs in which $\pi = 0.5, \pi^* = -0.5, \gamma = 5, c = 0$. The nonzero skewness estimates reported in tables 1.5-1.10 indicate that distribution of parameter estimates are not symmetric around the mean parameter estimate and most often skewed in the positive direction. The general result from this experiment is that usually the NLS performs poorly for sample sizes of 100 (which corresponds the sample size available for many macroeconomic time series) and improves for sample sizes higher than 300. In applications of STAR models with reasonable sample sizes one needs to interpret inference based on asymptotic theory with caution.

1.6 Diagnostic Checking of Estimated STAR model

This section discusses some diagnostic tests which can be used to evaluate estimated STAR models. In particular, diagnostic tests for residual autocorrelation, remaining nonlinearity, and parameter constancy will be discussed as developed in Eitrheim and Teräsvirta (1996), Lundbergh, Teräsvirta, and van Dijk (1999), and van Dijk and Franses (1999).

1.6.1 Tests for serial autocorrelation

In order to facilitate the review consider the STAR model of order p,

$$y_t = S(x_t; \pi) + u_t \tag{1.32}$$

where $x_t = (1, \tilde{x}_t')', \tilde{x}_t = (y_{t-1}, \cdots, y_{t-p})'$ as before and $S(x_t; \pi)$ is given in (1.26), is called the skeleton of the model. As shown in Eitrheim and Teräsvirta (1996) an LM-test for k-th order serial dependence in u_t can be obtained as TR^2, where R^2 is the coefficient of determination from the regression of $\tilde{u}_t$ on $\partial S(x_t, \hat{\pi})/\partial\pi$ and k lagged residuals $\hat{u}_{t-1}, \cdots, \hat{u}_{t-k}$. Hats indicate that the relevant quantities are estimates under the null hypothesis of serial independence of u_t. The resulting test statistic is denoted by $LM_S(k)$, is χ^2 distributed with k degrees of freedom. As shown in Eitrheim and Teräsvirta (1996), this test is a generalization of the LM-test for serial correlation in an $AR(p)$ model of Breusch and Pagan (1979), which is based on the auxiliary regression

$$\hat{u}_t = \sum_{i=1}^{p} \alpha_p y_{t-p} + \sum_{i=1}^{k} \hat{u}_t + v_t \tag{1.33}$$

where now $\hat{u}_t$ is the residuals from $AR(p)$ model. In a linear $AR(p)$ model (without an intercept) $S(x_t;\pi) = \sum_{i=1}^{p} \pi_i y_{t-i}$, and $\frac{\partial S(x_t;\hat{\pi})}{\partial \pi} = (y_{t-1}, \cdots, y_{t-p})'$. In the case of STAR model, the skeleton is given by $S(x_t;\pi) = \pi_1' x_t(1 - F(z_t,\gamma,c)) + \pi_2' x_t S(z_t,\gamma,c)$. Hence, in this case the parameter vector is $\pi = (\pi_1, \pi_2, \gamma, c)$ and the relevant partial derivatives $\frac{\partial S(.)}{\partial \pi}$ can be obtained in a straightforward manner, for details see Eitrheim and Teräsvirta (1996). The nonlinear function $S(x_t;\pi)$ needs to be twice differentiable in order for the above testing procedure to work.

1.6.2 Testing for remaining nonlinearity

It is important to assess whether the estimated nonlinear model adequately captures the nonlinearity in the time series under investigation. An intuitive method to examine this question is to apply a test for no remaining nonlinearity in the estimated model(s). In the case of STAR models, an approach is to specify the alternative hypothesis of remaining nonlinearity as the presence of an additional regime. This approach is suggested by Eitrheim and Teräsvirta (1996). For instance, one can test the null hypothesis that a two regime model is adequate against the alternative that a third regime is necessary. Eitrheim and Teräsvirta (1996) develop an LM statistic to test a two regime STAR model against the alternative of an additive 3-regime model which can be written as,

$$y_t = \pi_1' x_t + (\pi_2 - \pi_1)' x_t F_1(z_{1t}, \gamma_1, c_1) + (\pi_3 - \pi_2)' x_t F_2(z_{2t}, \gamma_2, c_2) + u_t \tag{1.34}$$

where $F_1(.)$ and $F_2(.)$ are the transition functions given either in (1.3) or (1.4) and where $c_1 < c_2$ is also assumed. The null hypothesis of a two regime STAR model can be expressed as either $H_0 : \gamma_2 = 0$ or $H_0 : \pi_3 = \pi_2$. This testing problem suffers from a similar identification problem as the problem of testing the null hypothesis of linearity against the alternative of a two-regime STAR model discussed in section 4. The proposed solution is the same, namely approximating the transition function $F_2(z_{2t}, \gamma_2, c_2)$ around $\gamma_2 = 0$. In the case of a third order approximation, it is shown in Eitrheim and Teräsvirta (1996) that the resulting auxiliary model will be

$$y_t = \phi_0' x_t + (\pi_2 - \pi_1)' x_t F_1(z_{1t}, \gamma_1, c_1) + \phi_1' \tilde{x}_t z_{2t} + \phi_2' \tilde{x}_t z_{2t}^2 + \phi_3' \tilde{x}_t z_{2t}^3 + \eta_t \qquad (1.35)$$

where the parameters ϕ_i, $i = 0, 1, 2, 3$, are functions of the parameters π_1, π_2, γ_2 and c_2. The null hypothesis H_0': $\gamma_2 = 0$ in (1.34) translates into $H_0'' : \phi_1 = \phi_2 = \phi_3 = 0$ in (1.35). The test statistic is computed as TR^2 from the auxiliary regression of the residuals obtained from estimating the model under the null hypothesis $\hat{u}_t$ on the partial derivatives of the regression function with respect to the parameters in the two-regime model, π_1, π_2, γ_1 and c_1, evaluated under the null hypothesis, and the auxiliary regressors $\hat{x}_t z_{2t}, i = 1, 2, 3$. The resulting test statistic is shown in Eithrheim and Teräsvirta (1996) to have an asymptotic χ^2 distribution with $3p$ degrees of freedom. The statistic is denoted by $LM_{AMR,3}$, where the subscript AMR is used to indicate that this statistic is designed as a test against an additive multiple regime model.

van Dijk and Franses (1999) derived an LM-type statistic for testing the null of a two-regime STAR model against the alternative of a four regime STAR model by using the same procedure as above. The null hypothesis is the two-regime STAR model given in (1.2) and the alternative now is given by the following multiple regime STAR model developed in van Dijk and Franses (1999);

$$\begin{aligned} y_t = {} & [\pi_1' x_t (1 - F(z_{1t}, \gamma_1, c_1)) + \pi_2' x_t F_1(z_{1t}, \gamma_1, c_1)][1 - F_2(z_{2t}, \gamma_2, c_2)] \qquad (1.36) \\ & + [\pi_3' x_t (1 - F_1(z_{1t}, \gamma_1, c_1)) + \pi_4' x_t F_1(z_{1t}, \gamma, c_1)] F_2(z_{2t}, \gamma_2, c_2) + u_t \end{aligned}$$

In this model the relationship between y_t and its lagged values are given by a linear combination of four linear AR models, each associated with a particular combination of $F_1(z_{1t})$ and $F_2(z_{2t})$ being equal to 0 or 1. This model is called Multiple Regime STAR (MRSTAR) model and is discussed in detail in van Dijk and Franses (1999). The test statistic developed in van Dijk and Fransess (1999) involves replacement of second transition function $F_2(z_{2t}, \gamma_2, c_2)$ by a third order Taylor approximation to render the auxiliary regression

$$\begin{aligned} y_t = \phi_0' x_t + (\pi_2 - \pi_1)' x_t F_1(z_{2t}, \gamma_1, c_1) + \phi_1' \tilde{x}_t z_{2t} + \phi_2' \tilde{x}_t z_{2t}^2 \\ + \phi_3' \tilde{x}_t z_{2t}^3 + \phi_4' \tilde{x}_t F_1(z_{1t}, \gamma_1, c_1) z_{2t} + \phi_5' \tilde{x}_t F_1(z_{1t}, \gamma_1, c_1) z_{2t}^2 \\ + \phi_6' \tilde{x}_t F_1(z_{1t}, \gamma_1, c_1) z_{2t}^3 + \eta_t \end{aligned} \tag{1.37}$$

The null hypothesis again can be stated as $H_0 : \gamma_2 = 0$ in (1.37). It becomes $H_0' : \phi_j = 0,\ j = 1, \cdots, 6$ which can be tested exactly the same way as above. The resulting test statistic denoted by $LM_{EMR,4}$i s asymptotically χ^2di stributed with $6(p+1)$ degrees of freedom, where the subscript EMR indicates that the statistic is designed as a test against an 'encapsulated' multiple regime model.

1.6.3 Testing parameter constancy

In order to assess the parameter stability in the estimated model LM type tests are developed in Lundbergh, Teräsvirta and van Dijk (1999). For this purpose they consider the MRSTAR model given in (1.37) with the second transition function F_2 being a function of time t rather than z_{2t}. In other words replacing the transition variable in the second transition function with a t gives rise to so called Time-Varying STAR (TVSTAR) model, which allows for both nonlinear dynamics of the STAR-type and time varying parameters. With this replacement the model in (1.37) becomes

$$\begin{aligned} y_t = [\pi_1' x_t (1 - F(z_t, \gamma_1, c_1)) + \pi_2' x_t F_1(z_t, \gamma_1, c_1)][1 - F_2(t, \gamma_2, c_2)] \\ + [\pi_3' x_t (1 - F_1(z_t, \gamma_1, c_1)) + \pi_4' x_t F_1(z_t, \gamma, c_1)] F_2(t, \gamma_2, c_2) + u_t. \end{aligned} \tag{1.38}$$

This model is discussed in detail in Lundbergh, Teräsvirta and van Dijk (1999). The relevance of this model here is that by testing the hypothesis H_0: $\gamma_2 = 0$, one tests for parameter constancy in the two-regime STAR model (1.2), against the alternative of smoothly changing parameters. The appropriate LM-type test statistic based on a relevant, say a j^{th}-order Taylor approximation of $F_2(t, \gamma_2, c_2)$, is denoted by $LM_{C,j}$ is similar to the $LM_{EMR,j}$ statistic with $z_{2t} = t$. They also note that the asymptotic theory works fine even if the transition variable is a non-stationary deterministic trend, see also Lin and Teräsvirta (1994).

1.7 Impulse response function analysis of estimated STAR model

Since parameter estimates generally do not provide much information about the dynamics of the estimated STAR model one needs to utilize alternative tools in order to characterize the dynamic behavior of the series under study. Impulse response functions (IRF) are convenient methods of evaluation of the properties of the estimated model, as they allow one to examine the effects of shocks u_t on future evolution of the time series under investigation and hence provide a measure of the response of y_{t+k} to an impulse ι at time t.

In the case of linear models IRFs are defined as the difference between two realizations of y_{t+k} which start from identical histories of the time series up to time $t-1$, denoted as ω_{t-1}. In one realization, the process is hit by a shock of size *iota* at time t, while in the other realization no shock occurs at time t. All shocks occur between the intermediate periods are set equal to zero in both realizations. This IRF is named by van Dijk and Teräsvirta (2000) as the traditional IRF and given by

$$TI_y(k, \iota, \omega_{t-1}) = E[y_{t+k} \mid u_t = \iota, u_{t+1} = \cdots = u_{t+k} = 0, \iota] - \qquad (1.39)$$

$$E[y_{t+k} \, |u_{\,t} = 0, u_{t+1} = \cdots = u_{t+k} = 0, \iota],$$

for $k = 0, 1, 2, \cdots$, where E denotes the expectation operator. The second conditional expectation in (1.40) is usually called the benchmark profile of the series. The IRF given in (1.40) has certain properties whenever the time series y_t follows a linear model. First of all it is *symmetric,* as such a shock of size $-\iota$ has an effect that is exactly opposite to that of a shock of size $+\iota$. Moreover, it is *linear* in the sense that the IRF is proportional to the size of the shock. Lastly, it is *history independent* as its shape does not depend on the particular history ω_{t-1}. These properties of traditional IRF function can be easily observed by considering an $AR(1)$ model. In the $AR(1)$ model, $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$, since $y_{t+k} = const. + \beta_1^k y_t + u_{t+k} + \beta_1 u_{t+k-1} + \cdots + \beta_1^k u_t$ one can easily show that $TI_y = \beta_1^k \cdots$, for $k = 0, 1, 2, \cdots$. From this equation it is trivial to observe the mentioned properties. As discussed in Koop et al. (1996) and Pesaran and Potter (1997) in general these somewhat simple properties do not hold when the time series follows a nonlinear model, for example a STAR model. It is shown that the impact of a shock depends not only on the history of the process but also on the sign and size of the shock. Furthermore, as shown in Pesaran and Potter (1997), when one wants to analyze the effect of a shock on the time series $k > 1$ periods ahead, the assumption that no shocks occur in the intermediate periods may give misleading inference concerning the propagation mechanism of the model. The assumption of no shocks in the intermediate periods for the linear models is justified by the existence of Wold representation of the linear time series,

$$y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j} \tag{1.40}$$

which shows that shocks in different periods do not interact. For nonlinear time series there does not exist Wold representation however. Nonlinear time series can be represented in terms of past and present shocks by means of the Volterra expansion,

$$y_t =] \sum_{j=0}^{\infty} \psi_j u_{t-j} + \sum_{j=0}^{\infty} \sum_{j=i}^{\infty} \zeta j i u_{t-j} u_{t-i} \tag{1.41}$$
$$+ \sum_{j=0}^{\infty} \sum_{j=i}^{\infty} \sum_{h=i}^{\infty} \zeta_{ji} u_{t-j} u_{t-i} u_{t-h} + \cdots,$$

as given in Granger and Teräsvirta (1993). From this representation of any nonlinear model it is obvious that the effect of the shock u_t on y_{t+k} depends on the shocks $u_{t+1}, \cdots, u_{t+k}$, as well as on the history of the shocks, $u_{t-1}, u_{t-2}, \cdots$. In order to deal with these problems Koop et al. (1996) developed so called the Generalized Impulse Response Function (GIRF). GIRF for a specific shock $u_t = \iota$ is defined as

$$GI_y(k, \iota_{t-1}, \omega) = E[y_{t+k} \, | u_{\ t} = \iota, \omega_{t-1}] - E[y_{t+k} \, | \omega_{\ t-1}], \tag{1.42}$$

for $k = 1, 2, \cdots$. Note that the expectations of y_{t+k} are conditioned only on the history and/or on the shock. In other words, the problem of dealing with shocks occurring in the intermediate periods is dealt with by averaging them out. That explains also why the benchmark profile is the expectation of y_{t+k} given only the history of the process ω_{t-1}. Therefore, in the benchmark profile the current shock is averaged out as well. This GIRF reduces to traditional IRF when the model is linear.

Koop et al. (1996) emphasize that the GIRF given in (1.42) is indeed a random variable. The GIRF is a function of ι and ω_{t-1}, which are realizations of the random variables u_t and the information set, Ω_{t-1}. In this framework, GIRF given in (1.42) can be written in a more general form as

$$GI_y(k, u_t, \Omega_{t-1}) = E[y_{t+k} \, | u_{\ t} \Omega_{t-1}] - E[y_{t+k} \, | \Omega_{\ t-1}] \tag{1.43}$$

The reformulation in (1.45) is flexible and useful for certain purposes as it allows one to consider a number of conditional versions of GIRF that can be obtained. For example, one might consider only a particular history ω_{t-1} and treat GI as a random variable in terms of u_t only, that is,

$$GI_y(k, u_t, \omega_{t-1}) = E[y_{t+k} \mid u_t, \omega_{t-1}] - E[y_{t+k} \mid \omega_{t-1}]. \tag{1.44}$$

It is also possible to reverse the roles of the shock and history by fixing the shock at $u_t = \iota$ and defining the GIRF as a random variable with respect to the history, Ω_{t-1}. Koop et al (1996) show that in general it is possible to compute GIRFs conditional on any particular subsets A and B of shocks and histories respectively.

The GIRFs can be utilized in several ways in analyzing the dynamic properties of the estimated model. They can be used to analyze the persistence of shocks. A shock $u_t = \iota$ is called *transient* at history ω_{t-1} if $GI_y(k, \iota, \omega_{t-1})$ becomes equal to zero as $k \to \infty$. If on the other hand, GI approaches a non zero finite value when $k \to \infty$ then the shock is said to be *persistent*. It is intuitive to think that if a time series process is stationary and ergodic, the effects of all shocks eventually converge to zero for all possible histories of the process. Hence the distribution of $GI_y(k, \iota, \omega_{t-1})$ collapses to a spike at 0 as $k \to \infty$. In contrast, for non-stationary time series the dispersion of the distribution of $GI_y(k, \iota, \omega_{t-1})$ is positive for all k. Koop et al. (1996) suggest that the dispersion of the distribution of $GI_y(k, \iota, \omega_{t-1})$ at finite horizons conveniently can be used to obtain information about the persistence of shocks. For instance, one can compare densities of GIRFs conditional on positive and negative shocks to find out whether there is a difference in terms of persistence for negative and positive shocks.

GIRFs can also be used to asses the significance of asymmetric effects over time. Potter (1994) defines a measure of asymmetric response to a particular shock $u_t = \iota$, given a particular history ω_{t-1}, as the sum of the GI for this particular shock and the GI for the shock of the same magnitude but with opposite sign, that is,

$$ASY_y(k, \iota, \omega_{t-1}) = GI_y(k, \iota, \omega_{t-1}) + GI_y(k, -\iota, \omega_{t-1}). \tag{1.45}$$

An alternative measure of asymmetry can be obtained by considering the distribution of the random asymmetry measures given above for each history and average across all possible histories to obtain

$$ASY_y^*(k,\iota) = E[GI_y(k,\iota,\omega_{t-1})] + E[GI_y(k,-\iota,\omega_{t-1})] \tag{1.46}$$
$$= E[y_{t+k} \mid u_t = \iota] + E[y_{t+k} \mid u_t = -\iota].$$

One problem in computing the GIRFs is that the analytic expressions for the conditional expectations are not available for $k > 1$. Therefore they need to be estimated. Koop et al. (1996) discusses in detail simulation methods to estimate GIRFs. In particular Monte Carlo or bootstrap methods are suggested for computation of GIRFs. For details see Koop et al. (1996).

1.8 Conclusion

This chapter reviewed the STAR models in reference to specification, estimation and inference. Both ESTAR and LSTAR models are discussed extensively. Issues pertaining to testing presence of STAR type nonlinearity, specification of autoregressive orders, estimation, diagnostic checking and inference procedures are discussed in some detail. The simulation experiments indicate that use of standard information criteria, say AIC or BIC may not always give the correct autoregressive order within the STAR models hence they need to be used cautiously. Both standard and heteroscedasticity consistent versions of STAR type nonlinearity tests have comparable power properties in detecting STAR type of nonlinearity. The performance of NLS in finite samples is analyzed by an extensive Monte Carlo experiments. The findings of the experiment indicate that NLS performs poorly for sample sizes of 100 but improves for sample sizes higher than 300.

BIBLIOGRAPHY

[1] Anderson, H. M.(1997), Transactions costs and non-linear adjustment towards equilirium in the US treasury bill market, *Oxford Bulletin of Economics and Statistics* **59**, 465–484.

[2] Berben, R.-P. and D. van Dijk (1999), Unit root testsand asymmetric adjustment, Econometric Institute Report 9902, Erasmus University Rotterdam.

[3] Breusch, T.S. and A.R. Pagan (1979), A simple test for heteroscedasticity and random coefficient variation, *Econometrica* **47**, 1287–94.

[4] Caner, M. and B. E. Hansen (2001), Threhold autoregression with a unit root, *Econometrica* **69** 1555-1596.

[5] Chan, K.S., J.D. Petrucelli, H. Tong, and S.W. Woolford (1985), A multiple threshold AR(1) model, *Journal of Applied Probability* **22**, 267–279.

[6] Dumas, B. (1992), Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**, 153–180.

[7] Enders, W. and C.W.J. Granger (1998), Unit root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business and Economic Statistics* **16**, 304–311.

[8] Eitrheim Ø. and T. Teräsvirta (1996), Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–76.

[9] Gallant, A. R. (1987), Nonlinear Statistical Models, New York: John Wiley

[10] Granger, C.W.J. and T. Teräsvirta (1993), Modelling Nonlinear Economic Relationships, Oxford: Oxford University Press.

[11] Hansen, B. E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**, 413–30.

[12] Huber, P.J. Robust Statistics, New York: John Wiley

[13] Jansen, D.W. and T. Teräsvirta (1996), Testing parameter constancy and super exogeneity in econometric equations, *Oxford Bulletin of Economics and Statistics* **58**, 735–768.

[14] Koop, G., M. H. Pesaran and S. M. Potter (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119–147.

[15] Lin , C-F.J. and T. Teräsvirta (1994), Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**, 211–228.

[16] Leybourne, S. P. Newbold, and D. Vougas (1998), Unit roots and smooth transitions, *Journal of Time Series Analysis* **19**, 83–97.

[17] Lundbergh, S., T. Teräsvirta (1998) Modelling economic high-frequency time series with STAR-GARCH models, *Working papers in Economics and Finance* **291**, Stockholm School of Economics.

[18] Lundbergh, S., T. Teräsvirta and D. van Dijk (1999), Time-varying smooth transistion autoregressive models, Stockholm School of Economics, unpublished muniscript.

[19] Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988), Täesting linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–9.

[20] Michael, P.,A.R. Nobay and D.A. Peel (1997), Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation, *Journal of Political Economy* **105**, 862–879.

[21] Pesaran, M. H. and S. M. Potter (1997), A floor and ceiling model of US output, *Journal of Economic Dynamics and Control* **21**, 661-695.

[22] Potter, S. M. (1994) Asymmetric economic propagation mechansisms, in W. Semmler (ed.), *Business cylces: Thoery and Empirical Methods*, Boston: Kluver, pp. 527–560.

[23] Pötcher, B.M. and I.V. Prucha (1997), Dynamic Nonlinear Econometric Models-Asymptotic Theory, Berlin: Springer-Verlag

[24] Taylor, M.P., D.A. Peel, and L. Sarno (2001), Non-linear in real exchange rates: towards a solution of the purchasing power parity puzzles, Working Paper, Centre for Economic Policy Research, London, UK.

[25] Teräsvirta, T. (1994), Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

[26] Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (editors), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

[27] Teräsvirta, T., D. Tjøtheim and C.W.J. Granger (1994), Aspects of modeling nonlinear time series, in R.F. Engle and D.L. McFadden (editors), *Handbook of econometrics, vol.IV*, Amsterdam: Elsevier Science.

[28] Teräsvirta, T. and H. M. Anderson (1992), Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* **7**, S119–S136.

[29] Tong, H.(1990), Non-linear Time Series: a Dynamical Systems Approach, Oxford: Oxford University Press.

[30] van Dijk, D., T. Teräsvirta and P.H. Franses (2000), Smooth transistion autoregressive models - a survey of recent developments, SSE/EFI Working paper series in Economics and Finace No. 380, Stockholm School of Economics.

[31] van Dijk, D., P.H. Franses and A. Lucas (1999), Testing for smooth transition nonlinearity in the presence of additive outliers, *Journal of Business and Economic Statistics* **17**, 217–235.

[32] van Dijk, D., and P.H. Franses (1999), Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics* **3**, 311-40.

[33] Wooldridge, J.M. (1990), A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17–43.

[34] Wooldridge, J.M. (1991), On the application of robust, regression-based specification tests, *Journal of Econometrics* **47**, 5–46.

Table 1.1: Lag selection frequencies in $AR(p)$ model

AR Order	AIC		BIC		HQC		LB	
p	T=250	T=500	T=250	T=500	T=250	T=500	T=250	T=500
1	734	728	984	993	906	938	875	870
2	120	114	13	6	67	43	8	6
3	62	68	2	1	12	15	10	14
4	35	37	1	0	11	3	15	14
5	25	28	0	0	2	0	15	21
6	24	25	0	0	2	1	77	75

Frequencies of lag length selection in $AR(p)$ models on series generated from ESTAR model (1.2) and (1.4), with $\pi_{1,0} = \pi_{2,0} = 0$, $\pi_{1,1} = 0.6$, $\pi_{2,1} = 0.3$, $c = 0.5, u_t \sim iid\, N(0,1)$.

Table 1.2: Parameter Specifications for the generated DGPs: All of the DGPs are generated with $c = 0$ and $\gamma = 5$

DGP	Conditional mean equation						Conditional Variance		
	$\pi_{1,0}$	$\pi_{2,0}$	$\pi_{1,1}$	$\pi_{2,1}$	$\pi_{1,2}$	$\pi_{2,2}$	ω	α	β
LSTAR(1)	-0.3	0.1	-0.5	0.5	.	.	.	.	.
LSTAR(1)-GARCH(1,1)	-0.3	0.1	-0.5	0.5	.	.	1	0.3	0.6
LSTAR(2)	-0.3	0.1	-0.5	0.3	0.5	-0.3	.	.	.
LSTAR(2)-GARCH(1,1)	-0.3	0.1	-0.5	0.3	0.5	-0.3	1	0.3	0.6
AR(1)	0.5	.	0.8	.	.	.	.	.	.
AR(1)-GARCH(1,1)	0.5	.	0.8	.	.	.	1	0.3	0.6
AR(2)	0.5	.	0.8	.	-0.4	.	.	.	
AR(2)-GARCH(1,1)	0.5	.	0.8	.	-0.4	.	1	0.3	0.6

Table 1.3: Empirical power of the linearity tests.

DGP	LS			HCC		
	LM_2	LM_3	LM_4	LM_2	LM_3	LM_4
Sample size: T=100						
STAR(1)	0.26	0.23	0.20	0.19	0.15	0.12
STAR(1)-G(1,1)	0.22	0.20	0.19	0.16	0.14	0.10
STAR(2)	0.56	0.50	0.45	0.39	0.33	0.25
STAR(2)-G(1,1)	0.62	0.57	0.53	0.44	0.39	0.31
Sample size: T=300						
STAR(1)	0.65	0.62	0.57	0.61	0.57	0.50
STAR(1)-G(1,1)	0.62	0.59	0.55	0.57	0.52	0.46
STAR(2)	0.98	0.99	0.99	0.96	0.98	0.94
STAR(2)-G(1,1)	1.00	0.99	0.99	0.98	0.98	0.97
Sample size: T=500						
STAR(1)	0.88	0.87	0.83	0.87	0.84	0.79
STAR(1)-G(1,1)	0.86	0.83	0.80	0.83	0.79	0.75
STAR(2)	0.99	1.00	0.99	0.98	0.99	0.97
STAR(2)-G(1,1)	1.00	1.00	1.00	1.00	1.00	1.00
Sample size: T=1000						
STAR(1)	0.99	0.99	0.99	1.00	0.99	0.99
STARE(1)-G(1,1)	1.00	1.00	1.00	1.00	1.00	0.99
STAR(2)	1.00	1.00	1.00	1.00	1.00	1.00
STAR(2)-G(1,1)	1.00	1.00	1.00	1.00	1.00	1.00

Note: **The LS stands for the standard least squares based versions of the LM-type tests, HCC refers to the Wooldridge version of the unknown heteroscedasticity consistent version of the tests. The empirical powers are computed at 5% significance level. The transition variable used in the linearity tests is y_{t-1}**

Table 1.4: Empirical size of the linearity test.

Sample size: T=300

DGP	LS			HCC		
	LM_2	LM_3	LM_4	LM_2	LM_3	LM_4
AR(1)	.044	.042	.043	.037	.037	.028
AR(1)-G(1,1)	.043	.036	.039	.044	.029	.028
AR(2)	.048	.034	.034	.039	.033	.022
AR(2)-G(1,1)	.048	.048	.044	.037	.029	.021
hline						
Sample size: T=500						
AR(1)	.049	.043	.037	.048	.039	.030
AR(1)-G(1,1)	.052	.041	.037	.051	.036	.028
AR(2)	.051	.045	.053	.050	.040	.042
AR(2)-G(1,1)	.053	.045	.045	.040	.028	.025
Sample size: T=1000						
AR(1)	.045	.040	.046	.048	.044	.041
AR(1)-G(1,1)	.052	.044	.044	.049	.048	.037
AR(2)	.056	.053	.050	.055	.048	.045
AR(2)-G(1,1)	.057	.056	.055	.050	.037	.035

Note: Each cell represents the proportion of rejections of the true null hypothesis of linearity at 5% significance level. LS columns give the standard least squares based tests and HCC columns give the Wooldridge type heteroscedasticity consistent versions of the tests. The transistion variable used in the linearity tests is y_{t-1}.

Table 1.5: Simulation Results on the finite sample performance of NLE of STAR models

Parm.	Mean Est	Mean S.E.	RMSE	BIAS	Skewness	Kurtosis
T=100						
π	1.043	1.143	1.717	0.430	1.370	5.360
π^*	0.850	0.322	0.172	-0.051	1.010	1.071
γ	4.605	1.981	6.651	3.605	2.144	5.329
T=300						
π	0.964	0.522	0.795	-0.036	1.376	3.274
π^*	0.885	0.188	0.092	-0.015	1.014	1.038
γ	3.657	1.900	5.123	2.657	2.313	5.091
T=500						
π	1.008	0.425	0.631	0.008	1.308	2.705
π^*	0.888	0.164	0.078	-0.012	1.008	1.020
γ	3.100	1.785	5.100	2.100	2.270	4.950

Key: Mean and RMSE, Bias, skewness and the kurtosis of NLS estimates of the parameters in the ESTAR model, with $\pi_1 = 1$, $\pi_1^* = 0.9$, $\gamma = 1$, $c = 0$ and $u_t \sim i.i.d.N(0, 1)$. The table is based on 2000 replications.

Table 1.6: Simulation Results on the finite sample performance of NLSE of STAR models

Parm.	Mean Est	Mean S.E.	RMSE	BIAS	Skewness	Kurtosis
T=100						
π	1.081	1.000	1.704	0.081	1.611	6.406
π^*	0.852	0.269	0.166	-0.048	1.005	1.044
γ	4.383	2.050	5.441	-0.617	2.138	5.246
T=300						
π	1.021	0.500	0.790	0.021	1.116	3.023
π^*	0.878	0.161	0.115	-0.022	1.006	1.022
γ	4.830	1.800	4.900	-0.170	2.036	4.850
T=500						
π	0.994	0.406	0.590	-0.006	1.039	2.636
π^*	0.883	0.160	0.106	-0.017	1.008	1.015
γ	4.885	1.650	4.225	-0.115	2.016	4.550

Mean and RMSE, Bias, skewness and the kurtosis of NLS estimates of the parameters in the ESTAR model, with $\pi_1 = 1$, $\pi_1^* = 0.9$, $\gamma = 5$, $c = 0$ and $u_t \sim i.i.d.N(0,1)$. The table is based on 2000 replications.

Table 1.7: Simulation Results on the finite sample performance of NLSE of STAR models

Parm.	Mean Est	Mean S.E.	RMSE	BIAS	Skewness	Kurtosis
T=100						
π	1.028	1.053	1.611	0.028	1.740	7.025
π^*	0.846	0.396	0.180	-0.054	1.015	1.057
γ	4.086	2.294	12.065	-10.914	2.258	5.958
T=300						
π	1.007	0.673	1.090	0.007	1.060	3.994
π^*	0.883	0.146	0.118	-0.017	1.009	1.022
γ	8.874	2.078	9.900	-6.126	2.006	4.395
T=500						
π	1.005	0.465	0.790	0.005	1.004	3.676
π^*	0.885	0.108	0.106	-0.015	1.008	1.011
γ	10.389	2.005	7.151	-4.611	2.120	4.255

Mean and RMSE, Bias, skewness and the kurtosis of NLS estimates of the parameters in the ESTAR model, with $\pi_1 = 1$, $\pi_1^* = 0.9$, $\gamma = 15$, $c = 0$ and $u_t \sim i.i.d.N(0,1)$. The table is based on 2000 replications.

Table 1.8: Simulation Results on the finite sample performance of NLSE of STAR models

Parm.	Mean Est	Mean S.E.	RMSE	BIAS	Skewness	Kurtosis
T=100						
π	0.960	0.439	1.527	-0.040	3.433	9.159
π^*	0.815	0.219	0.213	-0.085	1.022	1.073
γ	6.948	1.365	8.767	5.948	1.694	3.224
c	0.203	0.366	2.357	-0.297	0.142	2.524
T=300						
π	0.934	0.242	0.807	-0.066	1.851	4.186
π^*	0.878	0.210	0.157	-0.022	0.996	1.031
γ	5.875	1.335	7.168	4.875	1.249	2.933
c	0.440	0.326	2.119	-0.060	0.368	2.023
T=500						
π	0.975	0.171	0.545	-0.025	1.984	4.019
π^*	0.887	0.071	0.067	-0.013	0.771	1.055
γ	4.099	1.206	6.951	-3.099	1.118	3.349
c	0.515	0.289	1.847	-0.015	0.040	2.689

Mean and RMSE, Bias, skewness and the kurtosis of NLS estimates of the parameters in the ESTAR model, with $\pi_1 = 1$, $\pi_1^* = 0.9$, $\gamma = 1$, $c = 0.5$ and $u_t \sim i.i.d.N(0,1)$. The table is based on 2000 replications.

Table 1.9: Simulation Results on the finite sample performance of NLSE of STAR models

Parm.	Mean Est	Mean S.E.	RMSE	BIAS	Skewness	Kurtosis
T=100						
π	0.913	1.769	3.229	0.113	1.931	9.199
π^*	0.379	0.657	0.278	-0.021	1.056	4.117
γ	7.811	2.343	11.077	2.811	3.029	10.697
T=300						
π	0.901	0.866	1.944	0.101	2.004	5.082
π^*	0.393	0.442	0.202	-0.007	1.071	3.419
γ	6.611	2.176	7.783	1.611	2.902	6.179
T=500						
π	0.881	0.822	1.299	0.081	1.638	4.236
π^*	0.395	0.330	0.133	-0.013	1.016	1.686
γ	5.991	2.110	6.817	0.991	2.771	5.353

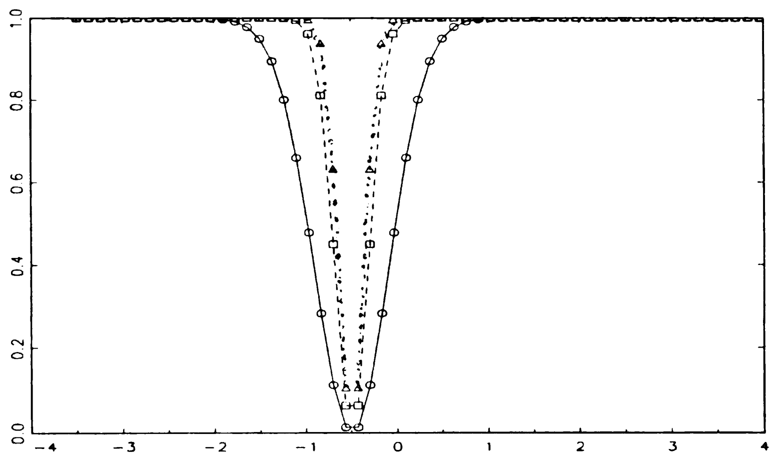
Mean and RMSE, Bias, skewness and the kurtosis of NLS estimates of the parameters in the ESTAR model, with $\pi_1 = 0.8$, $\pi_1^* = 0.4$, $\gamma = 5$, $c = 0$ and $u_t \sim i.i.d.N(0,1)$. The table is based on 2000 replications.

Table 1.10: Simulation Results on the finite sample performance of NLSE of STAR models

Parm.	Mean Est	Mean S.E.	RMSE	BIAS	Skewness	Kurtosis
T=100						
π	0.853	1.778	3.190	0.353	2.177	12.876
π^*	-0.480	0.441	0.223	-0.020	-0.803	2.323
γ	8.293	2.065	12.933	3.293	3.432	14.816
T=300						
π	0.724	1.121	1.800	0.224	2.094	6.876
π^*	-0.507	0.225	0.167	-0.007	-1.049	2.146
γ	6.684	1.175	8.286	1.684	3.174	7.559
T=500						
π	0.625	0.976	1.447	0.125	2.674	4.190
π^*	-0.504	0.215	0.112	-0.004	-0.509	1.726
γ	6.097	1.634	7.064	1.097	2.578	6.532

Mean and RMSE, Bias, skewness and the kurtosis of NLS estimates of the parameters in the ESTAR model, with $\pi_1 = 0.5$, $\pi_1^* = -0.5$, $\gamma = 5$, $c = 0$ and $u_t \sim i.i.d.N(0,1)$. The table is based on 2000 replications.

Figure 1.1: Examples of the exponential, logistic, functions for values of γ 3, 5, and 25 and threshold parameter $c = 0$.
a. Exponential



b. Logistic

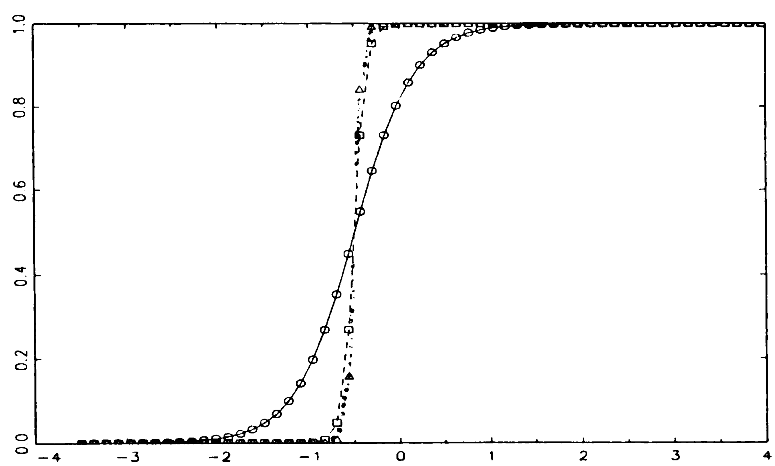
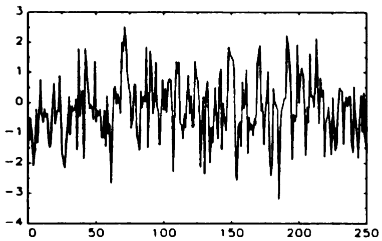
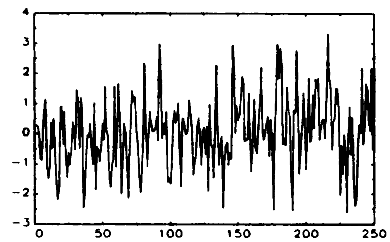


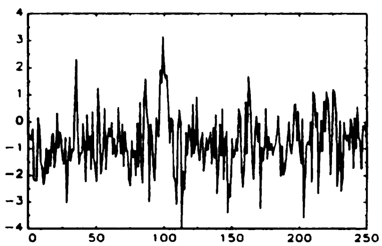
Figure 1.2: Sample realizations from the STAR models $\pi_{1,1} = -.3, \pi_{1,2} = 0.7, c = 0$ and $u_t \sim NID(0,1)$

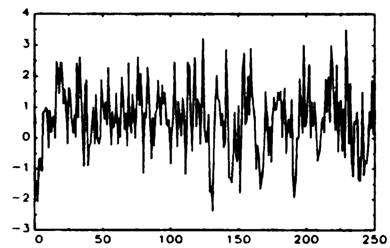
(a) $\pi_{1,0} = -0.5, \pi_{2,0} = 0.5$ (b)$\pi_{1,0} = 0.5, \pi_{2,0} = -0.5$



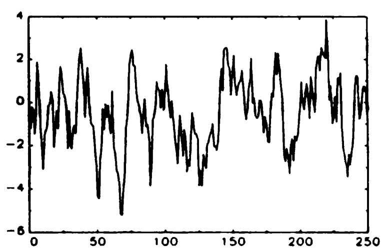


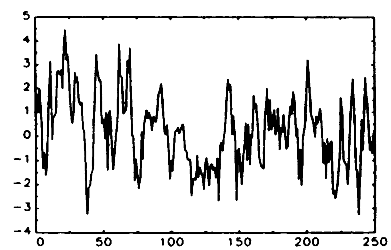
(c) $\pi_{1,0} = -1.5, \pi_{2,0} = 1.5$ (d)$\pi_{1,0} = 1.5, \pi_{2,0} = -1.5$





(e)$\pi_{1,0} = \pi_{2,0} = 0, \pi_{1,1} = 1, \pi_{2,1} = -0.3$ (f)$\pi_{1,0} = \pi_{2,0}, \pi_{1,1} = 1, \pi_{2,1} = -0.3$





Notes: The figures in 2a-2e are sample realizations from ESTAR model with the given parameter specifications, while figure in 2f is a sample realization from LSTAR model with quadratic logistic function given in (1.5) with the same parameter specification as in 2e, except, thresholds are specified to be $c_1 = 0, c_2 = 0.5$

CHAPTER 2

Review of long memory models for conditional mean and variance

2.1 Introduction: Definition and sources of long memory in economic time series

This chapter briefly discusses the properties of long memory process with particular attention given to fractionally integrated processes. Surveys of long memory processes, their statistical properties and applications in economics, finance and some other fields can be found in Baillie (1996), and Beran (1994).

Traditionally, long memory has been defined in the time domain in terms of decay rates of long-lag autocorrelations, or in the frequency domain in terms of rates of explosion of low-frequency spectra. A process with the long-lag autocorrelation function given by,

$$\gamma_k = c_\gamma k^{2d-1} \text{as } k \to \infty \tag{2.1}$$

is called a long memory process. The definition in (2.1) implies the following condition,

$$\lim_{T\to\infty} \sum_{j=-T}^{T} |\rho_j| = \infty. \tag{2.2}$$

That is, for a discrete time series, autocorrelation function, ρ_j is not absolutely summable. See for instance, McLeod and Hipel (1978).

In the spectral domain a long memory process is defined in terms of the behavior of the spectral density at low frequencies. A process is called long memory if the spectral density, $f_y(\omega) = c_f\omega^{-2d}$ as $\omega \to 0^+$. A more general definition, provided by Heyde and Yang (1997) in the frequency domain is simply, $f(\omega) = \infty$ as $\omega \to 0^+$. Note that the constants, c_γ and c_f can be replaced by so-called slowly varying functions, i.e., functions such that for any $t \in R$, $L(ty)/L(y) \to 1$ as $y \to \infty$ or $y \to 0$. Since knowing the covariances (or correlations and variance) is equivalent to knowing the spectral density, the long-lag autocorrelation definition in the time domain and low-frequency spectral definitions are equivalent under the conditions given, for example in Beran (1994, pp. 42-44).

A third definition of long memory involves the rate of growth of variances of partial sums,

$$S_T = \sum_{t=1}^{T} y_t.$$

A process is said to be a long memory process if $var(S_T) = O(T^{2d+1})$ for $d > 0$. In other words, a process is a long memory process if the growth rate of variances of its partial sums are in the order of T^{2d+1}. There is a connection between the variance-of-partial-sum definition of long memory and the spectral definition of long memory (and hence also the autocorrelation definition of long memory). In particular, because the spectral density at frequency zero is the limit of $\frac{1}{T}S_T$, a process has long memory in the generalized spectral sense of Heyde and Yang if and only if it has long memory for some $d > 0$ in the variance-of-partial-sum sense. Therefore, the variance-of-partial-sum definition of long memory is quite general.

It should be emphasized that these definitions are asymptotic in the sense that they characterize the ultimate behavior of the correlations, and variance of partial sums as lags and/or sample size approaches infinity. In general they do not specify the correlations and/or variance of the partial sums for any fixed finite lag and/or for any fixed finite sample size. In particular, both correlation definition and the spectral density definitions do not determine the absolute size of the correlations. In other words, each individual correlations can be arbitrarily small while the decay rate of correlations is slow.

There is a natural desire to understand the nature of various mechanisms that could generate long memory. Most econometric attention has focused on the role of aggregation. Granger (1980) considered the aggregation of $i = 1, \cdots, N$ cross-sectional components, $y_{i,t} = \alpha_i y_{i,t} + \epsilon_{i,t}$, where $\epsilon_{i,t}$ is white noise, and it is also assumed that for $i \neq j$ $\epsilon_{i,t}$ is independent of $\epsilon_{j,t}$ and α_i is also independent of $\epsilon_{j,t}$ for all i, j, t. As $N \to \infty$, it is shown in Granger (1980) that the spectrum of the aggregated process, $y_t = \sum_{i=1}^{N} y_{i,t}$ is approximately given by,

$$f_y = \frac{N}{2\pi} E[var(\epsilon_{i,t})] \int \frac{1}{|1 - \alpha \exp i\omega|^2} dF(\alpha),$$

where $F(\alpha) = \int_0^\alpha \frac{t^{p-1}(1-t)^{b-1}}{B(p,b)} dt$, is the cumulative density function governing the α_i's. Here, $B(p,b) = \int_0^1 \alpha^{p-1}(1-\alpha)^{b-1} d\alpha = \frac{\Gamma(p)\Gamma(b)}{\Gamma(p+b)}$, is the beta function, and p, $b > 0$. Upon assuming that α_i's are distributed as a Beta distribution with parameters (p, b),

$$dF(\alpha) = \frac{2}{B(p,b)} \alpha^{2p-1}(1-\alpha^2)^{b-1} d\alpha, \ 0 \leq \alpha 1,$$

then the kth autocovariance of y_t is

$$\gamma_y(k) = \frac{2}{B(p,b)} \int_0^1 \alpha^{2p+k-1}(1-\alpha^2)^{b-2} d\alpha = Ck^{1-b}.$$

Thus Granger (1980) shows that the aggregated series, y_t, is a long memory process in the sense that it is integrated of order $(1 - \frac{b}{2})$.

Recently, Lippi and Zaffaroni (1999) generalized Granger's result by replacing Granger's assumed beta distribution with weaker semi-parametric assumptions and obtained similar results. Chambers (1998) considers temporal aggregation in addition to cross sectional aggregation in both discrete and continuous time as the source of long memory.

An alternative source of long memory, which also involves aggregation, has been studied by Ciozek-Georges and Mandelbrot (1995), Taqqu, Willinger and Sherman

(1997), and Parke (1999). This source of long memory involves the distribution of the duration between consecutive events. In particular, the idea is based on the modelling of aggregate traffic computer networks. For illustration, consider the stationary continuous time binary series $S(t)$, $t \geq 0$ such that $S(t) = 1$ during "on" periods and $S(t) = 0$ during "off" periods. The lengths of the on and off periods are assumed to be independently and identically distributed (i.i.d) at all leads and lags. It is also assumed that on and off periods alternate. Under these assumptions, consider M sources, $S^m(t), t \geq 0$, $m = 1, \cdots, M$, and define the aggregate count in the interval $[0, tT]$ by

$$S_M(tT) = \int_0^{tT} (\sum_{m=1}^{M} S^m(v))dv.$$

Let $F_1(y)$ denote the c.d.f. of durations of on periods, and $F_2(y)$ be the c.d.f. of durations of off periods, and further assume the following for the tail of the distributions of on and off durations,

$$1 - F_1(y) \sim C_1 y^{-\alpha_1} L_1(y), \text{with } 1 < \alpha_1 < 2,$$

$$1 - F_2(y) \sim C_2 y^{-\alpha_2} L_2(y), \text{with } 1 < \alpha_2 < 2.$$

Thus the power-law tails imply infinite variance for the on and off durations. By letting first $M \to \infty$ and then $T \to \infty$ Cioczek-Georges and Mandelbrot (1995) and Taqqu et al. (1997) show that $S_M(tT)$ after being appropriately standardized, converges to a fractional Brownian motion. The regular Brownian motion, $B(r)$, is a continuous time stochastic process whose increments are independent Gaussian distributed. The fractional Brownian motion, $B_d(r)$ is regarded as the approximate $(-d)$ fractional derivative of regular Brownian motion, $B_d(r) = \frac{1}{\Gamma(d+1)} \int_0^r (r-y)^d dB(y)$. See Beran (1994) for details. Hence, the aggregate counts in the interval $[0, tT]$ is a long memory process.

Parke (1999) considers a closely related discrete-time error duration model. In particular, he assumes that the aggregate process, y_t is being generated by the following sum, $y_t = \sum_{s=-\infty}^{t} g_{s,t}u_s$, where $u_t \sim i.i.d.(0, \sigma^2)$, and $g_{s,t} = 1(t \leq s + n_s)$, where $1(.)$ is the indicator function, and n_s is the stochastic duration between consecutive errors. Assuming a probability law for the distribution of n_s that implies infinite variance for the durations, similar to above, leads y_t to be long memory.

An alternative route, that may lead to long memory, explored by Diebold and Inoue (2001), involves structural change or stochastic regime switching. They show how some simple stochastic regime switching models may produce realizations that appear to have long memory under conditions that ensure that as sample size increases the realizations tend to have just a few breaks. For illustration purposes consider the following mixture model,

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + v_t$$

$$v_t = \begin{cases} 0 & \text{w.p. } 1-p \\ w_t & \text{w.p. } p \end{cases}$$

where $w_t \sim iidN(0, \sigma_w^2)$ and $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$. They show that under the assumption that $p = O(T^{2d-2})$, $0 < d < 1$, y_t will be an $I(d)$ (integrated of order d) process. Diebold and Inoue (2001) show several other stochastic models under certain conditions (mostly assumptions that dictate how certain parameters, such as mixture probabilities vary with T) can generate realizations with long memory. Their theoretical results indicate that regime switching (structural change) and long memory are easily confused when only a small number of regime switches/breaks occurs. Guided by their theoretical results, they conduct extensive Monte Carlo analysis to verify how in finite samples with fixed-parameter stochastic regime switching models whose dynamics is either $I(0)$ or $I(1)$ one can obtain realizations that have long memory dynamics. Diebold and Inoue (2001) conjecture that threshold autoregressive (TAR), smooth transition autoregressive models (STAR) may have realizations with long memory once one allows thresholds to change appropriately with sample size.

2.2 Long Memory Models

This section discusses parametric models that are capable of capturing long memory phenomena in both the conditional mean and the conditional variance of a univariate series. In particular, the fractionally integrated autoregressive moving average (ARFIMA) model, developed by Granger and Joyeux (1980), Granger (1980), and Hosking (1981) for the conditional mean of a time series, and fractionally integrated autoregressive conditional heteroscedastic (FIGARCH) model due to Baillie etal. (1996) will be reviewed in terms of representation, specification, estimation, and inference.

2.2.1 The ARFIMA Model

Integrated autoregressive moving average (ARIMA) models were introduced by Box and Jenkins (1970). The theory of statistical inference for the ARIMA models is well developed, see for instance, Brockwell and Davis (1997), and Hamilton (1994). ARFIMA models are natural extensions of the ARIMA models. Therefore, let us first recall the definition of ARMA and ARIMA processes. To simplify the notation assume that $E(y_t) = \mu = 0$. Otherwise, y_t needs to be replaced by $y_t - \mu$ in the following formulas. First define the polynomials,

$$\phi(x) = 1 - \sum_{i=1}^{p} \phi_i x^i$$

$$\theta(x) = 1 + \sum_{i=1}^{q} \theta_i x^i$$

where p and q are integers. Assuming that all the solutions of polynomial equations, $\phi(x) = 0$ and $\theta(x) = 0$ are outside the unit circle, an $ARMA(p, q)$ model is defined to be the stationary solution of

$$\phi(L)y_t = \theta(L)u_t, \tag{2.3}$$

where L is the lag operator, and disturbances, u_t are usually assumed to have zero mean, $E(u_t = 0)$, and finite variance, $E(u_t^2) = \sigma_u^2$ and are serially uncorrelated, $E(u_t u_s) = 0$ for $t \neq s$. If equation (2.1) holds true for the dth difference $(1 - L)^d y_t$, then y_t is called an $ARIMA(p, d, q)$ process with the corresponding equation, now given by

$$\phi(L)(1 - L)^d y_t = \theta(L)u_t. \tag{2.4}$$

Note that $ARMA(p, q)$ model is encompassed by the $ARIMA(p, d, q)$ model in the sense that $ARMA(p, q)$ model is obtained from $ARIMA(p, d, q)$ model by letting $d = 0$. If $d \geq 1$, then the original series y_t is not stationary and hence to obtain a stationary process y_t needs to be differenced d times. Generalization of (2.4) to non-integer values of d gives the ARFIMA(p,d,q) model. Note that if d is an integer $(d \geq 0)$, then $(1 - L)^d$can be written as

$$(1 - L)^d = \sum_{i=0}^{d} \begin{pmatrix} d \\ k \end{pmatrix} (-1)^k L^k,$$

with the binomial coefficients

$$\begin{pmatrix} d \\ k \end{pmatrix} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)},$$

where $\Gamma(.)$ denotes the gamma function and is defined by

$$\Gamma(s) = \int_0^\infty \exp(-x) x^{s-1} dx.$$

Since the gamma function is defined for all real numbers, the binomial coefficients can be extended to all real numbers d. For any real number d, $(1-L)^d$i s defined by

$$(1-L)^d = \sum_{k=0}^{\infty} \begin{pmatrix} d \\ k \end{pmatrix} (-1)^k L^k = 1 - dL - \frac{d(1-d)L^2}{2!} - \cdots \tag{2.5}$$

$$= F(-d,1,1;L) = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} L^k,$$

where F stands for the hypergeometric function which is defined formally by

$$F(m,n,s;x) = \frac{\Gamma(s)}{\Gamma(m)\Gamma(n)} \sum_{j=0}^{\infty} \frac{\Gamma(m+j)\Gamma(n+j)}{\Gamma(s+j)\Gamma(j+1)}.$$

For all positive integers only the first $d+1$ terms are nonzero and hence, for positive integer d (2.6) is the usual dth difference operator while for non-integer d, the summation in (2.6) is genuinely over an infinite number of indices. Given (2.6) Granger and Joyeux (1980) and Hosking (1981) proposed the following definition for the ARFIMA model:

Definition 2.1 *Let y_t be a stationary process such that*

$$\phi(L)(1-L)^d y_t = \theta(L)u_t \tag{2.6}$$

for some $-\frac{1}{2} < d < \frac{1}{2}$. Then y_t is called an $ARFIMA(p,d,q)$ process.

The range that makes the $ARFIMA(p,d,q)$ process in (2.6) long memory is $0 \leq d < \frac{1}{2}$. The upper bound $d < \frac{1}{2}$ makes the process covariance stationary. For $d \geq \frac{1}{2}$ the $ARFIMA(p,d,q)$ process is not covariance stationary. In particular, the usual definition of the spectral density of y_t would lead to a non-integrable function. Whenever d falls in $[\frac{1}{2},1)$ then the process is considered to be covariance non-stationary. Moreover, the $ARFIMA(p,d,q)$ process given in (2.6) is invertible for values of $d > -\frac{1}{2}$ and have an infinite order autoregressive representation. For the range $-\frac{1}{2} < d < \frac{1}{2}$ the $ARFIMA(p,d,q)$ process is invertible and stationary and can be represented by both as an infinite order autoregressive or infinite order moving average process. These representations for the general $ARFIMA(p,d,q)$ are given in Sowell (1992). They are complicated functions of hypergeometric function. For $p=q=0$ the $ARFIMA(0,d,0)$ process is also called fractional white noise, see Baillie (1996). This is because a random walk is the discrete analog of the Brownian motion and similarly the discrete time version of fractional Brownian motion is the fractionally differenced white noise. Note that $ARFIMA(0,d,0)$ process is given by

$$(1-L)^d y_t = u_t. \tag{2.7}$$

In this case, the infinite order autoregressive and moving average representations are easy to obtain from (2.7) as shown in Hosking (1981). In particular, the infinite order autoregressive representation is,

$$y_t = \sum_{k=0}^{\infty} \pi_k y_{t-k} + u_t, \tag{2.8}$$

where the infinite order autoregressive weights are given in (2.6) and for $k \to \infty$,

$$\pi_k \sim \frac{1}{\Gamma(-d)} k^{-d-1}. \tag{2.9}$$

The infinite order moving average representation is obtained by use of the Wold decomposition, and given by,

$$\begin{aligned} y_t &= (1-L)^{-d} u_t = \sum_{k=0}^{\infty} \psi_k u_{t-k} \\ &= [1 + dL + \frac{d(d+1)L^2}{2!} + \frac{d(d+1)(d+2)L^3}{3!} + \cdots] u_t \end{aligned} \tag{2.10}$$

The infinite order moving average coefficients alternatively can be expressed by use of the gamma function. Since, $\Gamma(d+k) = \frac{d(d+1)(d+2)\cdots(d+k-1)}{\Gamma(d)}$ it follows that $\psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}$. When $k \to \infty$, the infinite order moving coefficients will be approximately equal to,

$$\psi_k \sim \frac{1}{\Gamma(d)} k^{d-1}. \tag{2.11}$$

Equation (2.6) can be interpreted in several ways. For instance, defining, $\tilde{y}_t = \phi^{-1}(L)\theta(L)u_t$, it can be written as

$$(1-L)^d y_t = \tilde{y}_t.$$

This representation means that an ARMA process is obtained after passing y_t through the fractional difference operator (or infinite linear filter) $(1-L)^d$. Alternatively, (2.6) can be written as

$$y_t = \phi(L)^{-1}\theta(L)y_t^*,$$

where y_t^* is an $ARFIMA(0,d,0)$ process defined in (2.7). In this representation, y_t is obtained by passing an $ARFIMA(0,d,0)$ process through an ARMA filter. Figures 1.a to 1.d show sample realizations of several ARFIMA processes with disturbances $u_t \sim iidN(0, 0.25)$ and the same long memory parameter $d = 0.3$. It is apparent from these graphs that many different types of dynamic behavior can be obtained. Figures 2.a to 2.d show the first fifty autocorrelations of the corresponding processes together with the 95 percent confidence intervals. As is evident from the figures the sample realizations are quite persistent in their autocorrelations in that there are very significant correlations in higher lags. The parameter d determines the long run behavior of the process while autoregressive and moving average parameters allow one to model short-run dynamics more flexibly. In this sense, $ARFIMA$ models are very flexible and parsimonious as they allow one to model both short run and long run behavior of a time series simultaneously.

The spectral density of an $ARFIMA$ process can be obtained directly from (2.6). Note that the spectral density of an $ARMA$ process, $\tilde{y}_t$ is given by;

$$f_{\tilde{y}}(\omega) = \frac{\sigma_u^2}{2\pi}\frac{|\theta(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2},$$

where ω is the angular frequency. Since the $ARFIMA$ process is obtained from a process $\tilde{y}_t$ with spectral density, $f_{\tilde{y}}$ by applying the infinite linear filter, $\sum_{k=0}^{\infty}\psi_k\tilde{y}_{t-k}$, then by a result from Priestley (1981, pp.243-66), the spectral density of y_t is equal to $|\Lambda(\omega)|^2 f_{\tilde{y}}(\omega)$, where $\Lambda(\omega) = \sum_{k=0}^{\infty} \psi_k e^{ik\omega}$. Hence, it follows from (2.6) that the spectral density of y_t will be;

$$f_y(\omega) = |1 - e^{i\omega}|^{-2d} f_{\tilde{y}}(\omega), \tag{2.12}$$

where $|1 - e^{i\omega}| = 2\sin(\frac{1}{2}\omega)$. Since, $\lim_{\omega \to 0} \omega^{-1}(\sin(\frac{1}{2}\omega) = 1$, the behavior of the spectral density of the process at low frequencies (alternatively, at high periods, or as sample size approaches infinity) will be given by

$$f_y(\omega) \sim \frac{\sigma_u^2}{2\pi} f_{\tilde{y}(0)} = \frac{\sigma_u}{2\pi} \frac{|\theta(1)|^2}{|\phi(1)|^2} |\omega|^{-2d}. \tag{2.13}$$

For $-\frac{1}{2} < d < 0$, $f_y(0) = 0$, and hence the sum off all autocorrelations is zero. For $d = 0$, spectral density reduces to that of an ordinary $ARMA(p, q)$ process with bounded spectral density. Long-range dependence, and/or long memory occurs when $0 < d < \frac{1}{2}$. To transform y_t into a process with bounded spectral density, the infinite linear filter, $(1 - L)^d$ needs to be applied.

Obtaining explicit expressions for all covariances for the $ARFIMA(p, d, q)$ process is relatively difficult, except in the case of $ARFIMA(0, d, 0)$ process. In this case, it is shown in Sowell (1992) that the covariances are given by the formula;

$$\gamma_k = \sigma_u^2 \frac{(-1)^k \Gamma(1 - 2d)}{\Gamma(k - d + 1)\Gamma(1 - k - d)} \tag{2.14}$$

The autocorrelations are given by,

$$\rho_k = \frac{\Gamma(1 - d)\Gamma(k + d)}{\Gamma(d)\Gamma(k + 1 - d)}. \tag{2.15}$$

By using the approximation, $\frac{\Gamma(k+d)}{\Gamma(k+1-d)} \approx k^{2d-1}$ for large k,ρ_k can be expressed asymptotically by

$$\rho_k \sim \frac{\Gamma(1 - d)}{\Gamma(d)} k^{2d-1} \text{ as } (k \to \infty) \tag{2.16}$$

To obtain the covariances of the general $ARFIMA(p, d, q)$ process as suggested in Beran (1994) one can use the covariances of the $ARFIMA(0, d, 0)$ process. This can be done by first recalling that y_t is obtained by passing an $ARFIMA(0,d,0)$ process, y_t^* through the linear filter,

$$\lambda(L) = \phi(L)\theta^{-1}(L) = \sum_{i=0}^{\infty} \lambda_i L^i.$$

Denoting the covariances of y_t^*, in the first step, calculate the coefficients λ_i by matching the powers of $\phi(L)\theta^{-1}(L)$ with those of $\lambda(L)$. In the second step the covariances of $ARFIMA(p,d,q)$ process, y_t are obtained from $\lambda(L)$ and the covariances γ_k^* by

$$\gamma_k = \sum_{i,l=0}^{\infty} \lambda_i \lambda_l \gamma_{k+i-l}^*. \tag{2.17}$$

See Chung (1994) for alternative derivation of autocorrelations of $ARFIMA(p,d,q)$ model. The asymptotic formulas for the covariances and autocorrelations are:

$$\gamma_k \sim C_{\gamma)}(d,\phi,\theta)|k|^{2d-1} \tag{2.18}$$

where

$$C_\gamma(d,\phi,\theta) = \frac{\sigma_u^2}{\pi}\frac{|\theta(1)|^2}{|\phi(1)|^2}\Gamma(1-2d)\sin d\pi.$$

and

$$\rho_k \sim C_\rho(d,\phi,\theta)|k|^{2d-1} \tag{2.19}$$

where

$$C_\rho(d,\phi,\theta) = \frac{C_\gamma(d,\phi,\theta)}{\int_{-\pi}^{\pi} f(\omega)d\omega}$$

2.3 Long memory volatility models

Risk is an important factor in financial markets. At a theoretical level, the Capital Asset Pricing Model (CAPM) developed by Sharpe (1964) and Merton (1973) indicates presence of a direct relationship between return and risk of an asset. Also an important determinant of an option is the risk associated with the price of the underlying asset, as measured by its volatility. One of the stylized facts of asset returns in financial markets is that volatilities of assets change over time. Periods of large price changes are followed by periods of relatively stable prices. This property of asset prices is referred to in the literature as *volatility clustering.* The time varying nature of the volatility was recognized early in 1960s, see for instance, Mandelbrot (1963a, 1963b) and Fama (1965). Econometric modelling of the volatility clustering phenomenon occurred relatively recently in 1980s. The Autoregressive Conditional Heteroscedasticity (ARCH) model introduced first by Engle (1982) and modified by Bollerslev (1986) and labelled as Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models and their extensions have become popular both among practitioners and researchers. GARCH models are able to describe certain properties of economic time series, such as volatility clustering and excess kurtosis. Although the GARCH model is able to capture the volatility clustering phenomenon well it is not able to capture certain other empirically relevant properties of financial time series. For instance, in the standard GARCH model the effect of a shock on volatility depends only on the shocks' size not sign. However, as observed in Black (1976) negative shocks or news may affect the volatility quite differently than positive ones. Hence, the sign of the shock may be relevant in understanding the dynamic nature of the volatility. Another example constitutes the persistence of the effects of shocks in the volatility process. As observed in Ding, Granger, and Engle (1993) sample autocorrelations of certain volatility measures, such as absolute and squared returns, decline at a hyperbolic rate. Standard GARCH models fail to account for this slow decay in the autocorrelations which is inherent in the volatility process. These considerations led several researchers to develop volatility models that are capable of modelling several aspects of volatility in financial markets. In this section, we will review GARCH class of models with particular attention given to parametric long memory volatility model of Baillie et al. (1996), namely the fractionally integrated GARCH, (FIGARCH) model.

In general, an observed time series y_t can be written as the sum of a predictable and an unpredictable component,

$$y_t = E[y_t \,|\Omega_{\,t-1}] + u_t, \tag{2.20}$$

where Ω_{t-1} is the information set consisting of all relevant information up to and including time $t-$ 1. In the previous section, different specifications (such as $ARIMA(p,q)$, or $ARFIMA(p,d,q)$ for the predictable or conditional mean $E[y_t|\Omega_{t-1}]$ have been discussed. In section 2.2, the unpredictable part or disturbance u_t is assumed to satisfy the white noise properties. In particular, it was assumed that u_t is both conditionally and unconditionally homoscedastic, that is, $E[u_t^2] = E[u_t^2|\Omega_{t-1}] = \sigma_u^2$ for all t. In the ARCH modelling of volatility, this assumption is relaxed, and replaced by the assumption that *the conditional variance of u_t can vary over time*, that is, $E[u_t^2|\Omega_{t-1}] = h_t$ for some nonnegative function $h_t \equiv h_t(\Omega_{t-1})$. Hence, the disturbances are conditionally heteroscedastic. Following Engle (1982), a convenient functional form is

$$u_t = z_t\sqrt{h_t} \tag{2.21}$$

where z_t independent and identically distributed with zero mean and unit variance. For convenience, it is usually assumed that z_t has a standard normal distribution. This latter assumption can be replaced with another distributional assumption, for example, following Bollerslev (1987) one may assume that z_t follows a student-t distribution with ν degrees of freedom. From (2.21) and the properties of z_t it follows that the distribution of u_t conditional upon the history Ω_{t-1} is either normal or student-t with mean zero and variance h_t. The unconditional variance of u_t is,

$$\sigma_u^2 \equiv E[u_t^2] = E[E[u_t^2|\Omega_{t-1}]] = E[h_t], \tag{2.22}$$

where the latter equality follows from the law of iterated expectations, assuming that the expectations exist. It follows that the unconditional variance of u_t should be constant, that is, the unconditional mean, $E[h_t] = constant$. Equations (2.21-2.22) specify the general representation of GARCH type of models. The complete specification involves how one assumes the conditional variance of u_t evolves over time. GARCH type models specify the conditional variance of u_t as such the specified model captures (some) of the empirically observed facts of the economic and financial time series.

2.3.1 The (G)ARCH Model

Engle (1982) introduced the class of Autoregressive Conditionally heteroscedastic (ARCH) models to capture the volatility clustering phenomenon that occurs in economic and financial time series. In the basic ARCH model, the conditional variance of the disturbance that occurs at time t is specified to be a linear function of the squares of past disturbances. The general $ARCH(q)$ model is given by

$$h_t = \omega + \sum_{j=1}^{q} \alpha_j u_{t-j}^2 \tag{2.23}$$

Obviously, the conditional variance h_t needs to be nonnegative. To guarantee nonnegativeness of the conditional variance, it is required that $\omega > 0$ and $\alpha_j \geq 0$ for all$j = 1, \cdots, q$. To understand why the ARCH model can describe volatility clustering, observe that model (2.21) with (2.23) basically states that the conditional variance of u_t is an increasing function of the disturbance/shock that occurred in the previous q periods with some nonnegative weights. Hence, if say u_{t-1}i s large in absolute value, u_t is expected to be large in absolute value as well. In other words, large (small) shocks tend to be followed by large (small) shocks of either sign. An alternative way to see the same thing is to note that the $ARCH(q)$ model can be written as an $AR(q)$ model for u_t^2. Adding u_t^2 to (2.23) and subtracting h_t from both sides gives

$$u_t^2 = \omega + \sum_{j=1}^{q} u_{t-j}^2 + v_t \tag{2.24}$$

where $v_t \equiv u_t^2 - h_t = h_t(z_t^2 - 1)$. Note that $E[v_t|\Omega_{t-1}] = 0$. Given the AR representation of $ARCH(q)$ process, the condition that needs to be satisfied in order for u_t^2 to be covariance stationary is that the roots of the lag polynomial $\alpha(L) = 1 - \alpha_1 L - \cdots - \alpha_q L^q$ need to be outside the unit circle. Moreover, the unconditional variance of u_t, or unconditional mean of u_t^2 can be obtained as

$$\sigma_u^2 \equiv E[u_t^2] = \frac{\omega}{1 - \sum_{j=1}^{q} \alpha_j} \tag{2.25}$$

Hence $\sum_{j=1}^{q} \alpha_j < 1$ in order for the unconditional variance to be well defined. Under these conditions, (2.24) can be rewritten as

$$\begin{aligned} u_t^2 &= \frac{\omega}{1 - \sum_{j=1}^{q} \alpha_j}(1 - \sum_{j=1}^{q} \alpha_j) + \sum_{j=1}^{q} \alpha_j u_{t-j}^2 + v_t \\ &= (1 - \sum_{j=1}^{q} \alpha_j)\sigma_u^2 + \sum_{j=1}^{q} \alpha_j u_{t-j}^2 + v_t \\ &= \sigma_u^2 + \sum_{j=1}^{q} (u_{t-j}^2 - \sigma_u^2) + v_t. \end{aligned} \tag{2.26}$$

Equation (2.26) shows that if u_{t-1}^2 is larger (smaller) than its unconditional expected value σ_u^2, u_t^2 is expected to be larger (smaller) than σ_u^2 as well.

ARCH model cannot only capture the volatility clustering of the time series under investigation but also their excess kurtosis which is common in economic and financial time series. From (2.21) it can be seen that the kurtosis of u_t is always greater than that of z_t,

$$E[u_t^4] = E[z_t^4]E[h_t^2] \geq E[z_t^4](E[h_t]^2) = E[z_t^4](E[u_t^2]^2),$$

where the inequality follows from Jansen's inequality. As shown by Engle (1982), for the $ARCH(1)$ model with normally distributed z_t the kurtosis of u_t is equal to

$$Kurt_u = \frac{E[u_t^4]}{E[u_t^2]^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2},$$

which is finite if $3\alpha_1^2 < 1$. It is clear that $Kurt_u$ is always larger than the normal value of 3.

To capture the dynamic patterns in conditional volatility adequately by means of an $ARCH(q)$ model, q needs often to be quite large. Hence it can be quite cumbersome to estimate the parameters in an $ARCH(q)$ model with large q, as nonnegativity and stationarity conditions need to be imposed. To reduce the computational problems one needs to impose some structure on the parameters, such as $\alpha_j = \alpha(q+1-j)/(q(q+1)/2), j = 1, \cdots, q$, which implies that the parameters of the lagged squared shocks/disturbances decline linearly and sum to α, see Engle (1982). An alternative method is suggested by Bollorslev (1986) which involves adding lagged conditional variances to the ARCH specification. For instance, adding p such conditional variances to the $ARCH(q)$ model results in the $GARCH(p, q)$ model,

$$\begin{aligned} h_t &= \omega + \sum_{j=1}^{q} \alpha_j u_{t-j}^2 + \sum_{j=1}^{p} \beta_1 h_{t-j} h_{t-j} \\ &= \omega + \alpha(L) u_t + \beta(L) h_t \end{aligned} \tag{2.27}$$

This model avoids the necessity of adding many lagged squared disturbance terms by adding lagged values of conditional variance terms. To see why a GARCH specification takes care of adding large number of lagged residual terms consider the $GARCH(1, 1)$ model,

$$h_t = \omega + \alpha_1 u_t^2 + \beta_1 h_{t-1}. \tag{2.28}$$

This model can be rewritten as,

$$h_t = \omega + \alpha_1 u_{t-1}^2 + \beta_1(\omega + \alpha_1 u_{t-2}^2 + \beta_1 h_{t-2}),$$

or by continuing the recursive substitution one can obtain,

$$h_t = \sum_{j=1}^{\infty} \beta_1^j \omega + \alpha_1 \sum_{j=1}^{\infty} \beta_1^{j-1} u_{t-j}^2. \tag{2.29}$$

This equation shows that the $GARCH(1, 1)$ model corresponds to an $ARCH(\infty)$ model with a particular parameter structure. This clearly illustrates why in most of the applications a low order, for instance a $GARCH(1, 1)$ model, is usually found to be general enough to capture the dynamic behavior of many economic and financial time series.

An alternative representation of a $GARCH(1,1)$ model can be obtained by adding u_t^2 to both sides of (2.28) and moving h_t to the right-hand side,

$$u_t^2 = \omega + (\alpha_1 + \beta_1)u_{t-1}^2 + v_t - \beta_1 v_{t-1}, \tag{2.30}$$

where again $v_t = u_t^2 - h_t$. This $ARMA(1,1)$ representation allows one to establish conditions for the covariance stationarity of the $GARCH(1,1)$ process. From (2.30) it is obvious that $GARCH(1,1)$ model is covariance stationary if and only if $\alpha_1 + \beta_1 < 1$. In this case the unconditional mean of u_t^2 - or unconditional variance of u_t - is equal to

$$\sigma_u^2 = \frac{\omega}{1 - \alpha_1 + \beta_1}. \tag{2.31}$$

The parameters in $GARCH(1,1)$ model need to satisfy $\omega > 0$, $\alpha_1 > 0$ and $\beta_1 \geq 0$ in order to guarantee that $h_t \geq 0$. Moreover, α_1 needs to be strictly positive in order for β_1 to be identified. This is because, if $\alpha_1 = 0$ in (2.30) both AR and MA polynomials become $1 - \beta_1 L$, hence when one rewrites the $ARMA(1,1)$ model for u_t^2 as an $MA(\infty)$ process polynomials will cancel out,

$$u_t^2 = \frac{1 - \beta_1 L}{1 - \beta_1 L} v_t = v_t,$$

which indicates that β_1 is not identified, see Bollerslev (1986) for details.

In the case of $GARCH(1,1)$ Bollerslev (1986) showed that the kurtosis of u_t under normality of z_t is given by

$$Kurt_u = \frac{3[1 - (\alpha_1 + \beta_1)]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2},$$

which is always larger than the normal value of 3. The autocorrelations of u_t^2 are derived in Bollerslev (1988) and are given by,

$$\rho_1 = \alpha_1 + \frac{\alpha_1^2 \beta_1}{1 - 2\alpha_1\beta_1 - \beta_1^2}, \tag{2.32}$$

$$\rho_k = (\alpha_1 + \beta_1)^{k-1} \rho_1 \text{ for } k = 2, 3, \cdots \tag{2.33}$$

The decay factor of autocorrelations is $\alpha_1 + \beta_1$. This means that if this sum is close to 1, the autocorrelations decline gradually still at an exponential rate. If the fourth moment of u_t does not exist (if $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 \geq 1$, as shown by Bollerslev 1986) then the autocorrelations of u_t^2 are time-varying. As shown by Ding and Granger (1996), if the $GARCH(1,1)$ model is covariance stationary but with infinite fourth moment, one can still compute the sample autocorrelations.

In the general $GARCH(p,q)$ model if all the roots of $1 - \beta(L)$ lie outside the unit circle, the model can be written as an infinite-order ARCH model,

$$\begin{aligned} h_t &= \frac{\omega}{1-\beta(1)} + \frac{\alpha(L)}{1-\beta(L)} u_t^2 \\ &= \frac{\omega}{1-\beta_1 \cdots - \beta_p} + \sum_{j=1}^{\infty} \delta_j u_{t-j}^2. \end{aligned} \tag{2.34}$$

To guarantee the nonnegativity of the conditional variance all δ_j need to be nonnegative. The $ARMA(m,p)$ representation of u_t^2 is given by

$$u_t^2 = \omega + \sum_{j=1}^{m} (\alpha_j + \beta_j) u_{t-j}^2 - \sum_{j=1}^{p} \beta_j v_{t-j} + v_t, \tag{2.35}$$

where $m = max(p,q), \alpha_j \equiv 0$ for $j > q$ and β_j for $j > p$. The $GARCH(p,q)$ model is covariance stationary if all the roots of $1 - \alpha(L) - \beta(L)$ lie outside the unit circle.

2.3.2 The IGARCH Model

In applications of the $GARCH(1,1)$ model to high frequency economic and financial data, it is usually found that the estimates of $\alpha_1$andβ_1 are such that their sum is close to or equal to 1. The $GARCH(1,1)$ model with restriction $\alpha_1 + \beta_1 = 1$ is referred to be the Integrated GARCH (IGARCH) model. The reason is that the restriction on these parameters leads a unit root in the $ARMA(1,1)$ representation of $GARCH(1,1)$ model. From equation (2.30) the ARMA representation of the model becomes,

$$(1-L)u_t^2 = \omega + v_t - \beta_1 v_{t-1}.$$

From (2.31) it can easily be seen that the unconditional variance of u_t is not finite. Therefore, the $IGARCH(1,1)$ model is not covariance stationary. Although, the autocorrelations of u_t^2 for an IGARCH model are not defined properly, Ding and Granger (1996) show that they are approximately equal to

$$\rho_k = \frac{1}{3}(1 + 2\alpha_1)(1 + 2\alpha_1^2)^{-k/2}.$$

The autocorrelations still decay exponentially. This is in sharp contrast to an $I(1)$ process, say for instance a random walk model, for which the autocorrelations are approximately equal to 1.

2.3.3 The FIGARCH Model

The properties of the conditional variance h_t as implied by the IGARCH model are not very attractive from an empirical point of view. The IGARCH model implies that a shock to the volatility process will have very persistent effects. The IGARCH model also implies that there is a linear trend in the future forecast of the volatility process, i.e. $E_t h_{t+k} = h_t + k\omega$, hence, the forecasts of future conditional variance increases linearly with the forecast horizon. This is not realistic from an empirical point of view. On the other hand, estimates of the $GARCH(1,1)$ model from high frequency financial time series invariably yield a sum of α_1 and β_1 close to 1, with α_1 small and β_1 large. From the $ARCH(\infty)$ representation of $GARCH(1,1)$ model, equation (2.29), it can be seen that the impact of a shock u_t on the conditional variance at a future date, h_{t+k} is given by $\alpha_1\beta_1^{k-1}$. With β_1 close to 1, the impact of a shock at time t on the conditional variance will decay very slowly as k gets larger and larger. Moreover, the autocorrelations for u_t^2 given in (2.33 and 2.34) are die out very slowly if the sum $\alpha_1 + \beta_1$ is close to 1, although the decay is still at an exponential rate. This can be seen from panel a of figure (2.3) which displays the autocorrelations for u_t^2 from a sample realization of $GARCH(1,1)$ with $\omega = 0.001$, $\alpha_1 = 0.2$, and $\beta_1 = 0.7$.

It is evident from the figure that the autocorrelations decay slowly but still the decay rate is too fast to mimic the observed autocorrelation patterns of empirical volatility processes. For example, Ding, Granger, and Engle (1993), deLima, Breidt, and Crato (1994), Baillie and Bollerslev and Mikkelsen (1996), Lobato and Sevin (1998), Dacorogna etal. (1993), Andersen and Bollerslev (1997), and Baillie, Çeçen, and Han, (2001), all report that the sample autocorrelations of absolute returns and power transformations of returns for various asset prices at different frequencies decline only at a hyperbolic rate. As this discussed in the previous section, this type of behavior of autocorrelations can be modelled by means of long memory or fractionally integrated processes.

Baillie, Bollerslev, and Mikkelsen (1996) propose the class of Fractionally Integrated GARCH (FIGARCH) models. The FIGARCH process is capable of modelling very slow hyperbolic decay in the autocorrelations of the volatility process quite flexibly. Re-writing the $ARMA(m,p)$ representation of the $GARCH(p,q)$ model as,

$$[1-\beta(L)-\alpha(L)]u_t^2 = \omega + [1-\beta(L)]v_t,$$

the $FIGARCH(p,\delta,q)$ model can be obtained by simply adding $(1-L)^\delta$ term on the left hand side of this $ARMA(m,p)$ representation. More explicitly, the $FIGARCH(p,\delta,q)$ model is given by

$$\phi(L)(1-L)^\delta u_t^2 = \omega + [1-\beta(L)]v_t, \tag{2.36}$$

where $\phi(L) = [1-\beta(L)-\alpha(L)](1-L)^{-\delta}$, all the roots of $\phi(L)$ and $[1-\beta(L)]$ lie outside the unit circle, and $0<\delta<1$. For $0<\delta<1$, $\phi(L)$ is an infinite order polynomial, while it is of order $m-1$ for $\delta=1$. As it is evident from (2.36) the FIGARCH model nests GARCH and IGARCH models in the sense that when $\delta=0$ the FIGARCH model reduces to the GARCH model while for $d=1$ it becomes an IGARCH model. Rearranging the terms in (2.36) an alternative representation for the FIGARCH model can be obtained as,

$$[1-\beta(L)]h_t = \omega + [1-\beta(L) - \phi(L)(1-L)^{\delta}]u_t^2. \tag{2.37}$$

From this representation, the conditional variance of u_t, or infinite ARCH representation of the FIGARCH process, is simply

$$\begin{aligned} h_t = \frac{\omega}{1-\beta(1)} + [1 - \frac{\phi(L)}{1-\beta(L)}(1-L)^{\delta}]u_t^2 \\ \equiv \frac{\omega}{1-\beta(1)} + \lambda(L)u_t^2, \end{aligned} \tag{2.38}$$

where $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \cdots$. For the $FIGARCH(p, \delta, q)$ process to be well defined and the conditional variance to be positive for all t, all the coefficients in the infinite ARCH representation in (2.38) need to be nonnegative, i.e. $\lambda_j \geq 0$ for $j = 1, 2, \cdots$. The general conditions for nonnegativity of lag coefficients in $\lambda(L)$ are not easy to establish, but as illustrated in Baillie et al. (1996) it is possible to show sufficient conditions in a case by case basis.

The FIGARCH process implies a slow hyperbolic rate of decay for the autcorrelations of u_t^2 as can be seen from panel b of figure 2.3 which displays the first fifty autocorrelations of u_t^2 from a sample realization of a $FIGARCH(1, \delta, 1)$ process. For $0 < \delta \leq 1$, $\lambda(1) = 0$ and hence the second moment of the unconditional distribution of u_t is infinite, and FIGARCH process is not covariance stationary similar to IGARCH processes. As argued in Baillie et al. (1996) just like the IGARCH processes it can be shown that FIGARCH processes are strictly stationary and ergodic for $0 < \delta \leq 1$. Baillie et al. (1996) show that it is possible to obtain impulse response coefficients from the definition given in (2.36). Specifically, the coefficients from the $\gamma(L)$ lag polynomial,

$$\begin{aligned} (1-L)u_t^2 = (1-L)^{1-\delta}\phi(L)^{-1}\omega + (1-L)^{1-\delta}\phi(L)^{-1}[1-\beta(L)]v_t \\ \equiv \zeta + \gamma(L)v_t. \end{aligned} \tag{2.39}$$

The long run impact of past shocks for the volatility process can be assessed in terms of the cumulative impulse response weights,

$$\gamma(1) = \lim_{k\to\infty} \sum_{j=0}^{k} \gamma_j = \lim_{k\to\infty} \lambda_k = F(\delta - 1, 1, 1; 1)\phi(1)^{-1}[1 - \beta(1)],$$

where $F(\delta - 1, 1, 1, 1; 1)$ is the hypergeometric function. For details, see Baillie et al. (1996). Since for $0 \leq \delta < 1$, $F(\delta - 1, 1, 1; 1) = 0$, shocks to the conditional variance of FIGARCH process will die out eventually in a forecasting sense similar to a GARCH process. But the shocks to the GARCH process dissipate at a fast exponential rate while shocks to the conditional variance of a FIGARCH process is much slower at a hyperbolic rate. In contrast, for $\delta = 1$, $F(\delta - 1, 1, 1; 1) = 1$ and hence cumulative impulse rates for a IGARCH process converge to the nonzero constant $\gamma(1) = \phi(1)^1[1 - \beta(1)]$. This implies that shocks to the conditional variance of the IGARCH process persist indefinitely. For an illustration, consider the basic $FIGARCH(1, \gamma, 0)$ model discussed in Baillie et al. (1996). This model can be written as

$$(1 - L)^\delta]u_t^2 = \omega + v_t - \beta_1 v_{t-1}.$$

Using the definition of $v_t = u_t^2 - h_t$, this can be rewritten as an $ARCH(\infty)$ process for the conditional variance as,

$$\begin{aligned} h_t &= \frac{\omega}{1 - \beta_1} + [1 - \frac{(1 - L)^\delta}{1 - \beta_1 L}]u_t^2 \\ &= \frac{\omega}{1 - \beta_1} + \lambda(L)u_t^2, \end{aligned}$$

where $\lambda(L) \equiv 1 - (1 - L)^\delta/(1 - \beta_1 L)$. By using the expansion (2.6) for $(1 - L)^\delta$, it can be shown that for large k

$$\lambda_k \approx [(1 - \beta_1)\Gamma(L)^{-1}]k^{\delta - 1}.$$

It is evident from this expression that the effect of u_t on h_{t+k} decays only at a hyperbolic rate as k increases.

2.4 ARFIMA-FIGARCH Model: Modelling long memory in both conditional mean and variance

A model that combines long memory processes for both the conditional mean and variance processes and allows one to model jointly the long memory in time series that may have long memory property in both its conditional mean and variance process is the $ARFIMA(P,d,Q) - FIGARCH(p,\delta,q)$. The ARFIMA-FIGARCH process can be expressed as,

$$\begin{aligned}
\Phi(L)(1-L)^d y_y &= \Theta(L)u_t \\
u_t &= z_t\sqrt{h_t} \\
\beta(L)h_t = \omega + [1-\beta(L) &- \phi(L)(1-L)^{\delta}]u_t^2
\end{aligned} \tag{2.40}$$

where $\beta(L)$, and $\phi(L)$ are the same as before, while $\Phi(L) = 1 - \Phi_1 L - \cdots - \Phi_P L^P$, $\Theta(L) = 1 + \Theta_1 L + \cdots + \Theta_Q L^Q$, and have all their roots outside the unit circle. Moreover, $E_{t-1}z_t = 0, E_{t-1}(z_t^2) = 1$. This model is capable of modelling both short run dynamics and long run properties of a time series in both conditional mean and variance very parsimoniously. Note that if $h_t = \omega$ then the model reduces to the $ARFIMA(p,d,q)$ model for the conditional mean process discussed above. If $p = q = d = 0$ the model becomes so called Martingale-FIGARCH process for the conditional mean and variance. The Martingale-FIGARCH model is appealing as it allows one to model random walk and highly persistent conditional second moments of many high frequency asset prices. The Martingale-FIGARCH model is fit to daily and high frequency exchange rate data (hourly of half-an hour data) by Baillie, Bollerslev and Mikkelsen (1996), and most recently by Baillie, Çeçen, and Han (2001). On the other hand, Baillie, Han and Kwon (2001) applied the $ARFIMA(p,d,q) - FIGARCH(P,\delta,Q)$ model to inflation series and obtained results that indicate presence of long memory dynamics in both the conditional mean and variance of the inflation series for several industrial countries. As noted in Baillie, Han and Kwon (2001), contrary to pure ARFIMA process, ARFIMA-FIGARCH process have an infinite unconditional variance for all d given $\delta \neq 0$.

2.5 Estimation and Inference

Several methods of estimating long memory parameter d have been suggested in the literature. The early methods are mostly heuristic in the sense that they are simple diagnostic tools used in detecting the presence of long memory. Most of these methods are discussed extensively in Beran (1994). More advanced and rigorous methods are developed to estimate long memory and parameters of long memory models discussed in the previous sections in both time and frequency domain. A complete review and discussion of them can be found in Baillie (1996) and Beran (1994) and references therein. In this section some of these methods, those mostly used among applied economists are discussed. In particular, semi-parametric estimation in the frequency domain (or least squares regression in the frequency domain) due to Geweke and Portar-Hudak (1983) and Robinson (1994, 1995), approximate maximum likelihood estimation in the frequency domain due to Whittle (1951) and Fox and Taqque (1986), and approximate maximum likelihood estimation (or nonlinear least squares estimation, or conditional sum of squares estimator) in the time domain due originally to Hosking (1984) in the context of ARFIMA processes, and Baillie and Chung (1993) in the context of ARFIMA-FI/GARCH processes, will be discussed in some detail within the context of the long memory models discussed in the previous sections.

2.5.1 Regression based estimation in the frequency domain

In the spectral domain, Geweke and Portar-Hudak (1983) suggested a semiparametric procedure to obtain an estimate of the fractional differencing parameter

d based on the slope of the spectral density function around the angular frequency. The spectral density of a stationary Gaussian long-memory time series y_t is given by

$$f(\omega) = |1 - \exp(-i\omega)|^{-2d} f(\omega)^* \tag{2.41}$$

where $d \in (-0.5, 0.5)$ and $f(\omega)^*$ is an even, positive continuous function on $[-\pi, \pi]$, bounded above and bounded away from zero with first derivative $f^{*\prime} = 0$ and second and third derivatives bounded in a neighborhood of zero. The function $f(\omega)^*$ endows the model (2.41) with a short-term correlation structure which is free of any parametrically imposed constraints. For this reason the semi-parametric model in (2.41) may be preferable to the assumption that the time series obeys an $ARFIMA(p, d, q)$ process with p and q finite, either known or unknown as in the $ARFIMA(p, d, q)$ model discussed above. Note the fact that the ARFIMA model is a special case of (2.41) that can be obtained by assuming $f(\omega)^*$ to be the spectral density of a stationary invertible $ARMA(p, q)$ process as in (2.12). The long memory parameter, d can be estimated semi-parametrically based on the first periodogram ordinates

$$I_j = \frac{1}{2\pi T j} T |\sum_{t=0}^{T-1} y_t \exp(i\omega_j t)|^2, \; j = 1, \cdots, m \tag{2.42}$$

where $\omega_j = 2j\pi/T$ and m is a positive integer. The semi-parametric estimator which is also known as GPH estimator in the literature, is given by $-\frac{1}{2}$ times the least squares estimate of the slope parameter in an ordinary linear regression of $\{\log I_j\}_{j=1}^{m}$ on the explanatory variable

$$x_j \equiv \log \|1 - \exp(-i\omega_j)\| = \log \|2 \sin(\frac{\omega_j}{2})\|,$$

together with a constant term. Therefore the GPH estimator can be written as

$$\hat{d}_{GPH} = \frac{-0.5 \sum_{j=1}^{m} (x_j - \overline{x}) \log I_j}{\sum_{k=1}^{m} (x_k - \overline{x})^2} \tag{2.43}$$

where $\overline{x} = \frac{1}{m}\sum_{k=1}^{m} x_k$. The GPH estimator can be motivated heuristically by noting that

$$\log I_j = (\log f_0^* - C) - 2dx_j + \log \frac{f_j^*}{f_0^*} + \epsilon_j,$$

where $\epsilon_j = \log(I_j/f_j) + C$, with $f_j = f(\omega_j)$, $f_j^* = f_j^*(\omega_j)$ and $C = 0.577216$ is the Euler's constant. It is assumed that $m \to \infty$, so that the variance of $\hat{d}_{GPH}$ will decrease to zero as $T \to \infty$, and also that $\frac{m}{T} \to 0$, so that bias due to the non-constancy of $\log(f_j^*/f_0^*)$ will tend to zero.

Although the GPH estimator is widely used in practice, its consistency for all $d \in (-0.5, 0.5)$ and asymptotic normality have only recently been proved by Hurvich et al. (1998). Robinson (1995) did prove consistency and asymptotic normality for a modified regression estimator which regresses $\{\log I_j\}_{j=l+1}^{m}$ on $\{x_j\}_{j=l+1}^{m}$, where l is a lower truncation point which tends to infinity more slowly than m. However, simulations (e.g. Hurvich and Beltaro, 1994) indicate that the modified estimator is typically outperformed in finite samples by the GPH estimator itself. The reason is that any bias reduction resulting from omission of the first l periodogram ordinates from the regression is more than offset by inflation of the variance (see Hurvich and Beltrao, 1994). Hurvitch et al (1998) show that the optimal (in the sense that it minimizes the theoretical mean squared error of the GPH estimator) choice of m is in the order of $O(T^{4/5})$. They present simulation results to asses the accuracy of their asymptotic theory on the mean squared error for finite sample sizes. Their findings indicate that the choice $m = T^{1/2}$, originally suggested by Geweke and Porter-Hudak (1983) and used extensively in the empirical literature, can lead to performance which is markedly inferior to that of asymptotically optimal choice in reasonably small samples.

The GPH estimation only allows one to estimate the long memory parameter. In a parametric model, such as in the case of $ARFIMA(p, d, q)$ model given in (2.6) all of the other parameters (i.e. ARMA parameters, variance, and the mean parameter) in principal can be estimated in the second step by any appropriate method, such as maximum likelihood once the series is filtered by the estimate of the long memory parameter, $\hat{d}_{GPH}$. A problem with this two-step approach is that the sampling distribution of estimators is not known yet. The problem may be much more serious in the models with GARCH or FIGARCH effects in the conditional variance of the process. Moreover, there is some evidence that in the case of autocorrelated disturbances the GPH estimator may have serious biases. See for instance Agiakloglu, Newbold, and Whoar (1993). The next subsection discusses methods that estimates jointly the long memory parameter and the $ARMA$ parameters.

2.5.2 Parametric Methods: Approximate Maximum Likelihood

It seems that if one is only interested in having an idea about the presence of long memory or not in a time series the GPH estimator may provide information about the presence of long memory. If on the other hand one needs to understand both short run and long run dynamics of a time series and use the model for describing the dynamic structure of the series and/or use the model for forecasting purposes, the GPH estimator obviously will not tell anything about the short term properties of the process. Methods which allow one to model the whole autocorrelation structure, or, equivalently, the whole spectral density at all frequencies, have to be used to characterize the short-run behavior of the series. One such approach is to use parametric models, such as the ARFIMA model in (2.6) and estimate parameters, for example, by maximizing the likelihood. One such method is the exact maximum likelihood estimator (MLE) of the $ARFIMA(p,d,q)$ model under the assumption that u_t is normally distributed. The exact MLE for the $ARFIMA(p,d,q)$ model is developed in Sowell (1992). Given the $ARFIMA(p,d,q)$ process in (2.6) the log-likelihood function is

$$\ell_E(y;\varphi) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log\det\Sigma(\varphi) - \frac{1}{2}y'\Sigma(\varphi)^{-1}y \tag{2.44}$$

where Σ is the variance-covariance matrix whose i, jth element is given by $\Sigma_{i,j} = \gamma_{|i-j|}$, y is the $T-$dimensional vector of observations on the process y_t, and $\varphi = (d,\ \phi_1 \cdots, \phi_p,\ \omega_1, \cdots, \omega_q,\ \sigma_u^2)'$, is the parameter vector in the $ARFIMA(p, d, q)$ model with known mean μ. The exact MLE of φ is obtained by maximizing (2.44) with respect to the $k = p + q + 2$ dimensional parameter vector. The consistency and normality of exact MLE of the $ARFIMA(p, d, q)$ model is established in Yajima (1985) and Dahlhaus (1989) for the Gaussian long memory processes. Although exact MLE of φ can in principal be obtained by the MLE procedure, in practice, exact MLE has serious computational problems. The exact MLE requires the inversion of a $T \times T$ matrix of nonlinear hypergeometric functions at each iteration of the maximization of the likelihood. To solve the computational problem an alternative approach is to maximize an approximation to the likelihood function. There are several alternative approximate MLE of the $ARFIMA(p, d, q)$ model under normality of disturbances. Two such approximate MLEs that are mostly used in empirical work are discussed here.

2.5.3 Whittle's approximate MLE

The two terms in (2.44) that depend on the parameter vector, φ are the logarithm of the determinant of the covariance matrix,

$$\log\det\Sigma(\varphi),$$

and the quadratic form

$$y'\Sigma(\varphi)^{-1}y.$$

The Whittle's approximate MLE uses the approximations for these terms in the log-likelihood function. In particular,

$$\lim_{T\to\infty} \log\det\Sigma(\varphi) = \log(2\pi)f(\omega_j).$$

and second term approximated by $I(\omega_j/f(\omega_j)$. Then the approximate log-likelihood is

$$\ell_W = \sum_{j=1}^{T-1} \log[(2\pi)f(\omega_j;\varphi)] + \sum_{j=1}^{T-1} \frac{I(\omega_j)}{f(\omega_j;\varphi)}, \tag{2.45}$$

where $\omega_j = 2\pi j/T - 1$, and $f(.)$ is the spectral density. An alternative approximate MLE is given by Fox and Taqque (1986) which numerically minimize the quantity

$$\sum_{j=1}^{m} \frac{I(\omega_j)}{f(\omega_j;\varphi)}, \tag{2.46}$$

where m is the number of frequencies used. For a detailed discussion of Whitlle's approximate MLE see Beran (1994) and references there.

2.5.4 Approximate MLE in the time domain

In this subsection estimation of long memory models will be discussed within the context of both $ARFIMA(p,d,q)$ model for the conditional mean process as well as the $FIGARCH(P,\delta,Q)$ model for conditional volatility. The setup of the technique is general enough to cover both types of long memory processes and the dual long memory model $ARFIMA(p,d,q) - FIGARCH(P,\delta,Q)$. To this end general principles are discussed first, and some remarks on specific models will be given.

Consider the $ARFIMA(p,d,q) - FIGARCH(P,\delta,Q)$ model given in (2.40). Under the assumption that disturbances are conditionally normally distributed the conditional log-likelihood can be written in the time domain is

$$\ell(u_1\cdots,u_T;\varphi) = -\frac{T}{2}\ln 2\pi - \sum_{t=1}^{T}[\ln h_t + \frac{u_t^2}{h_t}], \tag{2.47}$$

where $\varphi' = (\mu,\ \Phi_1\cdots\Phi_P,\ \Theta_1\cdots\Theta_Q, \omega\,\delta\,\beta_1\cdots\beta_p, \phi_1\cdots\phi_q)$. Since conditional normality of u_t is often not a very realistic assumption for many economic and financial time series, the resulting model fails to capture the kurtosis in the data. Instead, following Bollerslev (1987) one sometimes assumes that z_t is drawn from a (standardized) Student-t distribution. Note that the standardized Student-t distribution with ν degrees of freedom is,

$$f(z_t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)}(1+\frac{z_t^2}{\nu-2})^{-(\nu+1)/2}.$$

The Student-t distribution is symmetric around zero (and thus $E[z_t = 0]$), while it converges to the normal distribution as the number of degrees of freedom ν becomes larger. A further characteristic of the Student-t distribution is that only moments up to order ν exist. Hence, for $\nu > 4$, the fourth moment of z_t exists and is equal to $3(\nu-2)/(\nu-4)$. As this is larger than the normal value of 3, the unconditional kurtosis of u_t will also be larger than in the case where z_t followed a normal distribution. The number of degrees of freedom of the Student-t distribution can be estimated along with the other parameters of the model. Indeed any other distribution can be assumed. The parameters of the model under consideration then can be estimated by maximizing the log-likelihood corresponding with this particular distribution. As one can never be sure that the specified distribution of the disturbances is the correct one, an alternative approach is to ignore the problem and base the likelihood on the normal distribution as in (2.47). This method usually is referred to as quasi-maximum likelihood estimation (QMLE). In general, the resulting estimates are still consistent and asymptotically normal, provided that the models for the conditional mean and conditional variance are correctly specified. Li and McLeod (1986) have shown the consistency and asymptotic normality of QMLE for the $ARFIMA(P, d, Q)$-homoscedastic model with mean μ either known or zero. Dahlhaus (1988, 1989) and Moehring (1990) showed the same result with μ unknown. In particular, they show that the parameter estimates in the ARFIMA model with homoscedastic disturbances are asymptotically normal, with the ARFIMA parameter estimates being $T^{1/2}$ consistent while the QMLE of μ is $T^{1/2-d}$ consistent. For the conditional variance process, asymptotic normality and consistency have only been shown in specific cases. Weiss (1984, 1986) has demonstrated consistency and asymptotic normality for QMLE of $ARCH(q)$ model as in (2.24), while Bollerslev and Wooldridge (1992), Lee and Hansen (1994) and Lumsdaine (1996) have obtained the same result where h_t follows a $GARCH(1,1)$ under varying assumptions on the properties of z_t. Lumsdaine (1996) also illustrated consistency and asymptotic normality for the QMLE of $IGARCH(1,1)$ model. While simulation experiments for FIGARCH processes in Baillie and Bollerlev (1996) indicate consistency and asymptotic normality of the QMLE, a fully general theoretical treatment is not available yet. In the case of the more general models ARFIMA-GARCH and ARFIMA-FIGARCH, Baillie, Chung, and Tieslau (1996) and Baille, Han, and Kwon (2001) through simulations provide evidence that the QMLE is consistent and asymptotically normal.

As the true distribution of z_t is not assumed to be the same as the normal distribution which is used to construct the likelihood function, the standard errors of the parameters have to be adjusted accordingly. In particular, the asymptotic covariance matrix of $D_T(\hat{\varphi} - \varphi_0)$ is equal to

$$D_T^{-1} A(\varphi_0)^{-1} B(\varphi_0) A(\varphi_0) D_T^{-1}, \tag{2.48}$$

where $A(.)$ is the Hessian, i.e. the negative of the matrix of second-order partial derivatives of the log likelihood function with respect to the parameters in the model, $H(\varphi) \equiv -\partial \ell(u_1, \cdots, u_T; \varphi)^2 / \partial\varphi\partial\varphi'$, $B(.)$ is the expected value of the outer product of the gradient matrix,

$$\frac{1}{T} \sum_{t=1}^{T} E(\frac{\partial \ell(\varphi_0)}{\partial \varphi} \frac{\partial \ell(\varphi_0)}{\partial \varphi'}),$$

and D_T is a diagonal matrix with $diag(D_T) = [T^{1/2-d}, T^{1/2}, \cdots, T^{1/2}]$. The matrices

$A(.)$ and $B(.)$ can be consistently estimated by their sample analogs, namely,

$$A_T(\hat{\varphi}) = -\frac{1}{T}\sum_{t=1}^{T}(\frac{\partial \ell^2(\hat{\varphi})}{\partial\varphi\partial\varphi'}),$$

and

$$B(\hat{\varphi}) = \frac{1}{T}\sum_{t=1}^{T}(\frac{\partial \ell(\hat{\varphi})}{\partial\varphi}\frac{\partial \ell(\hat{\varphi})}{\partial\varphi'}).$$

As the first order conditions in maximization of the log likelihood will be nonlinear functions in the parameter of the models discussed here, an iterative optimization procedure has to be used to obtain the MLE $\hat{\varphi}$. The most frequently used iterative optimization procedures that can be used to estimate the parameters typically require the existence of first and second order derivatives of the log likelihood function with respect to φ -that is, the score $s(\varphi) \equiv \partial\ell/\partial\varphi$ and Hessian matrix $H(\varphi)$ defined above. For example, the iterations in the well known Newton-Raphson method take the form

$$\hat{\varphi}^k = \hat{\varphi}^{k-1} - \lambda(\sum_{t=1}^{T} H_t(\hat{\varphi}^{k-1})^{-1}\sum_{t=1}^{T} s_t(\hat{\varphi}^{k-1}), \tag{2.49}$$

where $\hat{\varphi}^k$ is the estimate of the parameter vector obtained in the mth iteration and the scalar λ denotes a step size. In the BHHH algorithm which is by far the most popular method to estimate GARCH and FIGARCH models, the Hessian $H_t(\hat{\varphi})$ in (2.49) is replaced by the outer product of the gradient matrix $B_t(\hat{\varphi}^{k-1})$ as given above.

2.6 Conclusion

This chapter provided a concise review of the long memory models for the conditional mean and variance of a time series. In particular, $ARFIMA(p,d,q)$ model for the conditional mean of a time series and $GARCH(p,q)$ and $FIGARCH(p,\delta,q)$ models for the conditional variance are discussed. The discussion is cast in terms of properties of the models and estimation of these models. Chapters 4 and 5 of the dissertation include applications of these models in commodity and stock markets.

BIBLIOGRAPHY

[1] Agiakloglu, C., P. Newbold and M. Wohar (1993), Bias in estimator of the fractional difference parameter, *Journal of Time Series Analysis* **14,** 235-246.

[2] Anderson, T. G. and T. Bollerslev (1997), Heterogenous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns, *Journal of Finance,* **3,** 975-1005.

[3] Aydogan, K. G.G. Booth (1988), Are there long cycles in common stock returns?, *Southern Economic Journal* **55**, 141-149.

[4] Baillie, R. T., (1996), Long Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics,* **73,** 5-59.

[5] Baillie, R. T., (1998), Comment *Journal of Business & Economic Statistics,* **16,** 273-276.

[6] Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen (1996), Fractionally integrated Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* **74,** 3-30.

[7] Baillie, R.T., C.-F. Chung, and M.A. Tieslau (1996), Analyzing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics* **11,** 23-40.

[8] Baillie, R. T. , Y. W. Han, and Tae-Go Kwon (2001), Further long memory properties of inflationary shocks, forthcoming, *Southern Economic Journal.*

[9] Baillie, R. T., A. A. Çeçen, and Y-W. Han (2001), High frequency Deutch mark-US dollar retruns: FIGARCH representations and non linearities, Manuscript, Department of Economics, Michigan State University, East Lansing, MI.

[10] Beran, J. (1994), Statistics for Long-Memory Processes, Chapman & Hall

[11] Black, F. (1976), The pricing of commodity contracts, *Journal of Financial Economics* **3,** 167-179.

[12] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307-327.

[13] Bollerslev, T. (1987), A conditionally heteroscedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* **69,** 542-547.

[14] Bollerslev, T. (1988), On the correlation structure of the generalized autoregressive conditional heteroscedastic process, *Journal of Time Series Analysis* **9,** 121-131.

[15] Bollerslev, T. and J. M. Wooldridge (1992), Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, *Econometric Reviews* **11,** 143-172.

[16] Bollerslev, T. and H.O.A. Mikkelsen (1996), Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73,** 151-184.

[17] Chambers, M. (1998), Long memory and aggregation in macroeconomic time series, *International Economic Review* **39,** 1053-1072.

[18] Cheung, Y. and Lai, K. (1995), A search for long memory in international stock market returns, *Journal of Internationtal Money and Finance,* **14,** 597-615.

[19] Cioczek-Georges, R., and B. B. Mandelbrot (1995), A class of micropulses and antipersistent fractional brownian motion, *Stochastic Processes and Their Applications* **60,** 1-18.

[20] Chung, C-F. and R.T. Baillie, 1993, Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models, *Empirical Economics* **18,** 791-806.

[21] Chung, C-F., (1994), A note on calculating the autocovariances of the fractionally integrated ARMA models, *Economic Letters, Economic Letters* **45,** 293-297.

[22] Crato, N. (1994), Some international evidence regarding the stochastic behavior of stock returns, *Applied Financial Economics,* **4,** 33-9.

[23] Dacorogna, M. M., U. A. Muller, R. J. Naglar, R. B. Olsen, and O. V. Pictet (1993), A geographical model for the daily and weekly seasonal volatility inthe foreign exchange markets, *Journal of International Money and Finance* **12,** 413-438.

[24] Dahlhaus, R. (1989), Efficent parameter estimation for self-similar processes, *Annals of Statistics* **17,** 1749-1766.

[25] Diebold, F. X., and A. Inoue (2001), Long memory and regime switching, *Journal of Econometrics* **105,** 131-159.

[26] Ding, Z., C.W.J. Granger, and R.F. Engle (1993), A long memory property of stock returns and a new model, *Journal of Empirical Finance,* **1**, 83-106.

[27] Ding, Z. and C. W. J. Granger (1996), Modeling volatility persistence of speculative returns: a new approach, *Journal of Econometrics* **73,** 185-215.

[28] Engle, R. F. (1982), Autoregressive conditionally heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econmetrica* **50,** 987-1007.

[29] Fama, E. F. (1965), The behavior of stock market prices, *Journal of Business* **38,** 34-105.

[30] Fox, R., and Taqqu, M. S. (1986), Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics,* **14,** 517-532.

[31] Granger, C., (1980), Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics,* **14,** 227-238.

[32] Granger, C., and R. Joyeux (1980), An Introduction to Long Memory Time Models and Fractional Differencing, *Journal of Time Series Analysis,* **1,** 15-29.

[33] Green, M. T., and Fieltz B. D. (1977), Long-term Dependence in Common Stock Returns, *Journal of Financial Economics,* **4,** 339-349.

[34] Gweke, J. and Portar-Hudak, S. (1983), The estimation and application of long memory time series models, *Journal of Time Series Analysis* **4,** 221-238.

[35] Hamilton, J. (1994), Time Series Analysis, Princeton, New Jersey: Princeton University Press.

[36] Hosking, J. (1981), Fractional Differencing, *Biometrika,* **68,** 165-176.

[37] Hosking, J. R. M. (1984), Modelling persistence in hydrological time series using fractional differencing, *Water Resources Research* **20,** 1898-1908.

[38] Hurst, H. (1951), Long Term Storage Capacity of Reservoirs, *Transactions of the American Society of Civil Engineers,* **116,** 770-799.

[39] Hyde, C. C., and Y. Yang (1997), On defining long range dependence, *Journal of Applied Probability* **34,** 939-944.

[40] Hurvich, C. M. and Beltrao, K. I. (1994), Automatic semiparametric estimation of the parameter of a long memory time series, *Journal of Time Series Analysis,* **15,** 285-302.

[41] Hurvich, C.M., Deo, R. and Brodsky, J. (1998), The mean squared error of Gweke and Portar-Hudak's estimator of the long memory parameter of a long-memory time series, *Journal of Time Series Analysis* **19,** 19-46.

[42] Lee, S. W. and B. E. Hansen (1994), Asymptotic theory for the $GARCH(1,1)$ quasi-maximum likelihood estimator, *Econometric Theory* **10**, 29-52.

[43] Lumsdaine, R. L. (1996), Consistency and asymptotic normality of quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, *Econometrica* **64,** 575-596.

[44] Lippi, M., and P. Zaffaroni (1999), Contemporaneous aggregation of linear dynamic models, Manuscript, Research Department, Bank of Italy.

[45] Lo, A. W. (1991), Long-term memory in stock market prices, *Econometrica,* **59,** 1279-313.

[46] Lobato, I. N., and Savin, N. E. (1998), Real and spurious long-memory properties of stock-market data, (with discusssion), *Journal of Business & Economic Statistics,,* **16,** 261-283.

[47] Mandelbrot, B., (1963a) New methods in statistical economics, *Journal of Political Economy* **71,** 421-40.

[48] Mandelbrot, B. (1963b) The variation of certain speculative prices, *Journal of Business* **36,** 394-419.

[49] Mandelbrot, B. B. (1971), When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models, *Review of Economics and Statistics,* **53,** 225-36.

[50] McLeod, A. I., and K. W. Hipel (1978), Preservation of the rescaled adjusted range, *Water Resources Research* **14,** 491-518.

[51] Parke, W. R. (1999), What is fractional integration? *Review of Economics and Statistics* **81,** 632-638.

[52] Robinson, P.M. (1990), Time series with strong dependence, Advances in econometrics, 6th world congress, Cambridge University Press, Cambridge.

[53] Robinson, P. M. (1994), Semiparametric analysis of long-memorytime series, *The Annals of Statistics* **22,** 515-539.

[54] Robinson, P. M. (1995), Log-periodgram regression time series with long-range dependence *Annals of Statistics* **23,** 1048-72.

[55] Robinson, P.M. and F.J. Hidalgo, (1997), Time series regression with long-range dependence, *Annals of Statistics* **27,** 77-104.

[56] Samarov, A. and M.S. Taqqu (1988), On the efficency of the sample mean in long memory noise, *Journal of Time Series Analysis* **9,** 191-200.

[57] Sowell, F. (1992), Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics,* **53,** 165-188.

[58] Taylor, S. (1986), Modelling Financial Time Series, John Wiley & Sons, New York.

[59] Taqqu, M. S., W. Willinger, and R. Sherman (1997), Proof of a fundamental result in self-similar traffic modelling, *Computer Communication Review* **27,** 5-23.

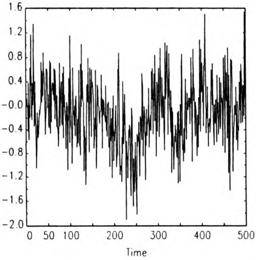
[60] Yajima, Y. (1985), On estimation of long memory time series models, *Australian Journal of Statistics* **27,** 303-320.

[61] Yajima, Y. (1991), Asymptotic properties of the LSE in a regression model with long-memory stationary errors, *Annals of Statistics* **19,** 158.

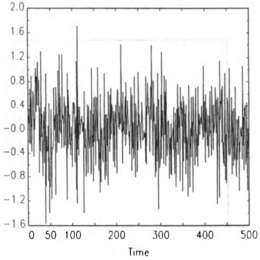
[62] Whittle, P. (1951) Hypothesis testing in time series analysis, Haftner: New York.

Figure 2.1: Sample realizations from $ARFIMA(p, d, q)$ processes

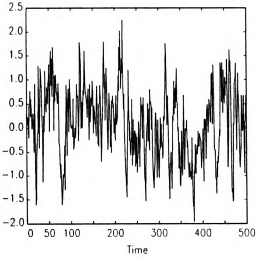
(a)$ARFIMA(0, 0.3, 0)$



(b)$ARFIMA(0, 0.3, 1)$



(c)$ARFIMA(1, 0.3, 0)$



(d) $ARFIMA(1, 0.3, 1)$

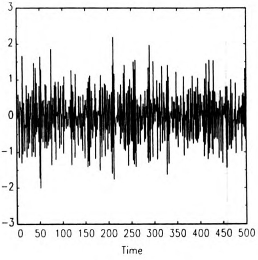
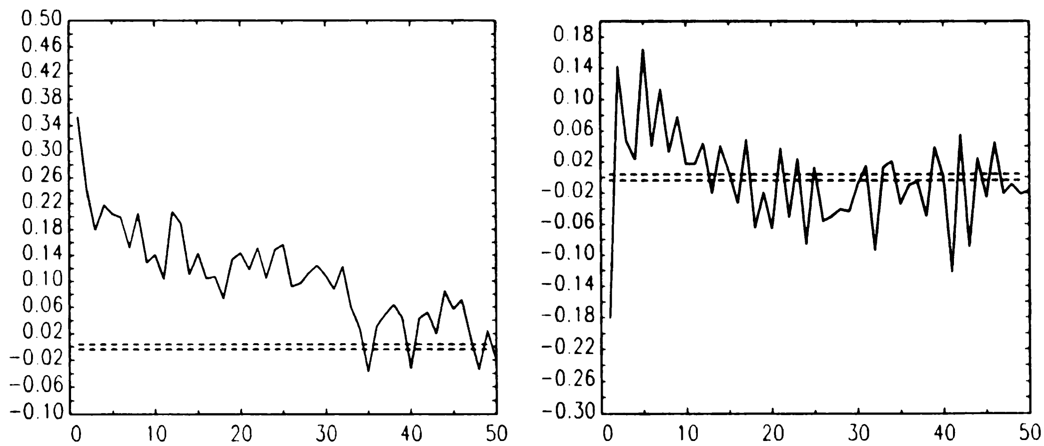


Figure 2.2: Autocorrelations of the Sample realizations from $ARFIMA(p, d, q)$ processes

(a)$ARFIMA(0, 0.3, 0)$ (b)$ARFIMA(0, 0.3, 1)$



(c)$ARFIMA(1, 0.3, 0)$ (d) $ARFIMA(1, 0.3, 1)$

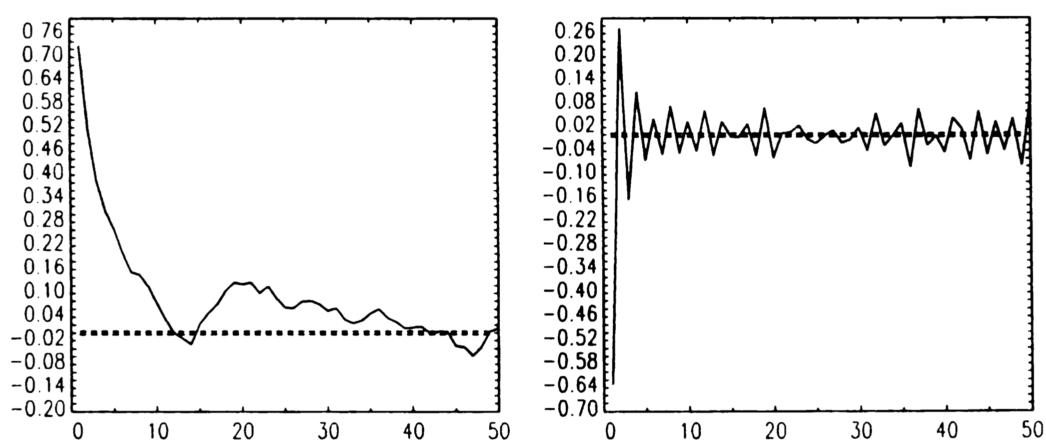
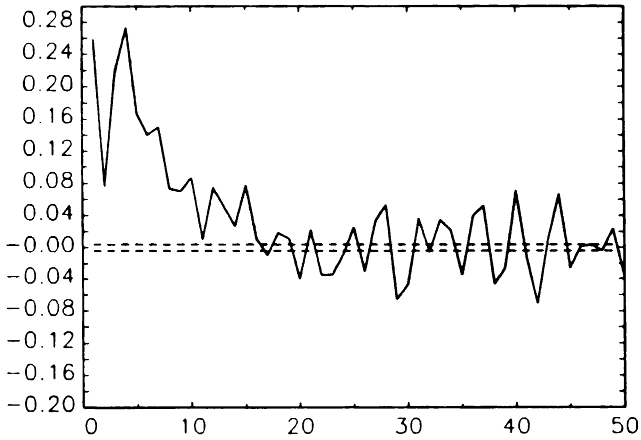
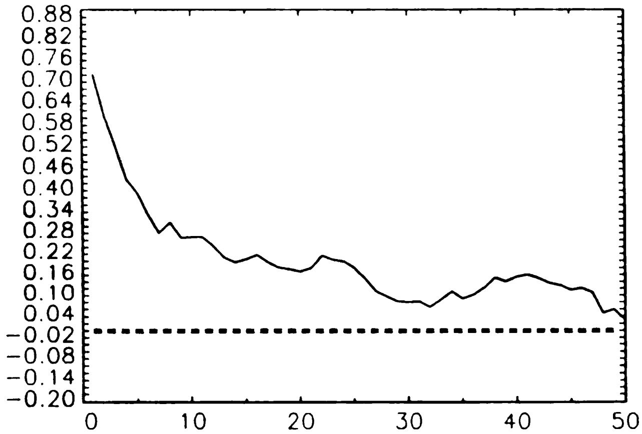


Figure 2.3: Autocorrelations of u_t^2 from sample realizations of GARCH(1, 1) and $FIGARCH(1, d, 1)$ processes
(a)$GARCH(1,1): \ h_t = 0.001 + 0.2u_{t-1}^2 + 0.7h_{t-1}$



(b) $FIGARCH(1, d, 1): \ (1 - 0.6L)h_t = 0.001 + [1 - 0.6L - 0.2(1 - L)^{0.35}]u_t^2$



CHAPTER 3

Persistence and Nonlinearity in Real Exchange Rates

3.1 Introduction

The purchasing power parity (PPP) condition states that a common basket of goods quoted in the same currency needs to cost the same in all countries. The condition rests on the assumption of perfect commodity arbitrage across countries. Although very few economists would believe that PPP holds true continuously in the real world, most would believe some form of PPP holds at least as a long-run relationship. Both traditional and new open economy macroeconomics based on intertemporal optimizing models assume some variant of PPP (Obstfeld and Rogoff, 1996). Apart from a constant term reflecting differences in units of measurement, real exchange rates are defined to be the deviation from PPP,

$$q_t = s_t - (p_t - p_t^*), \tag{3.1}$$

where s_t is the logarithm of the nominal exchange rate observed at time t, and p_t and p_t^* are the logarithms of the domestic and foreign price levels, respectively. A necessary condition for PPP to hold in the long run is that the real exchange rate needs to be stationary, not driven by permanent shocks.

Previous results from many single equation unit root tests indicate that, the unit root hypothesis in real exchange rates cannot be rejected in data from the free-floating period. Similarly, there is an absence of cointegration between nominal exchange rates and relative price levels, see Froot and Rogoff (1996), and Rogoff (1996), for recent surveys. Only from 1900 or further back is there evidence that real exchange rates are stationary, see for instance Diebold et al. (1991). To overturn this somehow puzzling empirical evidence, Pedroni (1995), Frankel and Rose (1996), Oh (1996), Wu (1996) and Lothian (1997) among others, applied panel data variants of standard unit root and cointegration tests. The idea behind these studies is to increase the power of the tests by increasing the sample size. These studies report evidence of mean reversion in real exchange rates for the floating era. One important critique of the panel data methods came from O'Connell (1998a). O'Connell's criticism centers on the failure of the panel data tests in controlling cross-sectional dependence in the data. He finds no evidence against the unit root in real exchange rate data for several countries when cross-sectional dependencies are taken into account. As noted by Rogoff (1996), the results of panel data and long-span studies seem to indicate a half-life of deviations from the PPP to be about three to five years. Since it is hard to believe that real shocks will account for the majority of short run volatility of real exchange rates and it is intuitive to think that nominal shocks can only have strong effects only a time period in which nominal wages and prices are sticky, then the apparent persistence of real exchange rates is puzzling, even if real exchange rates are mean reverting.

A recent strand of literature stresses the importance of allowing market imperfections in understanding the persistence in the adjustment of real exchange rates towards their long run equilibrium. General equilibrium models of real exchange rate determination developed in Dumas (1992) and in Sercu et al. (1995) take into account transaction costs and show that the adjustment of real exchange rates toward PPP is a nonlinear process. In these models, transaction costs create a band of inaction within which international price differentials are not arbitraged away, as only the price differentials exceeding transaction costs (outside the band) are profitable to arbitrage away. Therefore, the presence of transactions costs leads to the notion of different regimes in real exchange rates. In particular, the profits from commodity arbitrage, which is generally thought to be the ultimate force behind maintaining PPP, do not make up for the costs involved in the necessary transactions for small deviations from the equilibrium value. This means that there may exist a band around the equilibrium rate in which there is no tendency for the real exchange rate to revert to its equilibrium value. Whenever the rate is outside the band that is specified by the relevant costs, arbitrage becomes profitable, this in turn forces the real exchange rate back towards the band.

Several studies have tested and modelled the implications of transaction costs in real exchange rates. Micheal et al. (1997), use a long span of annual as well as quarterly data for the interwar period and report statistically significant evidence of nonlinearity in the adjustment of real exchange rates. Sarantis (1999), and Sarno (2000) reject linearity for several effective and bilateral real exchange rates respectively for a group of industrial countries over the floating period. Baum et al. (2001) fit the Exponential Smooth Transition Autoregressive (ESTAR) models to deviations from PPP which are obtained using the Johansen cointegration method on nominal exchange rates, home and foreign price levels. Taylor et al. (2001) report supportive evidence that the speed of convergence of real exchange rates towards their long run equilibrium increases with the size of the PPP deviation over the floating period for a number of US Dollar real exchange rates. On the other hand O'Connell (1998b) finds large deviations from PPP to be at least as persistent as small deviations.

The results of the literature seem to be unsettled and contentious in explaining the puzzling behavior of real exchange rates. Although, findings from the more recent studies that take nonlinearities into account are promising, there are certain issues that need to be investigated in judging the empirical success of these studies. Micheal et al. (1997), and Baum et al. (2001) test for cointegration in PPP, and subsequently apply the ESTAR model to the residuals from the cointegration relationship to analyze the adjustment process towards PPP. This approach may be questionable on the ground that if the residuals of PPP relationship follow a nonlinear process, the validity of the linear coinegration tests and interpretation of these residuals are doubtful. Moreover, the concept of equilibrium in nonlinear models may be different from that of linear models. To avoid these problems this chapter applies the Smooth Transition Autoregressive (STAR) models directly to the real exchange rate and then investigates the dynamic properties of the exchange rate process using well established statistical methods. Note also that theoretical models in Dumas (1992) and in Sercu et al. (1995), analyze directly the dynamic behavior of the real exchange rate process rather than the residuals that are obtained from a cointegration regression. Taylor et al. (2001), fit ESTAR models to the log real exchange rates, and then tested if there were any remaining nonlinearities left out. The problem with their approach is that the testing procedures in Taylor et al. (2001) departs from the original PPP by calling for further economic information about the other real exchange rates in the testing step, but has the drawback that this additional information is left aside in the univariate estimation of ESTAR models for the real exchange rate. For this reason, the stationarity evidence provided from their panel data tests may not be applied to univariate real exchange rates. If real exchange rates are nonstationary in the sample, then the results of their specification tests may also be questionable, as these tests are based on the assumption of stationary residuals. Moreover, since the transition variable used in their study was the lagged log real exchange rates, if the real exchange rates were nonstationary in their sample, then the process has a certain probability of being absorbed into a single regime. This in turn may invalidate the inference in the other regime.

Given the concerns discussed above, the purpose of the present chapter is twofold. One, to re-investigate more rigorously the threshold type nonlinear behavior in real exchange rates; two, to analyze carefully the persistence/mean reverting nature of real exchange rates when a nonlinearity of threshold form is allowed. More precisely, this chapter attempts to address the question to what extent does the presence of threshold dynamics in the real exchange rate resolve the puzzling evidence from unit root tests? To this end, this chapter carefully tests for the presence of threshold type nonlinearities. Three different forms of nonlinearity tests and their robust variants that take possible heteroscedasticity and outliers into consideration are applied. In addition to standard residual diagnostics, newly developed specification tests due to Eitrheim and Teräsvirta (1996), van Dijk and Franses (1999), and generalized impulse response functions, developed by Koop et al. (1996), are used as diagnostic tools to better evaluate the estimated models. The results of linearity tests and estimated STAR models provide evidence on the presence of threshold behavior in real exchange rates for several currencies but with the caveat that real exchange rates are still reasonably persistent when far away from PPP. This finding on persistence is similar to the findings of O'Conell (1998b) but contrary to Taylor et al. (2001), who employ a similar approach to modeling nonlinearity. The main reason for the different finding is that this chapter considers the first differences of real exchange rates, while Taylor et al. (2001) consider the levels. The simulation experiments on the power/size of the standard unit root and stationarity tests support the findings in that, these tests have power to detect nonlinear mean reversion in general. Hence, allowing transaction costs may not be able to solve the PPP puzzle alone.

The rest of this chapter is structured as follows. Section 3.2 discusses the issues relating to representation, testing and specification of the STAR model. Section 3.3 discusses nonstationarity and nonlinearity of real exchange rates and presents the simulation results on the power/size properties of the LM type linearity tests, unit root and stationarity tests. The data and empirical results are presented in section 3.4. In section 3.5, the dynamic behavior of real exchange rates is evaluated by analyzing the characteristic roots in different regimes and by estimating the generalized impulse response functions from the fitted ESTAR models. Finally section 3.6 concludes and discusses the implications of the empirical findings.

3.2 Modelling Nonlinearity by Smooth Transition Autoregressive Modes

The nonlinear dynamic behavior of real exchange rates in this chapter is modelled in terms of the STAR models that were discussed in chapter 1. In this section for the sake of completeness a brief overview the model is given. The STAR model for a univariate time series y_t, which is observed at times $t = 1-p, -p, \ldots, -1, 0, 1, \ldots, T-1, T$, is given by

$$y_t = (\pi_{1,0} + \pi_{1,1}y_{t-1} + \ldots + \pi_{1,p}y_{t-p})(1 - F(z_t;\gamma,c))$$
$$+(\pi_{2,0} + \pi_{2,1}y_{t-1} + \ldots + \pi_{2,p}y_{t-p})F(z_t;\gamma,c) + u_t, \tag{3.2}$$

where y_t is a stationary process with disturbances, u_t, which are martingale difference sequences with respect to the history of the time series up to time $t-1$, which is denoted by $\Omega_{t-1} = (y_{t-1}, \ldots, y_{1-p})$. This means that, $E[u_t|\Omega_{t-1}] = 0$. It is usually assumed that the conditional variance of u_t is constant, that is, $E[u_t^2|\Omega_{t-1}] = \sigma^2$. The transition function $F(z_t;\gamma,c)$ is a continuous function that is bounded between 0 and 1. The transition variable z_t can be a lagged endogenous variable, $z_t = y_{t-d}$ for a certain integer $d > 0$, as assumed most of the time in empirical studies. As discussed in chapter 1, the logistic and/or the exponential function are frequently used in empirical studies. Since the STAR models and their specification and estimation are discussed in chapter 1, we will briefly discuss the strategy as applied in this chapter.

In this study the autoregressive (AR) order is selected by a combined use of AIC, BIC, and Ljung-Box statistics for autocorrelation. Whenever these criteria do not agree on the appropriate lag order, the highest lag number is selected, because a low

AR order may not be able to take care of the possible serial correlation in the series which in turn might lower the power of the non-linearity tests. The usual practice in the literature is to first identify a linear $AR(p)$ model and then to estimate STAR models with the same specified order in each regime. This approach is somewhat problematic as the true AR order in a linear model may not be the same in a nonlinear STAR type of model. Simulation evidence reported in chapter 1 suggests that these criteria may fail to correctly select the true lag order in STAR models. In this chapter, whenever an estimate is found to be statistically insignificant then it has been removed and the model is re-estimated with different AR orders in each regime. Diagnostic tests are used to decide if the removal of a lag is appropriate or not.

Testing linearity against the STAR type of nonlinearity are carried out by use of the $LM-$ tests discussed in chapter 1. Standard, heteroscedasticity robust and outlier robust versions of LM_2, LM_3 and LM_4 are applied in this chapter. To specify the value of the delay parameter, d, the tests are performed for values of d ranging from 1 to 12. Following Teräsvirta (1994) the delay parameter is usually determined by $d = \arg\min P(d)$ for $1 \leq d \leq 12$, where $P(d)$ is the p-value of the LM_3 test. The choice between the LSTAR and the ESTAR model is usually done by a sequence of tests nested within the null hypotheses corresponding to the LM_3 and the LM_4 tests, see Teräsvirta (1994) and Escirbano and Jordå(1999). The type of regime switching implied by the LSTAR model can be convenient for modelling certain economic time series that exhibit asymmetries in terms of expansions and recessions. This is because in the LSTAR model, the two regimes correspond to the small and large values of the transition variable z_t relative to the threshold c. The ESTAR model may be better suited for modelling real exchange rates, as regimes in the ESTAR model are associated with small and large absolute values of the transition variable. In other words, properties of the ESTAR model allow symmetric adjustment of the real exchange rate for deviations above and below the equilibrium level. In the context of real exchange rates both models imply that there are distinct regimes in the exchange rate market, for example, an appreciating regime and a depreciating regime. The LSTAR model implies that real exchange rates behave differently in the two regimes, while the ESTAR model implies that the two regimes have rather similar dynamics, while the transition period can have different dynamics. In this chapter instead of, a priori, excluding LSTAR model as a possible model for the real exchange rates, the LSTAR models are also estimated along with the ESTAR models to check the adequacy of the ESTAR model. In all of the reported cases in section 3.4, the ESTAR model is found to better represent the dynamic behavior of real exchange rates. This way of selecting the appropriate STAR model and delay parameter is quite flexible and in general may be preferable to the strict application of the procedures described in Teräsvirta (1994) and Escirbano and Jordå(1999), as it allows one to compare the estimated models for each of the transition variables and functions. This approach is also suggested by Teräsvirta (1998). Another difference from the studies which apply STAR modelling to exchange rates is that this study estimates STAR type of models with different autoregressive orders in each regime. Given the results from linearity tests, several ESTAR and LSTAR models are estimated by nonlinear least squares (NLS). Under certain regularity conditions, which are discussed in Gallant (1987) Pötcher and Prucha (1997) among others, the NLS estimates are consistent and asymptotically normal. The estimation is performed by using constrained maximum likelihood library of Gauss. The Newton-Raphson algorithm is used in optimization. Apart from the standard diagnostic analysis of residuals the diagnostic tests developed by Eitrheim and Teräsvirta (1996) and van Dijk and Franses (1999) are applied. For details, see chapter 1.

3.3 Nonlinearity, Non-stationarity and Real Exchange Rates

The application of linearity tests and of the STAR models presumes stationary time series. An issue that deserves particular attention in modelling real exchange rates by STAR type models involves the treatment of non-stationarity. The recent empirical literature argues that standard unit root tests fail to detect mean reverting behavior of real exchange rates as the the true data generating mechanism (DGP) for the real exchange rates follow a nonlinear model of the STAR type. This idea rests on the following re-parameterization of the real exchange rates;

$$\Delta q_t = (\alpha + \rho q_{t-1} + \sum_{j=1}^{p-1} \pi_{1,j} \Delta q_{t-j})(1 - F(z_t, \gamma, c)) + \\ (\alpha' + \rho' q_{t-1} + \sum_{j=1}^{p-1} \pi_{2,j} \Delta q_{t-j}) F(z_t, \gamma, c) + u_t. \tag{3.3}$$

Note that equation (3.3) indicates that when the process is in the middle regime, (that corresponds to $F(.) = 0$ in the ESTAR model) the behavior of real exchange rates is mostly determined by the value of ρ and when the process is in the outer regime (that corresponds to $F(.) = 1$in the ESTAR model) the behavior is mostly determined by the value of ρ'. Hence, for small deviations from PPP the coefficient ρ will govern the adjustment process whereas for large deviations from PPP the coefficient ρ' becomes more and more important. In this sense, STAR models of the form (3.3) are consistent with the predictions of equilibrium models of real exchange rate determination in the presence of transactions costs. In particular, the larger the deviation from PPP, the stronger the tendency to move back to equilibrium, provided that the estimates of ρ and ρ' are such that ρ is even positive while ρ' is negative. These conditions will ensure the global stationarity of the real exchange rates generated from model in (3.3). If the true DGP of real exchange rates is given by the model in (3.3), then unit root tests which are based on a linear $AR(p)$ model of the augmented Dickey-Fuller regression form

$$\Delta q_t = (\alpha^* + \rho^* q_{t-1} + \sum_{j=1}^{p-1} \pi_j^* \Delta q_{t-j}) \tag{3.4}$$

may not be able to detect the mean reverting behavior of real exchange rates, as the estimates of the parameter ρ^* in (3.4) will tend to be a combination of ρ and ρ'. Thus, failure to reject the unit root hypothesis on the basis of a linear model does not necessarily invalidate long-run PPP. That is, the unit root hypothesis $H_0 : \rho^* = 0$ may not be rejected against the stationary linear alternative hypothesis $H_1 : \rho^* < 0$, even though the true DGP is a nonlinear globally stable process. Given this possibility of non-rejection of the unit root hypothesis when in fact the true process is nonlinearly mean reverting, it is worthwhile to investigate the frequency with which the hypothesis of a unit root can be rejected using standard test procedures when, under the null hypothesis, the data generating process is a mean reverting STAR process. This may shed some light on understanding the power/size properties of the standard tests and may reveal information on the reasons why previous research has resulted in non-rejection of unit root null or rejection of stationary null for real exchange rates over the floating period.

Since, a priori, it is not known, whether or not real exchange rates are stationary, it is also worthwhile to investigate the frequency with which the hypothesis of nonlinearty is rejected when the true DGP is a linear unit root and/or stationary process. This is important as the linearity tests and estimation of STAR models assume that the time series under study is stationary. Results of this experiment combined with the results of the experiment on the power/size of unit root/stationarity tests will guide us in testing and estimating the STAR models in the subsequent sections.

To investigate the size of linearity tests, data is generated from $AR(p)$ model. To investigate the power/size properties of unit root and stationarity tests the data is generated from the $ESTAR$ model with $p = 1$ and $p = 2$. The parameters in

ESTAR models are specified so that the generated series are globally stationary even though they may behave as a random walk in the middle regime. In all experiments, disturbances are generated from independent and identically distributed Gaussian innovations with zero mean and unit variance. Starting values are set equal to zero and in each replication the first 100 observation is discarded in order to remove the possible effects of starting values. A sample size of 305 observations is generated from $AR(p)$ and $ESTRAR(p)$ models as this corresponds to the sample size used in this study. The results are given in tables 3.1 and 3.2. Table 3.1 gives the empirical rejection frequencies of the F variants of LM type tests. Linearity tests and corresponding $p-$values are computed and compared with the 5% significance level. Both levels and first differences are used in computing the tests. The first values in the table are the empirical size of tests when the level of the generated data is used while the values in the square brackets correspond to the size of tests when first difference of the data is used. Tests are computed given the true lag order of 2. Experiments are conducted with different p values. Since the results are similar only results from $p = 2$ are reported. The results from table 3.1 indicate that for the values of the AR parameter which make the $AR(p)$ model stationary the standard versions of $LM-$type tests have estimated empirical sizes closer to the nominal size of 5%. As the the coefficients in $AR(p)$ processes take values so that the processes become near unit root or a pure unit root process the empirical size of the tests worsens and becomes unity. This means that the $LM-$type tests may spuriously suggest presence of nonlinearity even though the true DGP is a linear process. The results also indicate that first differencing the series in general improve the size of the tests.

The results in table 3.2 indicate that the ability of Phillips-Perron (1988) (PP), Augmented Dickey-Fuller (ADF) and KPSS tests to reject nonstationarity when nonstationarity is false depend on the parametric specification for the true data generating process (DGP). When the true DGP is a STAR model with near unit root or unit root behavior in the middle/inner regime and stationary in the outer regime such that the process is globally stable then the unit root tests and stationarity tests have good power and size properties in terms of detecting global stationarity of the series. However, when the root of the autoregressive parameter in the outer regime approaches unity then the ability of ADF and PP tests declines in detecting nonlinear mean reversion. This indicates that the power of the ADF and PP tests depend on the behavior of the process in the outer regime as the global behavior of the time series in an ESTAR model is dictated by the roots of the autoregressive polynomial in the outer regime. As the autoregressive parameter(s) in the outer regime approaches to unity, the ESTAR model becomes more and more persistent and hence the ADF and the PP lose power in detecting the global stationarity of the process while the power of KPSS rises as KPSS has power against persistent but stationary alternatives.

3.4 Empirical Results

3.4.1 The Data

The data used in this study consists of monthly observations on consumer price indices for Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Switzerland, the UK, and the US and end-of-period spot exchange rates for Belgian franc, Canadian dollar, French franc, German mark, Italian lira, Japanese yen, Dutch guilder, Swiss franc, the UK pound against the US dollar. All data cover the sample period from 1973M03 to 1998M07 and derived from the International Monetary Fund's *International Financial Statistics* data compact disks. The logarithmic real exchange rate series constructed with these data as in equation (3.1), with s_t taken as the logarithm of the dollar price of currency, p_t as the logarithm the US price level, and p_t^* as the logarithm of the price level of the relevant country.

PP, due to Phillips, and Perron (1988), KPSS, due to Kwiatkoski, Phillips Schmidt, and Shin (1992), statistics in both levels and first differences are used to evaluate the nonstationarity-stationarity nature of real exchange rates. The results are given in Table 3.3. The results from the table indicates that for all series the real exchange rates are non-stationary, and clearly have a unit root. The log differenced real exchange rates are all stationary. Combined with the results from the simulation experiments reported above the first difference logarithmic real exchange rates are going to be used in analyzing the nonlinear behavior of the real exchange rate series over the free floating period in the rest of the study.

3.4.2 Nonlinearity tests and STAR model specification

The $p-$values for linearity tests with the maximum AR lag determined by combined use of AIC, BIC and LB statistics, are reported in table 3.4. Following the suggestion in Teräsvirta (1994, 1998) F-variants of linearity tests are used as they have more power in finite samples. Each table gives three versions of each of the LM-type tests discussed above. Each row in table 3.4 gives the transition variable(s) for which at least one of the p-values from any version of the test is less than 0.10. One of the striking result from table 3.4 is that for some of the currencies (especially for Belgian franc, the British pound, Dutch guilder, French franc, Italian lira and Japanese yen) the standard variant of the tests indicate presence of very significant nonlinearity while either HCC or OR or both variants have highly insignificant p-values, indicating either the results from LS variants may be spurious in the sense that a finding of nonlinearity possibly due to either presence of heteroscedasticity, outliers or both, or robust variants are not able to detect nonlinearity. There is almost no evidence of nonlinearity at any reasonable level of significance for the British pound and Swiss franc for the sample in this study from HCC variants of the tests. For all other currencies either some or all of the tests indicate the presence of STAR type of nonlinearity at either 5% or 10% significance levels. In some of these cases evidence from HCC and/or OR versions of nonlinearity tests on the presence of STAR form nonlinearity is not very strong. In these cases it is not clear how to conclude about the presence of nonlinearity. An approach is to estimate STAR models for all of the delay parameters for which nonlinearity is suggested by the LS versions of the nonlinearity tests and then let the diagnostic and specification tests reveal the relevance of the nonlinear model for the data. This approach is intuitive, because if there is no STAR type of nonlinearity in the data, either the estimation procedure would fail (indicating threshold type of nonlinearity is not being identified) or else, in the case of curve fitting, the fitted model would fail to pass at least some of the diagnostic and specification tests. This is the approach taken in the remaining part of this chapter.

3.4.3 Results from the Estimated STAR Models

For all currencies, both ESTAR and LSTAR models are estimated for each of the transition variable for which some evidence of nonlinearity is obtained from linearity tests. LSTAR models are used for comparison purposes to check if the ESTAR models appropriately model the dynamics of real exchange rates as suggested by economic intuition. Consistent with the intuition, in all cases the ESTAR model is found to represent the dynamics better than the LSTAR model. The estimated models for the Belgian franc, British pound, Dutch guilder and Swiss franc either failed in the estimation stage or failed to pass the diagnostic tests, especially the presence of remaining nonlinearity and presence of serial correlation tests. Hence no results for these currencies are reported in the following. The selection of the model with the appropriate transition variable is done by use of diagnostic statistics. The use of diagnostic tests in selecting the appropriate transition variable and function is quite flexible and in general should be preferred as it allows one to compare the estimated models for each of the transition variables and functions. For example for the French franc and Italian lira the LS versions of the tests indicated the presence of strong nonlinearity especially at $d = 1$ while other versions suggested that these findings are probably due to the presence of heteroscedasticity or outliers. Despite this, both LSTAR and ESTAR models were estimated with $d = 1$ and it was found that there were considerable nonlinearities left out for higher delay parameters, and significant correlations are found in the residuals. Hence these and several other estimated models were discarded as they failed to pass the diagnostic tests. On the other hand, for the German mark, consistent with the results of the LS variant of linearity tests, the ESTAR model with delay parameter $d = 1$ is found to be the best one. STAR models of the form given in (3.3) are estimated without any restriction. The hypothesis that the process is white noise in the outer regime as suggested by economic theory, is tested by testing the null of, $H_0 : \rho^* = -1,\ \pi_1^* = \cdots = \pi_p = 0$, in (3.3). This hypothesis implies that real exchange rates, although they can behave as random walks or even have explosive paths within the neighborhood of a threshold level, become increasingly mean reverting with the absolute size of the deviations from equilibrium level. In all of the estimated models this hypothesis is rejected significantly. Those parameters which are found to be nonsignificant are deleted and the model is re-estimated. The model best fits the data in terms of adequate diagnostic properties selected and reported.

Tables 3.5 and 3.6 present the results from five of the countries. The ESTAR model is found to be an adequate representation for the rates reported. This implies that real exchange rates move from high or low levels towards the middle level or their normal level in a similar fashion. Diagnostic statistics are satisfactory in all cases. The γ estimates vary across countries, with the speed of adjustment for some real exchange rates being much higher than others. The estimated values for γ for all series are found to be significantly different from zero. The estimate of threshold parameter, $\hat{c}$ is found to be indistinguishable from zero.

In order to better evaluate the estimated models, panels of Figure 3.1 display the graphs of the estimated transition function versus time and threshold variable. The figures reveal that transition functions, visit each of the extreme regimes in general. This means that real exchange rates behave in a nonlinear fashion in that they visit extreme regimes quite often and a linear representation that ignores this behavior will not be appropriate to fully understand the dynamic behavior of real exchange rates. It can be observed from the panels of figure 3.1 that the Dutch guilder and Italian lira rates spend most time during the sample period closer to the outer regime, while German mark, Canadian dollar and Japanese yen rates stay closer to the middle regime. The estimated transition functions over threshold variable indicate that transition between regimes is relatively fast. That is to say that real exchange rate differences adjust to shocks rapidly as the slope of the transition functions for all currencies are high. The estimated transition functions in general provide evidence of nonlinearity for all of the series.

3.5 Further Analysis of the Dynamics of Estimated Star Models: Characteristics Roots and GIRFs

To gain some insights into the dynamic behavior of real exchange rates this section examines the dynamics of estimated models first by computing the characteristic roots from estimated equations and second by analyzing the propagation mechanism of shocks to real exchange rate process through use of generalized impulse response functions (GIRF). Characteristic roots are obtained by solving the equation

$$\lambda^p - \sum_{j=1}^{p}[\pi_{1,j}(1 - F(z_t, \gamma, c)) + \pi_{2,j}F(z_t, \gamma, c)]\lambda^{p-j} = 0. \tag{3.5}$$

For illustration two extreme regimes are considered, namely $F = 0$, (middle regime) and $F = 1$ (outer regime). Characteristic roots are computed for the level series. Table 3.7 gives roots for each regime. The striking result is that for all of the series the modulus is equal to unity in the middle regime. This implies that the real exchange rates will behave as if they are a unit root process in this regime. Although for all the series, the modulus in the outer regime is less then one, albeit they are very close to unity. This implies that, although real exchange rates tend toward the stationary equilibrium as time passes, the speed with which they tend to the equilibrium level is very slow. In other words, when ta real exchange rate is in the outer regime it will adjust towards its equilibrium level, but most probably the size of the adjustment is very small hence it takes for a long time for the real exchange rate to revert back to its respective equilibrium path. The rest of this section further investigates this implied persistence by means of GIRFs developed by Koop et al. (1996).

Impulse response functions (IRF) for a linear model and a nonlinear model are different. An IRF for a linear model is *symmetric,* as such a shock of size $-\delta$ has an effect that is exactly opposite to that of a shock of size $+\delta$. Moreover, it is *linear* in the sense that the IRF is proportional to the size of the shock. Lastly, it is *history independent* as its shape does not depend on the particular history ω_{t-1}. As discussed in Koop et al. (1996) and Pesaran and Potter (1997), in general, properties of IRFs from a linear model do not carry to IRFs from a nonlinear model. Koop et al. (1996) show that the impact of a shock depends not only on the history of the process but also on the sign and size of the shock when the time series follows a nonlinear process such as a STAR model. Furthermore, as shown in Pesaran and Potter (1997), when one wants to analyze the effect of a shock on the time series $k > 1$ periods ahead, the assumption that no shocks occur in the intermediate periods may give misleading inference concerning the propagation mechanism of the model. GIRF for a specific shock $u_t = \delta$ is defined as

$$GI_y(k, \delta, \omega_{t-1}) = E[y_{t+k} \,|u_{\,t} = \delta, \omega_{t-1}] - E[y_{t+k} \,|\omega_{\,t-1}], \qquad (3.6)$$

for $k = 1, 2, \cdots$. Note that the expectations of y_{t+k} are conditioned only on the history and/or on the shock. In other words, the problem of dealing with shocks occurring in the intermediate periods is dealt with by averaging them out. That explains also why the benchmark profile is the expectation of y_{t+k} given only the history of the process ω_{t-1}. Therefore, in the benchmark profile the current shock is averaged out as well. This GIRF reduces to traditional IRF when the model is linear. Koop et al. (1996) emphasize that the GIRF given in (3.6) is indeed a random variable. The GIRF is a function of δ and ω_{t-1}, which are realizations of the random variables u_t and the information set, Ω_{t-1}.

The GIRFs can be utilized in several ways in analyzing the dynamic properties of the estimated model. They can be used to analyze the persistence of shocks. A shock $u_t = \delta$ is called *transient* at history ω_{t-1} if $GI_y(k, \delta, \omega_{t-1})$ becomes zero as $k \to \infty$. If on the other hand, $GIRF$ approaches a non zero finite value when $k \to \infty$ then the shock is said to be *persistent*. It is intuitive to think that if a time series process is stationary and ergodic, the effects of all shocks eventually converge to zero for all possible histories of the process. Hence the distribution of $GI_y(k, \delta, \omega_{t-1})$ collapses to a spike at 0 as $k \to \infty$. In contrast, for non-stationary time series the dispersion of the distribution of $GI_y(k, \delta, \omega_{t-1})$ is positive for all k. Koop et al. (1996) suggest that the dispersion of the distribution of $GI_y(k, \delta, \omega_{t-1})$ at finite horizons conveniently can be used to obtain information about the persistence of shocks. GIRFs can also be used to assess the significance of asymmetric effects over time. One difficulty in computing the GIRFs is that the analytic expressions for the conditional expectations are not available for $k > 1$. Therefore they need to be estimated. Koop et al. (1996)discusses in detail simulation methods to estimate GIRFs. In particular Monte Carlo or bootstrap methods are suggested for computation of GIRFs. In this study, conditional expectations are simulated realizations that are obtained from iteration of the estimated ESTAR model, randomly by drawing with replacement from the estimated residuals of the model, and then averaging over 5000 random draws over

$h = 0, 1, 2, \cdots, 60$. For each combination of history and initial shock, we compute generalized impulse responses for horizons $k = 1, 2, \cdots, N$ with $N = 60$. More explicitly, the conditional expectations in (3.9) are estimated as the means over 5,000 realizations of Δq_{t+k} with and without using the selected initial shock to obtain Δq_t and using randomly sampled residuals of the estimated ESTAR models elsewhere. All generalized impulse responses are initialized such that they equal $\iota/\hat{\sigma}_u$ at $k = 0$.

There are different ways of obtaining GIRFs. One way is to estimate GIRFs for each history vector. Alternatively one could estimate GIRFs by estimating conditional expectations for each history ω_{t-1} and then average the obtained sequences over all possible drawings from ω_{t-1}. A third way is to estimate GIRFs by setting the conditioning vector to $\omega_{t-1}^0 = E[\omega_{t-1}]$. GIRFs from all of these strategies are computed. The mean GIRFs from histories that correspond to the upper 10 percent quintile of the estimated transition function are given in the panels of figure 3.2. GIRFs are computed for the levels of the real exchange rates by cumulating the impulse responses from the logarithmic difference of the real exchange rates for each horizon. Inspection of the generalized impulse response functions reveal that for all of the series, shocks to innovations in real exchange rates do not dissipate as the horizon increases. That is, consistent with a modulus that is around unity, a shock will have quite persistent effects in that real exchange rates do not return to their equilibrium path in a short period of time. This is in contrast to the argument that real exchange rates should be mean reverting when deviations from the equilibrium level implied by the PPP condition are large. This result indicates that although, the presence of transaction costs may lead to nonlinear type of behavior that can be modelled appropriately by ESTAR models, it does not necessarily imply that real exchange rates are anti-persistent.

3.6 Conclusion

The use of three different nonlinearity tests and their robustified variants against heteroscedasticity and outliers indicated presence of STAR type nonlinearities at different transition variables for most of the currencies considered in this study. The results from nonlinearity tests also revealed the importance of evaluating the estimated STAR model in different respects, as a finding from nonlinearity tests may be due to some other property of the data. In turn, several different diagnostic tests are utilized in evaluating the estimated STAR models. For the Belgian franc, British pound, and French franc rates, estimated models did not pass all the diagnostic tests, especially tests of remaining nonlinearity and tests for serial correlation in the residuals despite the evidence of nonlinearity from the LM tests.

Further evaluation of the dynamic behavior of real exchange rates from estimated STAR models revealed that shocks to real exchange rates have quite persistent effects which is consistent with a non-stationary process. This finding is consistent with the results of the simulation experiments on the power and size of PP, ADF and KPSS statistics which indicated that unit root and stationarity tests are capable of detecting a globally stationary process even if the true DGP is a nonlinear one. The findings here support the findings of O'Connell (1998b), in that small deviations from PPP can be as persistent as large deviations. The identified threshold type of nonlinearity may indicate that a certain component of real exchange rates may have the tendency to behave as nonlinearly mean reverting but apparent persistence indicates that either the nominal exchange rates or the relative prices converge too slowly. As such, the presence of transaction costs by themselves are not able to induce real exchange rates converge to long run equilibrium levels. The general equilibrium models that incorporate transaction costs, such as Dumas (1992) and Sercu et al. (1995) indicate that real exchange rates spend most of the time away from equilibrium. Still, they presume that relative prices and nominal exchange rates converge to the long run equilibrium at the same rate. Since in these models adjustments in relative prices are the main force that cause real exchange rates to revert to equilibrium, the findings here raise the question of why adjustments in relative prices are not able to induce real exchange rates to move to equilibrium faster? Perhaps, as argued by Engel and Morley (2001) nominal exchange rates and relative prices have different speeds of adjustment and persistence of real exchange rates can be explained by persistence of nominal exchange rates rather than relative prices. An interesting issue that may worth investigating is the persistence and nonlinear behavior of nominal exchange rates and relative prices separately as this may reveal important information on the adjustment dynamics and speed with which nominal exchange rates and relative prices converge to their long run equilibrium levels. Given the observed strong correlation between nominal and real exchange rates it is possibly the relative prices that have the threshold type of mean reversion rather than the nominal exchange rates. This issue is left for future research.

BIBLIOGRAPHY

[1] Baum, C., Berkoulas, and M. Caglayan (2001), Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era, *Journal of International Money and Finance* **20,** 379-399.

[2] Diabold, F. X., S. Husted, and M. Rush 1991. Real exchange rates under the gold standard. Journal of Political Economy 99 (6), 1252-1271.

[3] Diabold, F. X., and A. Inoue 1999. Long memory and structural change, Working paper, Stern School of Business, NYU.

[4] Dumas, B. (1992), Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**, 153–180.

[5] Engel, C., and J.C. Morley (2001) The adjustment of prices and the adjustment of the exchange rate, NBER Working Paper 8550, Cambridge, MA.

[6] Eitrheim Ø. and T. Teräsvirta (1996), Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–76.

[7] Escirbano, A. and O. Jordå(1999), Improved testing and specification of smooth transition regression models, in P. Rothman (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston: Kluwer, pp. 289–319.

[8] Frankel, J.A., and A.K. Rose (1996), Mean reversion within and between countries: a panel project on purchasing power parity, *Journal of International Economics* **40,** 2-9-224.

[9] Froot, K.A., K., Rogoff (1995), Perspectives on PPP and long-run real exchange rates, in Grossman, G., Rogoff, K. (Eds.), *Handbook of International Economics*, North-Holland, Amsterdam (chap 32).

[10] Gallant, A. R. (1987), Nonlinear Statistical Models, New York: John Wiley.

[11] Granger, C.W.J. and T. Teräsvirta (1993), Modelling Nonlinear Economic Relationships, Oxford: Oxford University Press.

[12] Koop, G., M. H. Pesaran and S. M. Potter (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119-147.

[13] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* **54,** 159-178.

[14] Lothian, R. (1997), Multi-country evidence on the behavior of purchasing power parity under the current float, *Journal of International Money and Finance* **16**, 19-35.

[15] Lundbergh, S., T. Teräsvirta (1998) Modelling economic high-frequency time series with STAR-GARCH models, *Working papers in Economics and Finance* **291**, Stockholm School of Economics.

[16] Lundbergh, S., T. Teräsvirta and D. van Dijk (1999), Time-varying smooth transistion autoregressive models, Stockholm School of Economics, unpublished muniscript.

[17] Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988), Täesting linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–9.

[18] Micheal, P., R.A. Nobay, and D.A. Peel (1997), Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation, *Journal of Political Economy* **105**, 862–879.

[19] Obstfeld, M. and A. M. Taylor (1997), Nonlinear aspects of goods-market arbitrage and adjustment: Hecksher's commodidty points revisited, *Journal of the Japanese and International Economics* **11**, 441–479.

[20] Obstfeld, M. K. *Rogoff, Foundations of International Macroeconomics,* MIT Press, Cambridge, Massachusetts.

[21] O'connel, P.G.J. (1998), The overvaluation of purchasing power parity, *Journal of International Economics* **44,** 1-19.

[22] O'Connel, P.G.J. (1998), Market frictions and real exchange rates, *Journal of International Money and Finance* **17**, 71–95.

[23] Oh, K.Y. (1996), Purchasing power parity and unit root tests using panel data, *Journal of International Money and Finance* **15,** 405-418.

[24] Pedroni, P. (1995), Panel cointegration: asymptotic and finite sample propereties of pooled time series tests with an application to the PPP hypothesis, Unpublished Working Paper 95-013, Department of Economics, Indiana University, Bloomington, IN.

[25] Pesaran, M. H. and S. M. Potter (1997), A floor and ceiling model of US output, *Journal of Economic Dynamics and Control* **21**, 661-695.

[26] Phillips, P. and C.B.P. Perron (1988), Testing for a unit root in time series regression, *Biometrika* **75,** 335-346.

[27] Pötcher, B.M. and I.V. Prucha (1997), Dynamic Nonlinear Econometric Models-Asymptotic Theory, Berlin: Springer-Verlag

[28] Rogoff (1996), The purchasing power parity puzzle, *Journal of Economic Literature* **34,** 647-668.

[29] Sarantis, N. (1999), Modelling non-linearities in real effective exchange rates, *Journal of International Money and Finance* **18**, 27–45.

[30] Sarno, L. (2000), Systematic sampling and real exchange rates, *Weltwirtschaftliches Archiv* **136,** 24-57.

[31] Sarno, L. and M. P. Taylor (1999), The economics of exchange rates, Cambridge and Newyork: Cambridge University Press.

[32] Sercu, P., R. Uppal, and C. Van Hulle (1995), The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity, *Journal of Finance* **10**, 1309–19.

[33] Taylor, M.P., D.A. Peel, and L. Sarno (2001), Non-linear in real exchange rates: towards a solution of the purchasing power parity puzzles, Working Paper, Centre for Economic Policy Research, London, UK.

[34] Teräsvirta, T. (1994), Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

[35] Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (editors), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

[36] van Dijk, D., P.H. Franses and A. Lucas (1999), Testing for smooth transition nonlinearity in the presence of additive outliers, *Journal of Business and Economic Statistics* **17**, 217–235.

[37] van Dijk, D., and P.H. Franses (1999), Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics* **3**, 311-40.

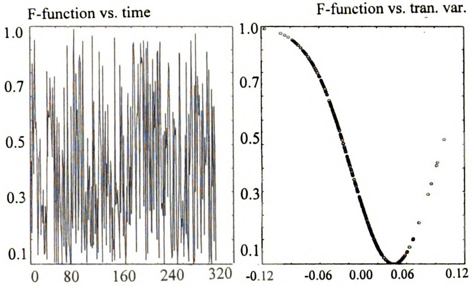
[38] Wooldridge, J.M. (1990), A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17–43.

[39] Wooldridge, J.M. (1991), On the application of robust, regression-based specification tests, *Journal of Econometrics* **47**, 5–46.

[40] Wu, Y. (1996), Are real exchange rates nonstationary? Evidence from panel-data tests, *Journal of Money Credit, and Banking* **28**, 54-63.

Figure 3.1: Estimated Transition Function versus Time and Threshold Variable

(a)Canadian Dollar



(b)Dutch Guilder

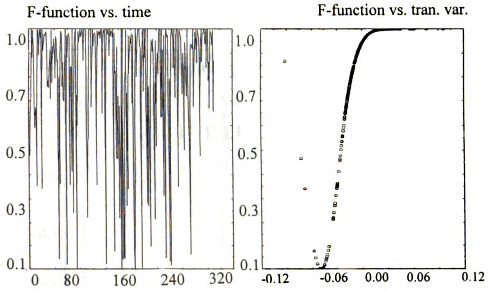
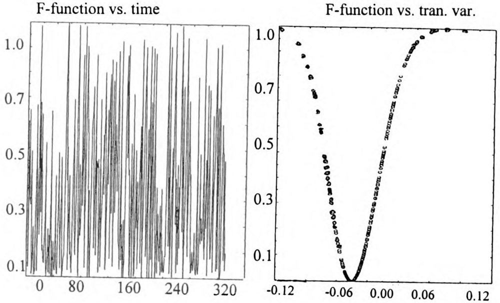


Figure 3.1 (cont'd).

(c) German Mark



(d) Italian Lira

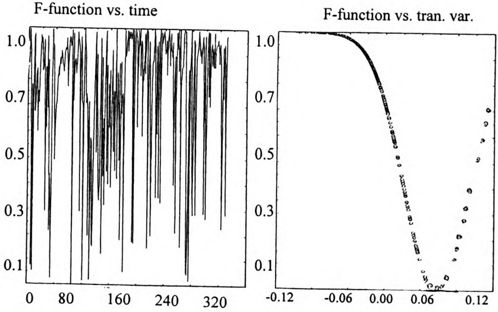


Figure 3.1 (cont'd).

(e) Japanese Yen

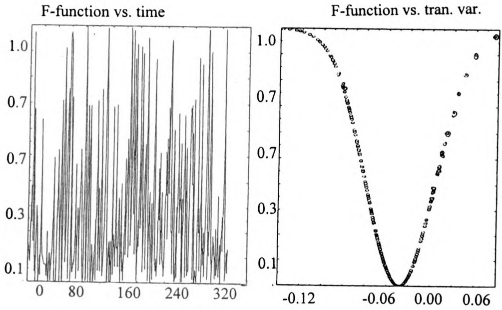
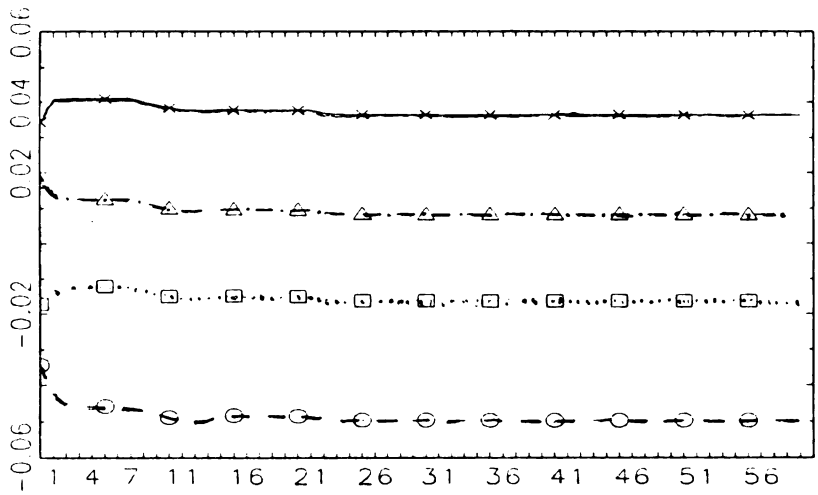


Figure 3.2: Generalized Impulse Response Functions from estimated ESTAR models
(a) Canadian Dollar



(b) Dutch Guilder

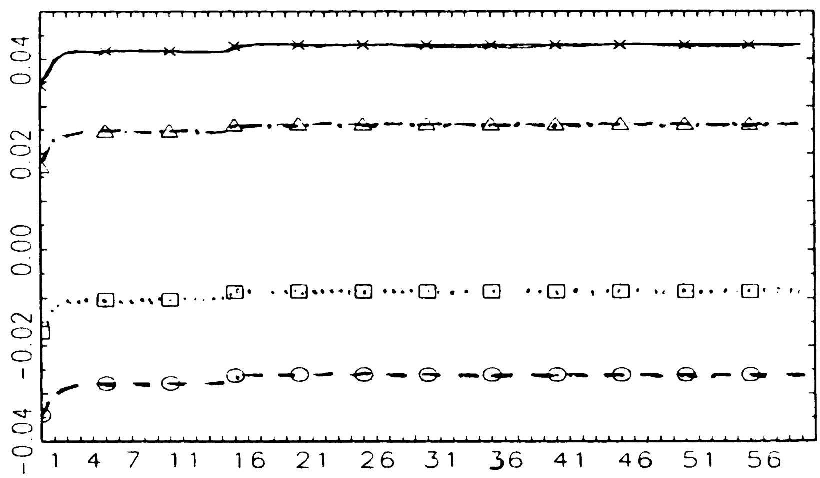
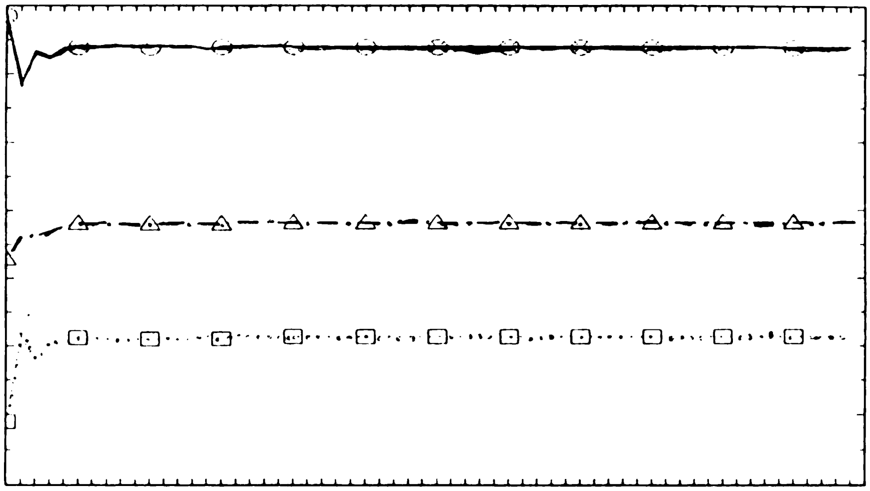


Figure 3.2 (cont'd)

(c) German Mark



(d) Italian Lira

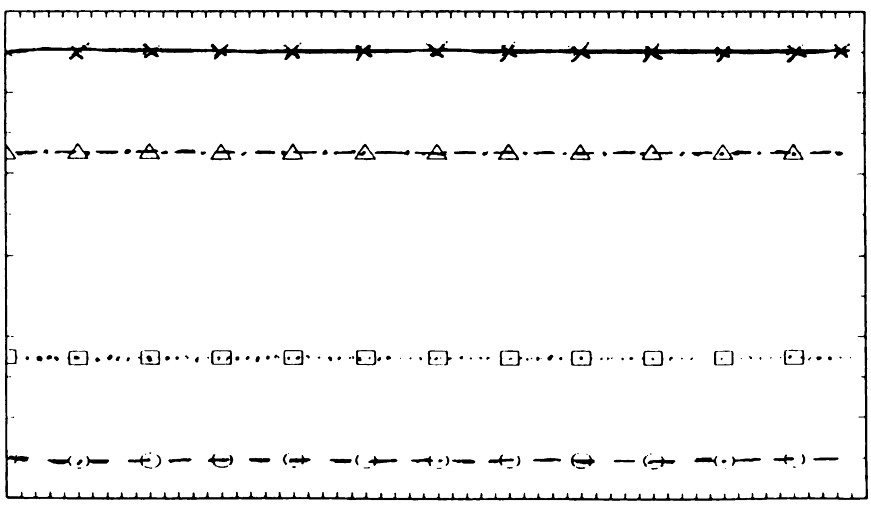
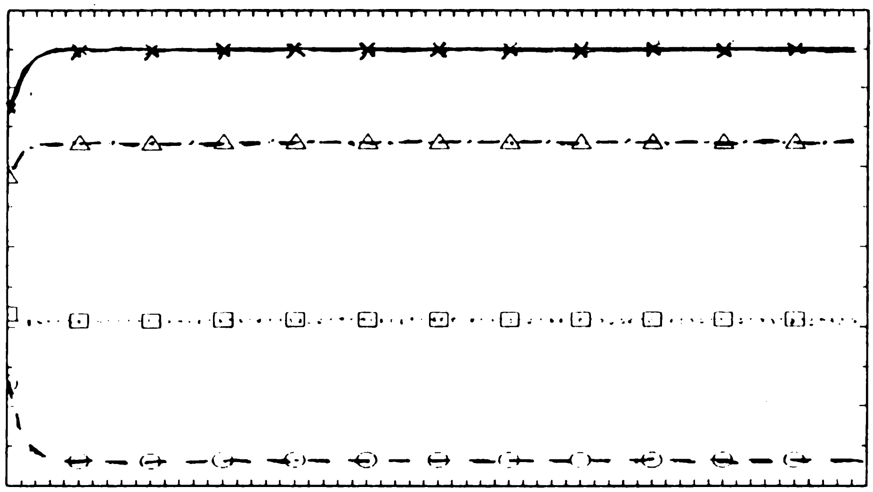


Figure 3.2 (cont'd).

(e) Japanese yen



Note: The mean GIRFs from shocks of 10%, (solid lines with star), 5%, dotted lines with triangles), -5% (dots with squares), and -%10(dashes with circles) are given for the histories that correspond to the outer regime. Note that shocks are standardized by dividing the standard error of the residuals from estimated models.

Table 3.1: Empirical rejection frequencies of linearity tests, Sample size=305.
Model Design: $y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t$, $u_t \sim i.i.d.N(0,1)$

Parameter	Rejection frequencies		
	LM_2	LM_3	LM_4
$\rho_1 = 0.3$, $\rho_2 = 0.6$	0.077[0.041]	0.067[0.040]	0.064[0.044]
$\rho_1 = 1.0$, $\rho_2 = 0.0$	0.105[0.044]	0.098[0.039]	0.108[0.046]
$\rho_1 = 0.7$, $\rho_2 = 0.3$	0.110[0.047]	0.090[0.048]	0.097[0.049]
$\rho_1 = 0.3$, $\rho_2 = 0.7$	0.292[0.045]	0.262[0.043]	0.247[0.049]
$\rho_1 = 0.5$, $\rho_2 = 0.5$	0.193[0.046]	0.162[0.040]	0.165[0.045]
$\rho_1 = 0.7$, $\rho_2 = 0.4$	0.999[0.997]	1.000[1.000]	1.000[1.000]

Notes: The rejection frequencies are obtained computation F variants of LM tests and corresponding p-values 5000 times. Since the true data generating model is linear these frequencies indicate the empirical sizes of the tests. The nominal significance level taken is %5. Squared bracketed values correspond to the first differenced series.

Table 3.2: Empirical rejection frequencies for ADF PP and KPSS tests
Model Design:
$y_t = \pi_{1,1} y_{t-1}(1 - F(y_{t-1}, 5, 0)) + [\pi_{1,2} y_{t-1} F(y_{t-1}, 5, 0)] + u_t$, $u_t \sim iidN(0,1)$

Parameter specification	Rejection frequency		
	KPSS	PP	ADF
$\pi_{1,1} = 0.9, \pi_{1,2} = -0.5$	0.067	0.990	0.970
$\pi_{1,1} = 1, \pi_{1,2} = -0.5$	0.071	0.899	0.900
$\pi_{1,1} = 1, \pi_{1,2} = -0.1$	0.355	0.997	0.990
$\pi_{1,1} = 1.1, \pi_{1,2} = -0.5$	0.085	0.994	0.991
$\pi_{1,1} = 1.2, \pi_{1,2} = -0.5$	0.120	0.991	0.995
$\pi_{1,1} = 1.0, \pi_{1,2} = 0.5$	0.800	0.845	0.840
$\pi_{1,1} = 1.0, \pi_{1,2} = 0.7$	0.870	0.835	0.830
$\pi_{1,1} = 1.0, \pi_{1,2} = 0.95$	0.850	0.540	0.520
$\pi_{1,1} = 1.1, \pi_{1,2} = 0.95$	0.880	0.480	0.475
Model Design: $y_t = [\pi_{1,1} y_{t-1} + \pi_{1,2} y_{t-2}](1 - F(y_{t-1}, 5, 0)) + [\pi_{2,1} y_{t-1} + \pi_{2,2} y_{t-2}]F(y_{t-1}, 5, 0) + u_t, u_t \sim iidN(0,1)$			
	KPSS	PP	ADF
$\pi_{1,1} = 0.6$, $\pi_{1,2} = 0.4$, $\pi_{2,1} = 0.4, \pi_{2,2} = -0.6$	0.104	0.890	0.992
$\pi_{1,1} = 0.4$, $\pi_{1,2} = 0.6$, $\pi_{2,1} = 0.4, \pi_{2,2} = -0.6$	0.344	0.995	0.994
$\pi_{1,1} = 0.7$, $\pi_{1,2} = 0.3$, $\pi_{2,1} = 0.4, \pi_{2,2} = -0.6$	0.059	0.996	0.992
$\pi_{1,1} = 0.3$, $\pi_{1,2} = 0.7$, $\pi_{2,1} = 0.4, \pi_{2,2} = -0.6$	0.613	0.998	0.993
$\pi_{1,1} = 0.3$, $\pi_{1,2} = 0.7$, $\pi_{2,1} = 0.4, \pi_{2,2} = 0.4$	0.815	0.722	0.720
$\pi_{1,1} = 0.3$, $\pi_{1,2} = 0.7$, $\pi_{2,1} = 0.6, \pi_{2,2} = 0.3$	0.828	0.718	0.715

Note: Rejection frequencies are based on 5000 replications.

Table 3.3: Results on unit root and stationarity tests:PP, and KPSS

Currency	level		first difference	
	PP	KPSS	PP	KPSS
Belgian franc	-1.351	0.997	-16.299	0.091
Canadian dollar	-1.504	2.812	-14.253	0.180
French franc	-1.534	1.354	-17.046	0.206
German Dmark	-1.882	3.217	-16.259	0.166
Italian lira	-2.589	3.239	-15.102	0.438
Japanese yen	-0.483	3.695	-12.532	0162
Dutch guilder	-1.397	3.088	-16.612	0.100
Swiss franc	-2.226	3.205	-15.950	0.228
British pound	-2.941	2.706	-11.586	0.312

Notes: The reported values for the PP test are based on the regression of the time series on a constant and its lagged value. The lag truncation for the Bartlett kernel is obtained from the formula $floor(4(\frac{T}{100})^{2/9})$. The 1% and 5% critical values are -3.454 and -2.871 respectively for the PP tests. The reported values for the KPSS test are based on a regression of the series on a constant only. The 1% and 5% critical values for the KPSS tests are 0.739 and 0.463 respectively. PP statistic test the null hypothesis of a unit root against the alternative of stationarity while the KPSS statistic has the null of covariance stationarity against non-stationarity.

Table 3.4: p-values of LM tests for star type of nonlinearity in monthly logarithmic differences of real exchange rates.

d	LS			HCC			OR		
	LM_2	LM_3	LM_4	LM_2	LM_3	LM_4	LM_2	LM_3	LM_4
Belgian franc, $p = 2$									
1	0.0094	0.0005	0.0040	0.3631	0.5446	0.2361	0.7686	0.5849	0.0290
9	0.0828	0.1255	0.0628	0.2762	0.4011	0.2579	0.0318	0.0182	0.0597
11	0.1208	0.2529	0.0912	0.1143	0.2820	0.0851	0.0134	0.0016	0.0188
British pound, $p = 3$									
3	0.0478	0.1351	0.1260	0.5300	0.8183	0.3079	0.2695	0.0433	0.0671
5	0.0663	0.1536	0.0242	0.3113	0.5744	0.1923	0.3186	0.0492	0.4010
Canadian dollar, $p = 1$									
8	0.0971	0.2434	0.0964	0.0699	0.1791	0.0728	0.2052	0.0797	0.3906
10	0.2462	0.0970	0.0792	0.3092	0.3097	0.1735	0.0778	0.0623	0.1340
Dutch guilder, $p = 2$									
1	0.0199	0.0007	0.0096	0.4889	0.5581	0.3328	0.4807	0.1423	0.0790
9	0.2268	0.2120	0.1761	0.4380	0.7388	0.4993	0.0467	0.0640	0.0688
11	0.0740	0.1985	0.0468	0.1161	0.2429	0.0691	0.0350	0.0035	0.0477
French franc, $p = 1$									
1	0.0575	0.0147	0.0575	0.1025	0.1153	0.1025	0.0453	0.0923	0.0627
5	0.4571	0.1114	0.3617	0.2203	0.0346	0.2047	0.0254	0.0514	0.0697
11	0.1462	0.0703	0.0468	0.3514	0.3636	0.2592	0.0108	0.0026	0.0288
German mark, $p = 1$									
1	0.0032	0.0001	0.0032	0.0723	0.1506	0.0723	0.0373	0.3271	0.0032
5	0.1411	0.0331	0.3912	0.0454	0.0404	0.3588	0.0383	0.4653	0.0863
9	0.1719	0.2021	0.2524	0.2533	0.4151	0.5244	0.0175	0.0632	0.0475
Italian lira, $p = 2$									
1	0.0278	0.0027	0.1901	0.1643	0.0422	0.5446	0.0023	0.0040	0.0021
7	0.0377	0.0150	0.0039	0.1813	0.1709	0.0863	0.0071	0.0043	0.0164
9	0.0228	0.0589	0.0450	0.1479	0.2457	0.0870	0.0217	0.0058	0.0360
11	0.0512	0.1480	0.0557	0.1044	0.2245	0.0787	0.0620	0.0166	0.0901
Japanese yen, $p = 3$									
1	0.0387	0.0538	0.1588	0.0929	0.0768	0.2206	0.2055	0.2918	0.1987
8	0.1970	0.4093	0.1128	0.0814	0.2169	0.0500	0.1797	0.0255	0.2454
11	0.3895	0.1872	0.1080	0.1504	0.0931	0.0596	0.0746	0.0452	0.1076
Swiss franc, $p = 1$									
4	0.0294	0.0904	0.1872	0.1862	0.1262	0.1981	0.4382	0.5533	0.2568
12	0.2384	0.0445	0.2205	0.5011	0.1578	0.4964	0.0920	0.8255	0.1221

Key:LS, HCC, and OR stand for Least squares, Heteroscedasticity Consistent and Outlier Robust variants of the LM tests described in the paper. The column d gives those delay parameters, and hence the transition variables, for which most of the p-values from three variants of LM-type tests are less than 0.1.

Table 3.5: Estimation Results from ESTAR models: Sample size: 291 (after adjusting end points).

Parameters	Parameter Estimates for each currency				
	CD	DG	GM	IL	JY
$\pi_{1,0}$	0.003	-0.073	.	.	.
	(0.001)	(0.034)	.	.	.
$\pi_{1,1}$	0.271	-1.138	0.610	0.592	0.223
	(0.103)	(0.520)	(0.158)	(0.302)	(.128)
$\pi_{2,0}$	0.002	0.013	0.035	0.063	.
	(0.001)	(0.008)	(.017)	(0.026)	.
ρ'	-0.024	-0.010	-0.034	-0.008	-0.002
	(0.012)	(0.007)	(0.017)	(0.004)	(0.001)
$\pi_{2,1}$	.	.	-0.385	.	0.425
	.	.	(0.187)	.	(0.179)
γ	25.091	21.473	15.508	12.578	6.661
	(1.116)	(0.935)	(0.495)	(1.494)	(2.636)
c	0.016	-0.067	-0.002	0.077	0.046
	(0.132)	(0.483)	(0.113)	(0.594)	(0.340)
Skewness	-0.036	0.310	0.184	0.587	-0.558
Kurtosis	2.830	3.846	3.213	4.259	3.982
$p_{LM(6)}$	0.385	0.408	0.681	0.348	0.491
$p_{LM(12)}$	0.178	0.526	0.819	0.293	0.421
$p_{ARCH(6)}$	0.685	0.254	0.650	0.158	0.464
$p_{ARCH(12)}$	0.147	0.446	0.627	0.338	0.667
d	8	1	1	9	8

HCC standard errors are given underneath the parameter estimates. Transition variable and the transition function are indicated in the first row of the table along with the currency. d stands for the transition variable used in the estimation. The rows corresponding to $p_{LM(6)}$ and $p_{LM(12)}$ give p-values from LM, statistics for 6th and 12th order serial correlations in residuals. The rows corresponding to $p_{ARCH(6)}$ and $p_{ARCH(12)}$ report the $p-$values for the presence of ARCH effects up to 6th and 12th orders in the residuals. d gives the lag value of the transition variable.

Table 3.6: Tests for remaining nonlinearity and parameter constancy

p-Values from LM_{AMR} test: HCC version					
Tr. var	CD	DG	GDM	IL	JY
y_{t-1}	0.813	0.139	.	0.942	0.955
y_{t-2}	0.670	0.027	0.141	0.372	0.561
y_{t-3}	0.444	0.596	0.455	0.373	0.278
y_{t-4}	0.012	0.129	0.060	0.705	0.680
y_{t-5}	0.318	0.799	0.182	0.552	0.108
y_{t-6}	0.367	0.688	0.702	0.331	0.717
y_{t-7}	0.854	0.154	0.138	0.481	0.443
y_{t-8}	0.914	0.908	0.600	0.763	0.642
y_{t-9}	0.644	0.688	0.853	0.664	0.738
y_{t-10}	0.282	0.367	0.917	0.569	0.477
y_{t-11}	0.100	0.392	0.721	0.165	0.072
y_{t-12}	0.707	0.318	0.919	0.614	0.633
p-Values from LM_{EMR} test: HCC version					
y_{t-1}	0.651	0.098	.	0.950	0.760
y_{t-2}	0.304	0.106	0.251	0.519	0.241
y_{t-3}	0.768	0.828	0.521	0.244	0.168
y_{t-4}	0.042	0.288	0.173	0.872	0.454
y_{t-5}	0.408	0.405	0.160	0.540	0.247
y_{t-6}	0.415	0.398	0.589	0.468	0.848
y_{t-7}	0.779	0.427	0.339	0.194	0.441
y_{t-8}	.	0.746	0.751	0.890	0.460
y_{t-9}	0.179	0.460	0.693	0.081	0.737
y_{t-10}	0.556	0.344	0.976	0.894	0.683
y_{t-11}	0.316	0.590	0.872	0.413	0.197
y_{t-12}	0.729	0.432	0.694	0.843	0.477
p-Values from LM_{Cj} tests for parameter constancy					
Statistics			p-Values		
LM_{C1}	0.869	0.544	0.406	0.379	0.854
LM_{C2}	0.900	0.331	0.519	0.231	0.945
LM_{C3}	0.529	0.305	0.456	0.140	0.987

Table 3.7: Characteristic Roots in extreme regimes

Currency	Regime	Characteristic Roots	Modulus
CD	M	1.000, 0.271	1.000
	O	0.976	0.976
DG	M	1.000, -1.138	1.138
	O	1.00, 0.077	1.00
GM	M	1.000, 0.610	1.00
	O	0.976, 0.395	0.976
IL	M	1.000, 0.592	1.000
	O	0.992	0.992
JY	M	1.000, 0.285	1.000
	O	0.967, 0.285	0.967

*Note:*M stands for the middle regime, and O for the outer regime.

CHAPTER 4

Long Memory in Commodity Markets

4.1 Introduction

In accord with the efficient markets hypothesis, asset price returns and exchange rate returns exhibit very little serial correlation. On the other hand their volatilities contain a much richer structure in that certain transformations of asset price and exchange rate returns have an extremely persistent distinct form of autocorrelation. There is considerable evidence that shows that conditional volatility of returns of asset prices and returns of exchange rates display long memory. Ding et al. (1993), de Lima and Crato (1993), Bollerslev and Mikkelsen (1996), Granger and Ding (1996), have shown that asset price return volatilities have long memory property. On the other hand, Baillie et al. (1996) have shown that exchange rate volatility displays long memory property. Previous literature has found daily commodity series to be well described by martingale-GARCH(1,1) models, see for example, Baillie and Myers (1991).

The purpose of this chapter is to examine daily commodity futures and cash returns for several primary commodities and their volatilities, particularly, their squared and absolute returns as well as intra-daily ranges. The subject of this chapter is modelling volatility in commodity markets. At a substantive level, one may be interested in forecasting the volatility in these markets. Moreover, knowledge of the dynamic properties of return volatilities may have implications on the dynamic nature of commodity prices, and forecasting optimal hedge ratios. This is because a finding of time dependency in second conditional moments of cash and future commodity returns will imply that optimal hedge ratios should be time dependent as well. See for instance Baillie and Myers (1991). The results of this study may be helpful in comparing the dynamic features of commodity markets with that of stock and foreign exchange markets. This in turn may have implications for theoretical modelling of the prices in these markets. This study tries to answer the following questions. Do daily commodity cash and future prices have *long memory* property, with cash and future returns being approximately uncorrelated, and with very persistent autocorrelation in certain proxies for the volatility, such as, for example, squared and absolute returns and intradaily ranges?

Granger and Ding (1995), using the results of Luce (1980), showed that the expected absolute return and any power transformation of this return, may be interpreted as a measure of risk. Hence, volatility literature routinely uses absolute or squared returns as volatility proxies. In this chapter, following Garman and Klass (1980), Parkinson (1980) and Anderson and Bollerslev (1998), we consider a third proxy, namely range, defined here as the difference between the highest and lowest log asset price during a discrete sampling interval. It is by now well known that the conditional distribution of log absolute and squared returns are far from Gaussian. On the other hand, Alizadeh, Brandt, and Diebold (1999)show both theoretically and empirically that log range is approximately Gaussian, in sharp contrast to popular volatility proxies, such as log absolute and/or squared returns. There is considerable literature on both absolute and squared returns in stock and exchange rates markets, but little attention has been paid to extreme value volatility proxies. Range as a proxy for volatility has been appreciated in the business press, which routinely displays high and low prices. One potential problem in the use of range as a proxy for volatility is the downward bias in the range induced by discrete sampling (Rogers and Satchell 1991). However, as Alizadeh, Brandt, and Diebold (1999) and Anderson and Bollerslev (1998) show on days with substantial price reversals, return-based proxies underestimate daily volatility, as the closing price is not very different from the opening price, despite the large intraday price fluctuations. The range in this sense may better reflect the intraday volatility. In this chapter, the long memory property of absolute and squared returns as well as intraday range will be analyzed. If intraday log range exhibits long range dependence then this may support the findings of Anderson and Bollerslev (1998) and Alizadeh, Brandt, and Diebold (1999)and motivate consideration of intraday log range in modelling financial market volatility.

We utilize the Fractionally Integrated GARCH (FIGARCH) model of Baillie et al. (1996) to model the dynamics of volatility in commodity cash and futures returns. Since the GARCH model attempts to account for volatility persistence, but has the feature that persistence decays relatively fast, we use it as a benchmark and compare its results with the FIGARCH model, as the latter model is capable of modelling very long temporal dependencies in the conditional variance of a process. In order to better asses the presence of long memory in the volatility of commodity future and cash returns, this chapter also models absolute returns, squared returns, and intraday ranges using the Fractionally Integrated Autoregressive Moving Average (ARFIMA) model of Granger and Joyeux (1980), and Hosking (1981). Moreover, estimates of the long memory parameter for the volatility proxies from semi-parametric methods are also obtained. Particularly, the GPH estimator from Geweke and Portar-Hudak (1983), and a local Whitlle estimator based on Fox and Taqque (1986) are used.

The rest of the chapter is organized as follows. Section 4.2 describes the data and examines the empirical autocorrelations of the series. Section 4.3 presents and discusses the results from the estimation of the FIGARCH models for daily cash and future return volatilities. Results from the estimation of the ARFIMA models and nonparametric methods for squared and absolute returns are discussed in section 4.4. The last section provides the conclusion.

4.2 The Data

We analyze cash and future prices on commodities, coffee, corn, gold, silver, soybean, and unleaded gasoline. The data is obtained from the Chicago Mercantile Exchange. The data set consists of the daily observations for each commodity. The sample period differs for each commodity. The sample periods for each of the commodities are the following; coffee, 03/20/84-12/29/00; corn, 03/20/85-03/14/01; gold, 04/21/75-03/31/00; silver, 12/26/89-12/26/97; soybean, 03/20/80-12/29/00; and unleaded gasoline, 04/25/86-12/29/00. Each contract starts trading well before the delivery month. Except for gold and silver, for all commodities we consider the contract that expires in March of each year. For gold, the December contract, and for silver, the April contract are used.

Following the standard practice, the returns are defined as $R_t = 100 \times \Delta \ln(P_t)$, where P_t is the price (either cash or future) at date t, absolute returns as $|R_t|$, and squared returns as R_t^2. Daily returns are computed for each contract and then combined to obtain a series of future returns. In estimation, dummy variables are included to see if contract expiration dates have any statistically significant effect on the return and volatility dynamics. For none of the commodities were the estimated coefficients of dummy variables significant. Following Parkinson (1980), range is defined by

$$RR_t = \frac{\ln(P_t^h) - \ln(P_t^l)}{2\ln 2},$$

where P_t^h and P_t^l are the highest and lowest prices at day t, respectively.

Panels of figures 4.1 and 4.2 give the graphs of the daily cash and future returns, absolute returns and squared returns, as well as intraday range for the commodity futures over each sample period. It appears from the graphs that for all commodities, relatively volatile periods, characterized by large price changes, alternate with more tranquil periods in which prices remain more or less stable. This indicates that large cash and future returns (both positive and negative) seem to occur in clusters and so does volatility. The volatility clustering phenomenon which is typical of stock prices and exchange rates, seems to occur in the commodity markets as well.

Summary statistics for the future and cash returns are given in table (4.1). The table indicates that most of the series have small negative means and medians equal to zero over their respective sample periods. One of the usual ways of getting an idea of the distribution of a time series y_t is to look at the *kurtosis* and the *skewness* and compare them with that of a normal random variable. The last two columns of table 4.1 indicate that the kurtosis of all returns are much larger than that of a normal random variable. This reflects the fact that the tails of the distribution of these return series are fatter than the tails of the normal distribution. This in turn means that large realizations occur more often than one might expect for a normally distributed variable.

Since any symmetric distribution has a skewness equal to zero, table 4.1 indicates that the distribution of the daily cash returns has some asymmetry. From table 4.1 it is seen that all of the future returns and three out of six cash returns (silver, soybean, and unleaded gasoline) have negative skewness. This implies that for those commodities, the left tail of the distribution is fatter than the right tail, or large negative returns tend to occur more often than large positive ones. The analysis here indicates that daily future and cash return distributions are far from being normal. This finding is consistent with the distributions of daily returns for stock price returns and exchange rate returns.

Table (4.2) gives the summary statistics for return based and range based volatility proxies for the commodity futures. For almost all commodities, intraday volatility has a lower sample variance and skewness compared with absolute and squared returns. Squared returns always have the highest kurtosis. It seems that not only return based volatility proxies but also log range is far from being normal, a result in contrast to the findings of Alizadeh, Brendt, and Diebold (1999).

Table (4.3) reports the results from the Phillips-Perron test (PP) from Phillips and Perron (1988), and the KPSS test, due to Kwiatoski et al. (1992). The PP tests the null hypothesis of a unit root, $I(1)$, against the alternative of $I(0)$, while KPSS tests the null of an $I(0)$ against the alternative of an $I(1)$ process. As shown in Lee and Schmidt (1996) the KPSS test has power against the long memory alternative as well. Both tests indicate that commodity futures and cash prices are non-stationary and possibly have a unit root, while daily cash and future returns are stationary. The PP test indicates that all of the volatility proxies are stationary. The KPSS test, on the other hand, rejects the null of $I(0)$ for the squared future returns and absolute returns for coffee, gold, soybeans, and unleaded gasoline. Combined with the results of the PP test, this may indicate long memory behavior in the future squared and absolute returns for these commodities. The KPSS test also rejects its null for coffee, gold, silver, and soybeans intraday ranges. Hence, there is some evidence from the unit root and stationarity tests that volatility proxies may have long memory behavior for some of the commodity future returns. The KPSS test rejects its null for coffee and gold squared cash returns, and for the absolute returns of coffee, gold, soybean and unleaded gasoline at the 5 percent level. Hence, evidence of long memory for the cash squared and absolute returns is not that strong compared to future squared and absolute returns.

To gain further insight on the dependence structure of the series, panels of figures 4.3 and 4.4 display the first 100 autocorrelations for the daily log cash and future returns, absolute returns, squared returns, and intraday range together with two-sided 5 percent critical values ($\pm 1.96/\sqrt{T}$) where T is the respective sample size. It is seen that the autocorrelations of the future and cash returns are very small, even at low lags and for a majority of lags they are within the 5 percent intervals. Hence, autocorrelations of returns mimic the autocorrelation structure of a stationary process. By contrast, for the absolute and squared returns, and the intraday ranges the autocorrelations start off at a moderate level but remain significantly positive for a substantial number of lags. Moreover, autocorrelation in the absolute returns is generally somewhat higher than the autocorrelation in the squared returns and for all commodities autocorrelations in absolute returns hardly become insignificant at all lags considered. This illustrates what has become known as the 'Taylor property' (see Taylor, 1986, pp.52-55), that is, when calculating the autocorrelations for the series R_t^δ for various values of δ, one almost invariably finds that autocorrelations are largest for $\delta = 1$.

As is evident from the graphs, autocorrelations for absolute returns are not only larger than those of squared returns but also much more persistent in the sense that they decay much more slowly. Moreover, autocorrelations for intraday range are usually higher than those of absolute and squared returns and more persistent. The autocorrelations in absolute and squared returns and intraday range seem to mimic the correlation properties of a long memory process rather than a short memory stationary process for which autocorrelations decay to zero at an exponential rate. As is evident from the graphs, the autocorrelations in absolute and squared returns and intraday range decay very slowly, indicating that the linear association between distant observations is persistent and autocorrelations decay at a hyperbolic rate. This behavior of autocorrelations is consistent with time series models with long memory or long range dependence. The above described characteristics of autocorrelations in log commodity future and cash prices are in conformity with the findings from the stock and foreign exchange markets. For example, see Ding and Granger (1993), Baillie et al. (1996), Bollerslev et al. (1996).

4.3 Results from GARCH and FIGARCH Models

A class of parametric models that is capable of modelling volatility clustering and the persistence in the autocorrelations of absolute and squared cash returns is the Fractionally Integrated Generalized Autoregressive Heteroscedastic (FIGARCH) model of Baillie et al. (1996). The details of volatility models are discussed in chapter 2.

In the light of the discussion in chapter 2, conditional variance of commodity cash and future returns are modelled by GARCH/FIGARCH processes. The robust Wald statistic is used to check if the estimated FIGARCH model better represents the long memory property of the data compared to a GARCH specification. Results of the estimated $ARMA(p,q) - FIGARCH(P,\delta,Q)$ models for future and cash returns are presented in tables (4.4)-(4.7). The conditional mean specification for cash and future returns varies across different commodities. An $MA(1)$ specification found to be satisfactory for modelling the conditional mean of cash and future returns for all commodities except coffee. For the conditional mean of coffee cash and future returns an $MA(3)$ found to be a better specification. The estimate of long memory parameter, δ, for daily future and cash returns are significantly different from zero. Various tests for specification of the models were performed. In particular, the last row of the tables (4.5 and 4.7) give the robust Wald test values of a stationary $GARCH(1,1)$ model under the null hypothesis against a $FIGARCH(1,\delta,1)$ model under the alternative hypothesis. In each of the commodities, the robust Wald test values indicate clear rejection of the null hypothesis when compared with the critical values of a χ^2di stribution with one degree of freedom. For none of the commodities did the estimated $GARCH$ models performed better than the $FIGARCH$ models. The sum of the estimates of α and β in the $GARCH$ models are found to be close to one for all commodities, indicating that the volatility process is highly persistent. In all cases the standardized residuals exhibit less skewness and kurtosis than the returns. Perhaps of greater importance, the Ljung-Box statistic, Q, fails to reject the null hypothesis of independently and identically distributed standardized residuals and squared standardized residuals for most of the commodities. One striking result from table 4.7 is the finding of dual long memory in both conditional mean and conditional variance of the coffee cash returns. As the table indicates, an $ARFIMA(0, d, 1) - FIGARCH(1, \delta, 0)$ model seems to fit the coffee cash returns better than the other specifications. Although the estimate of the long memory parameter is small, it is significantly different from zero.

To obtain some insight into the volatility in the commodity markets, panels of figure 4.5 present the commodity future returns together with the estimated conditional variances from the $FIGARCH$ models. As the figures indicate, the estimated models do very well in describing in sample volatility in the commodity markets. The $FIGARCH$ models are quite accurate in estimating the time dependence and clustering in the volatility.

In the $FIGARCH$ model, taking out the mean parameters, the squared error term coincides with the squared return. Hence, the $FIGARCH$ model estimates provide evidence that the squared returns exhibit long memory. As indicated in section 4.2 the autocorrelations of squared returns, absolute returns, and intraday range seemed to mimic the autocorrelation structure of a long memory process. Moreover, the results of the unit root and stationarity tests indicated that the volatility proxies are neither unit root nor stationary. A result that can be interpreted as evidence of long memory. To further analyze the long memory in the proxies for the volatility tables 4.8 and 4.9 present the results from the GPH estimates for different number of periodogram ordinates and the table 4.10 reports results from the local Whittle estimation. The results show that both cash and future squared returns and absolute returns exhibit the long memory property with the estimates of the long memory parameter being significantly greater than zero and less than one. In most cases, the estimate is less than 0.5 indicating both long memory and stationarity. These findings are consistent with the $FIGARCH$ estimates. Interestingly, the intraday range also exhibits long memory usually the long memory parameter estimates usually greater than those of squared and absolute returns.

4.4 Conclusion

In this chapter, we analyzed daily commodity cash and future returns for certain primary commodities. The returns are modelled through the $GARCH$ and the $FIGARCH$ models. The chapter found evidence supporting the $FIGARCH$ models in the sense that the $FIGARCH$ models fit the data better than the $GARCH$ models. The $FIGARCH$ specification is able to capture both long and short run dynamic characteristics of the volatility process. The estimates of the fractional degree of integration parameter were found to be significantly different from zero. Robust Wald tests are used to test the $FIGARCH$ models against the $GARCH$ models and in all cases the tests rejected a $GARCH(1,1)$ model in favor of a $FIGARCH(1,\delta,1)$ model. This implies we need to consider time dependency and long term dependence in forecasting optimal hedge ratios. On the other hand this requires a bivariate FIGARCH modelling of cash and future returns. This is a potentially interesting question that may also raise interesting econometric issues that need to be studied in the future.

For each commodity the chapter also considered measures of risk or the volatility proxies, namely, squared returns, absolute returns, and the intraday range (or volatility). The sample autocorrelations, unit root and stationarity tests, and estimates from the semi-parametric methods, namely, the GPH estimates and the local Whittle estimates of the long memory parameter indicated presence of the long memory component in the volatility proxies. The findings here indicate that, in addition to squared returns and absolute returns, intraday range exhibits long memory property and it seems to be more persistent than the squared and absolute returns. The findings support the findings of Alizadeh et al. (1999) in that intraday range can be as good a proxy for the volatility as the squared and absolute returns.

The findings in this chapter indicates that on a practical level, one need to take into consideration the long memory in the conditional volatility of commodity cash and future returns in assessing the risk and return relations in these markets. The results also indicate that the optimal hedge ratios should be time dependent and one needs to consider taking the long memory dynamics in the conditional volatility in forecasting optimal hedge ratios. As shown in Baillie and Myers (1991) the optimal hedge ratios should be time dependent when there are GARCH effects. The findings in this chapter indicate that similar to Baillie and Myers (1991), one can improve in forecasting hedge ratios by considering the long memory in the conditional variance of cash and future returns.

BIBLIOGRAPHY

[1] Alizadeh, S., M. Brandt, and F. X. Diebold (1999), Range-based estimation of stochastic volatility models or exchange rate dynamics are more interesting than you think, Working paper, Stern School, NYU.

[2] Anderson, T. G. and T. Bollerslev (1998), Answering the skecptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review* **39,** 885-905.

[3] Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen (1996), Fractionally integrated Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* **74,** 3-30.

[4] Bollerslev, T. and H.O.A. Mikkelsen (1996), Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73,** 151-184.

[5] Baillie, R. T., and R. J. Myers (1991), Bivariate GARCH estimation of the optimal commodity futures hedge, *Journal of Applied Econometrics* **6,** 109-124.

[6] Crato, N. (1994), Some international evidence regarding the stochastic behavior of stock returns, *Applied Financial Economics,* **4,** 33-9.

[7] de Lima, P.J.F. and N. Crato (1993), Long range dependence in the conditional variance of stock returns, paper presented in August 1993 Joint Statistical Meeting, San Fransisco. Prooceedings of the Business and Economic Statistics Section.

[8] Ding, Z., C.W.J. Granger, and R.F. Engle (1993), A long memory property of stock returns and a new model, *Journal of Empirical Finance,* **1,** 83-106.

[9] Fox, R., and Taqqu, M. S. (1986), Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics,* **14,** 517-532.

[10] Granger, C., (1980), Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics,* **14,** 227-238.

[11] Granger, C., and R. Joyeux (1980), An Introduction to Long Memory Time Models and Fractional Differencing, *Journal of Time Series Analysis,* **1,** 15-29.

[12] Granger, C. Z. Ding (1995), Some properties of absolute return: an alternative measure of risk, *Annales D'Economie et de Statistque* **40,** 67-91.

[13] Granger, C. and Z. Ding (1996), Varities of long memory models *Journal of Econometrics* **73,** 61-77.

[14] Hosking, J. (1981), Fractional Differencing, *Biometrika,* **68,** 165-176.

[15] Garman, M. B., and M. J. Klass (1980), On the estimation of security price volatility from historical data,*Journal of Business* **53,** 67-78.

[16] Gweke, J. and Portar-Hudak, S. (1983), The estimation and application of long memory time series models, *Journal of Time Series Analysis* **4,** 221-238.

[17] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* **54,** 159-178.

[18] Luce, R. D. (1980), Several possible measures of risk, *Theory and Decision* **12,** 217-228.

[19] Parkinson, M. (1980), The extreme value method for estimating the variance of the rate of return, *Journal of Business* **53,** 61-65.

[20] Phillips, P. and C.B.P. Perron (1988), Testing for a unit root in time series regression, *Biometrika* **75,** 335-346.

[21] Rogers, L.C.G., and S. E. Satchell (1991), Estimating variance from high, low and closing prices, *Annals of Applied Probability* **1,** 504-512.

[22] Taylor, S. (1986), Modelling Financial Time Series, John Wiley & Sons, New York.

Figure 4.1: Cash returns, absolute and squared returns

a. **Coffee**

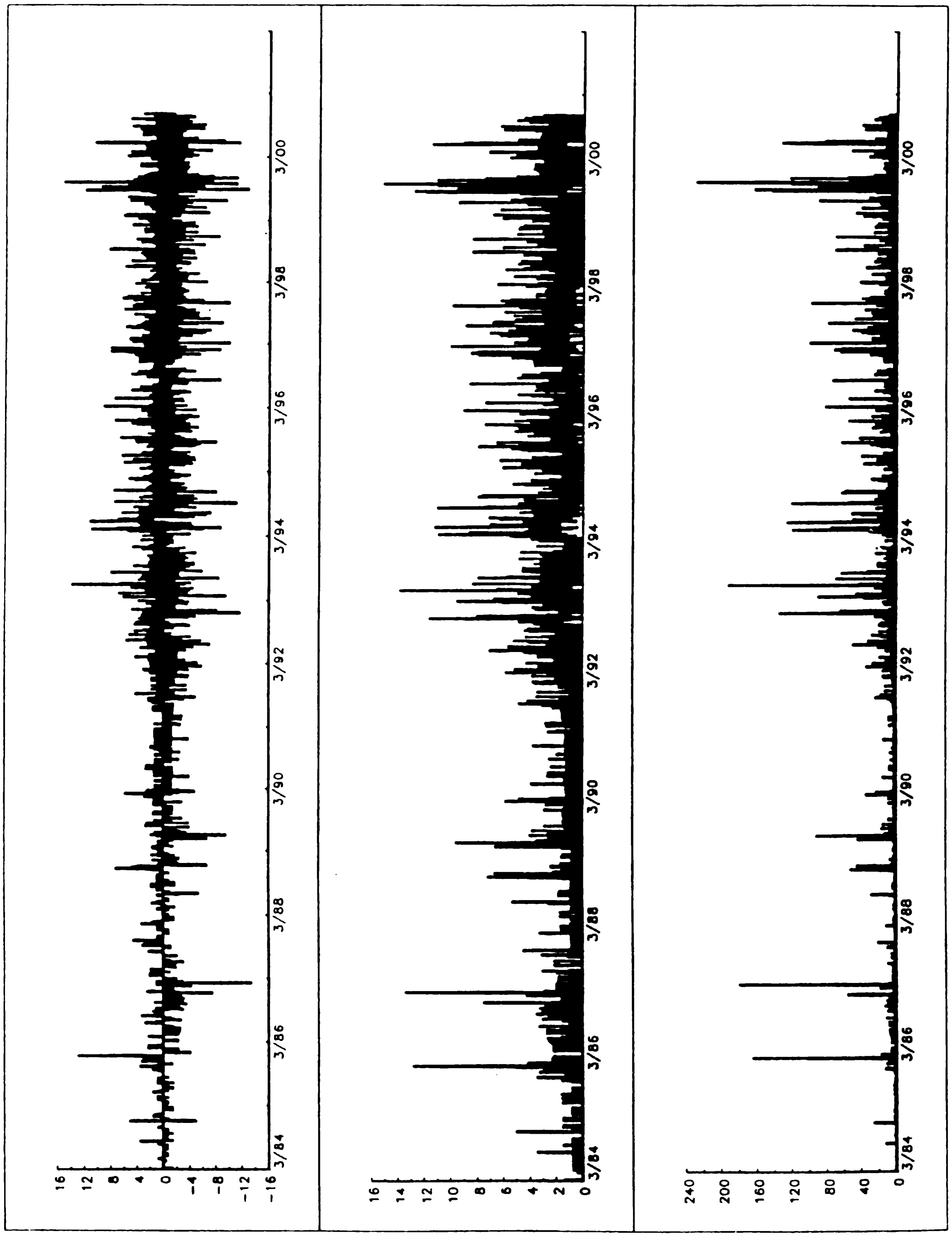


Figure 4.1 (cont'd).

b. Corn

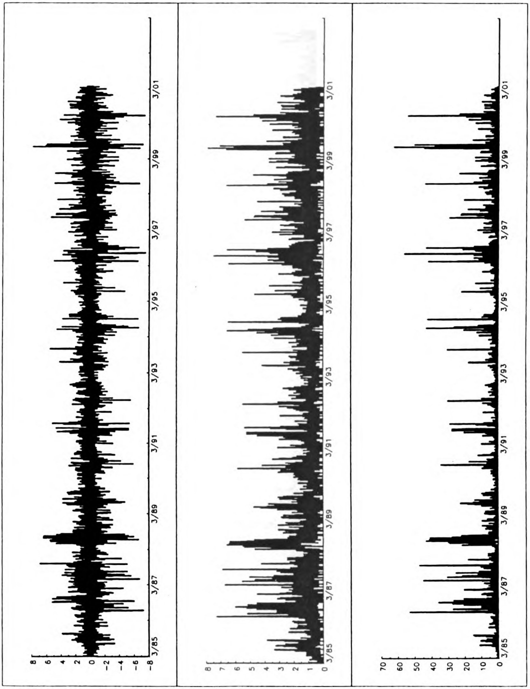


Figure 4.1 (cont'd).

c. Gold

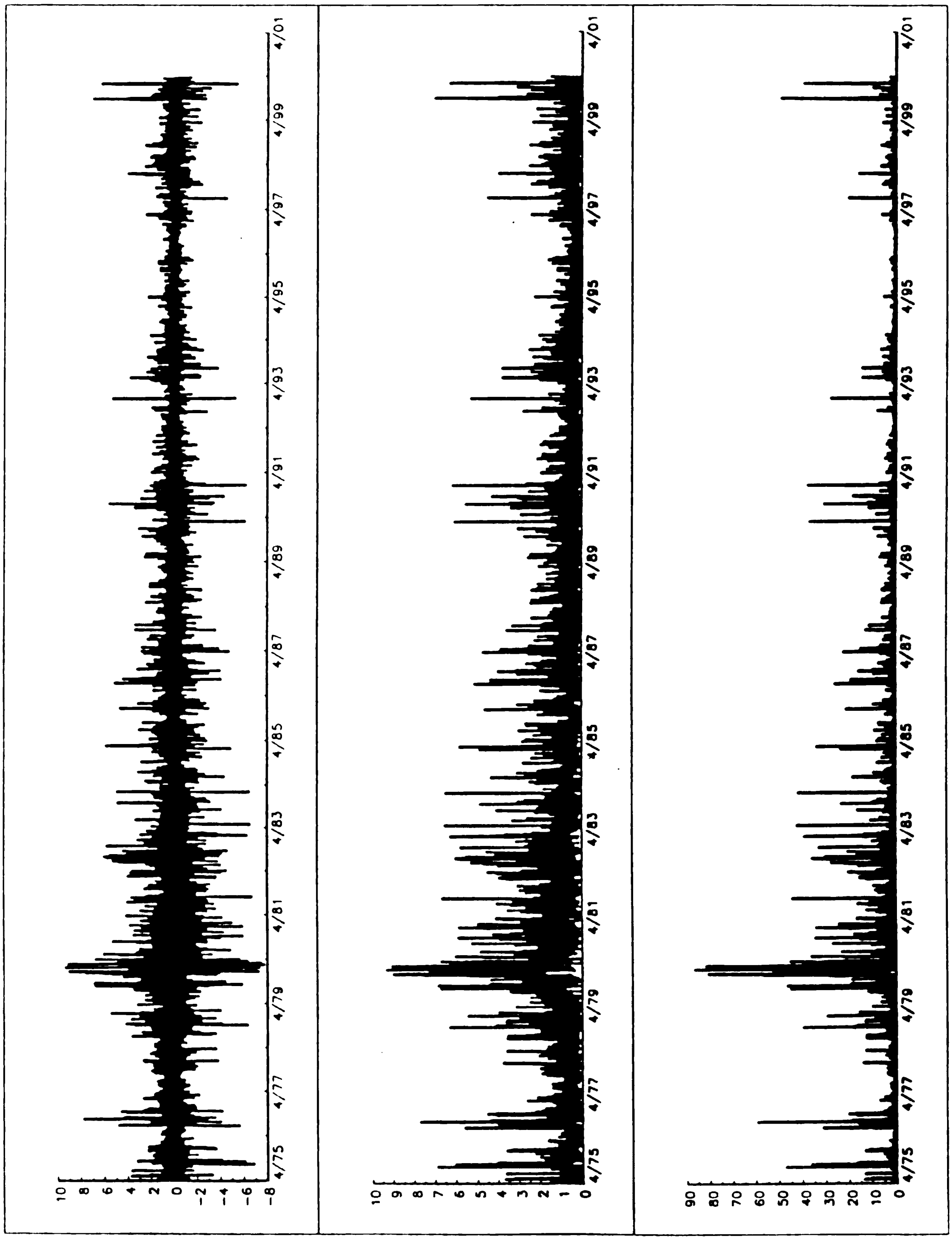


Figure 4.1 (cont'd).

d. Silver

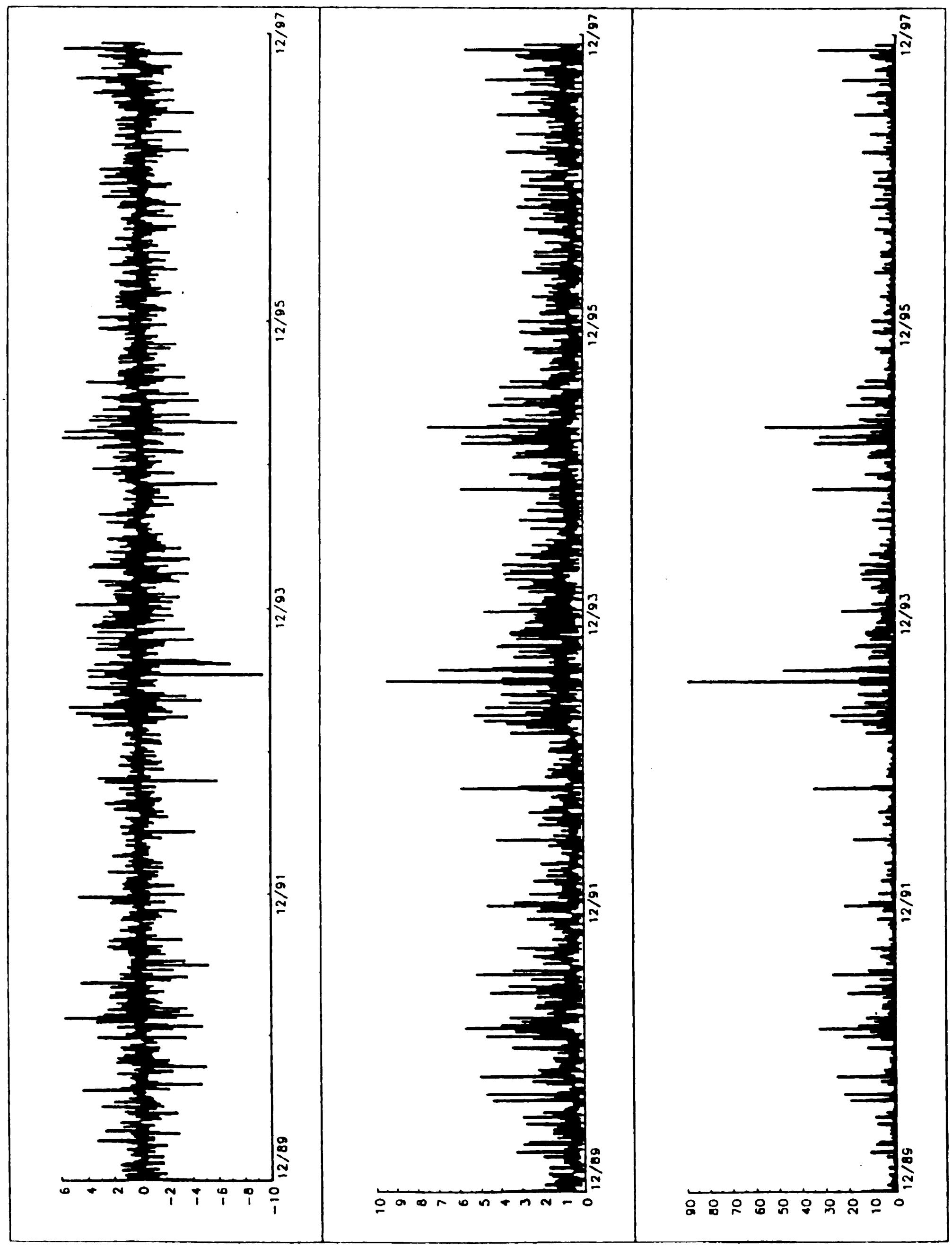


Figure 4.1 (cont'd).

e. Soybean

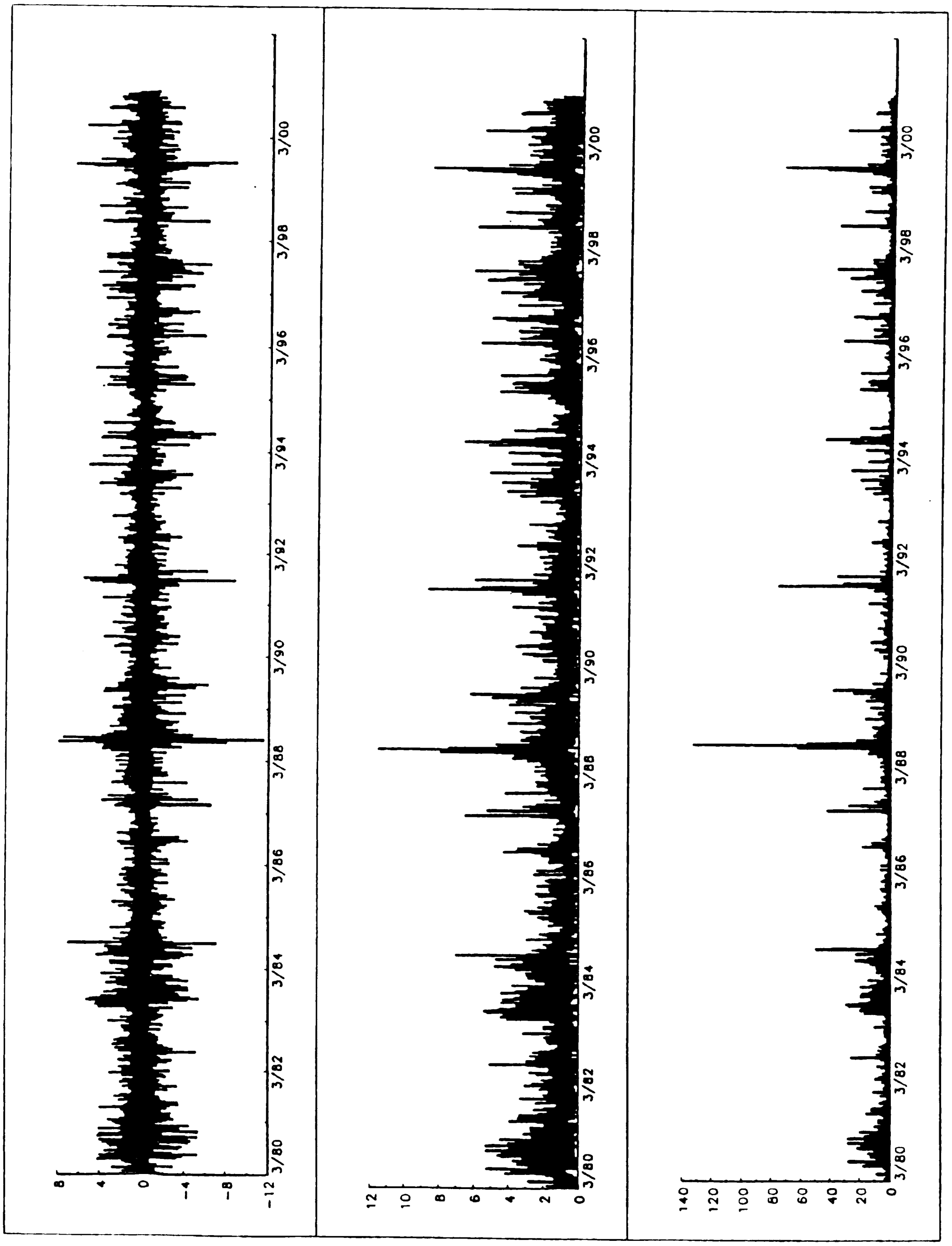


Figure 4.1 (cont'd).

f. **Unleaded Gasoline**

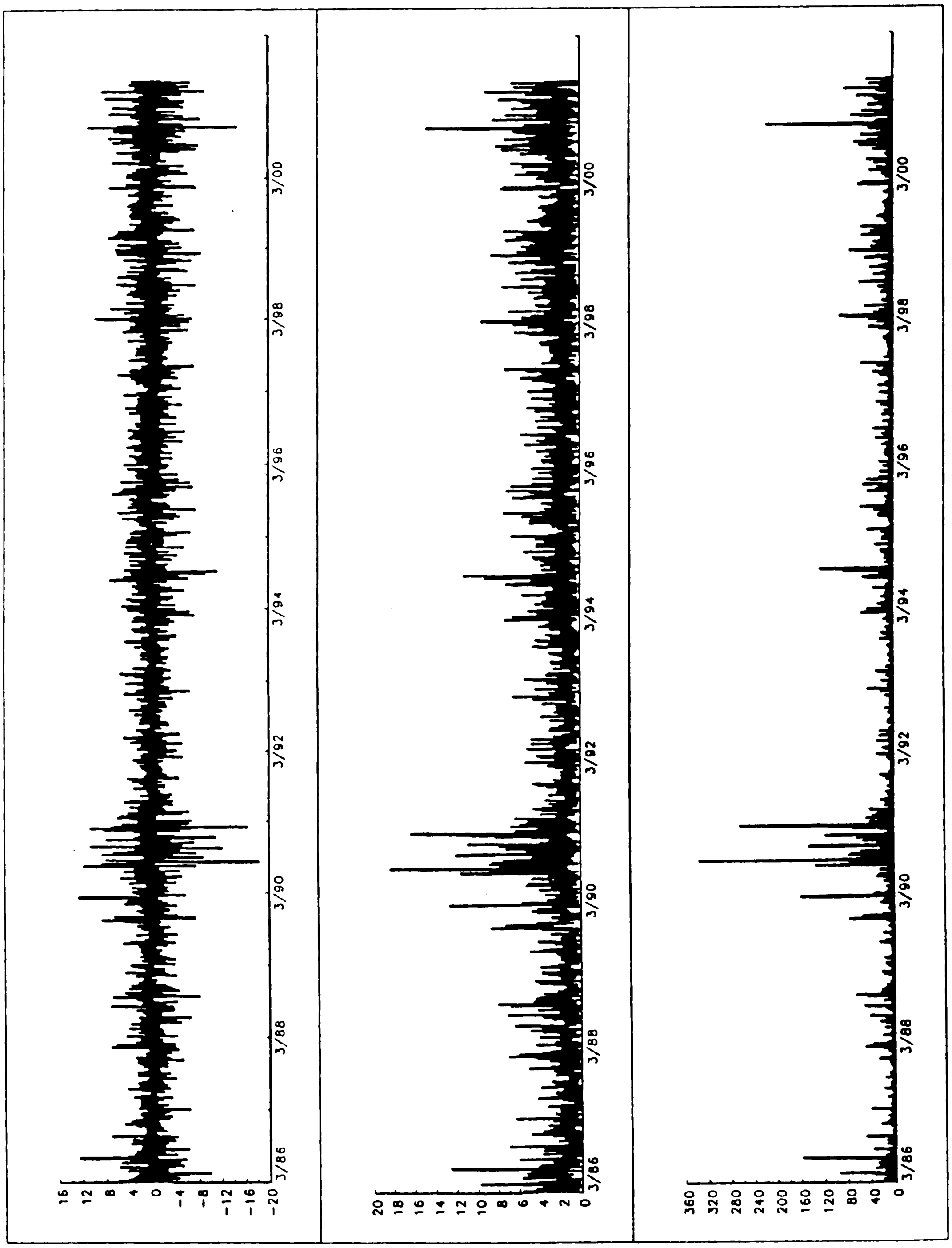


Figure 4.2: Commodity future returns, absolute returns, and intraday range

a. Coffee

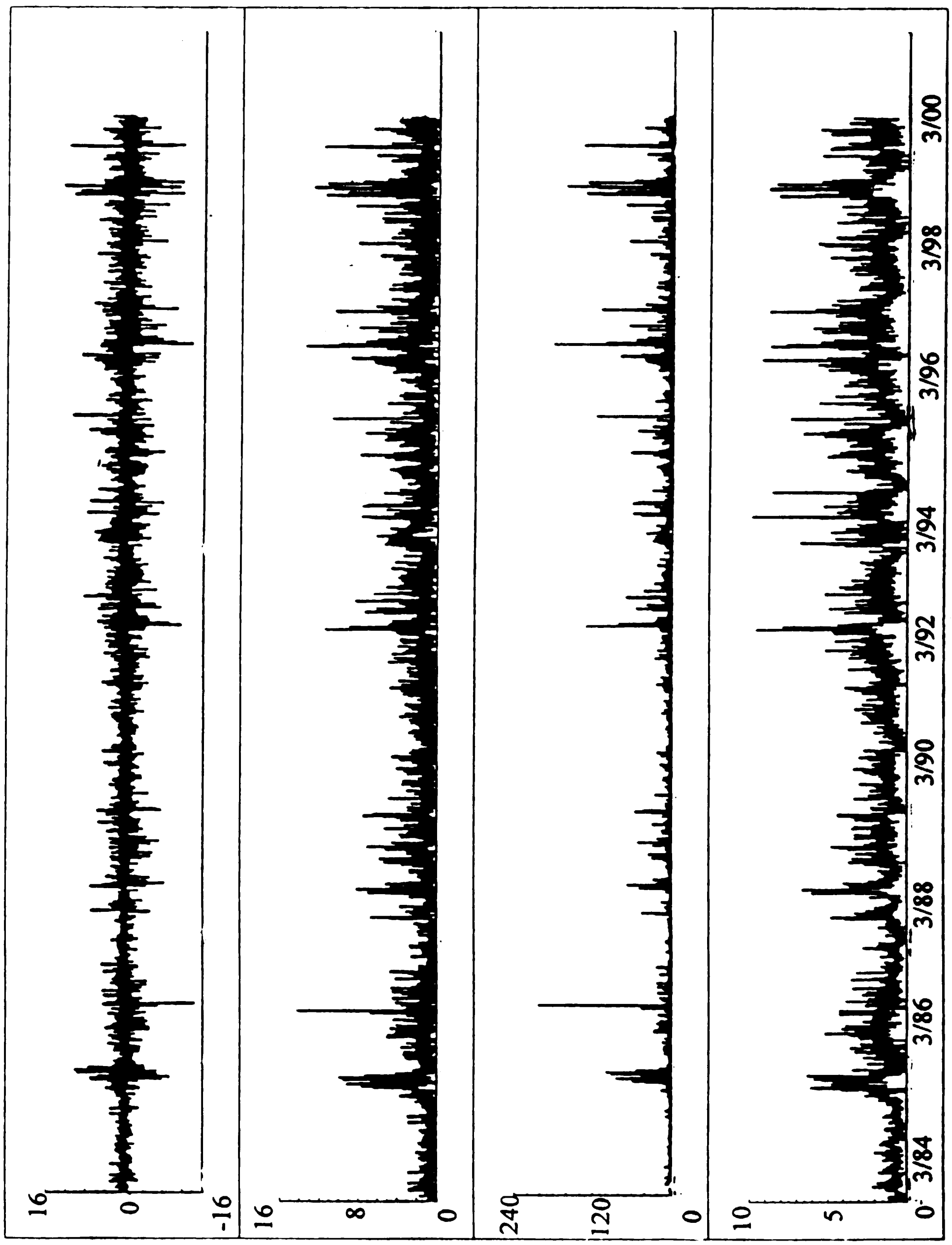


Figure 4.2 (cont'd).

b. Corn

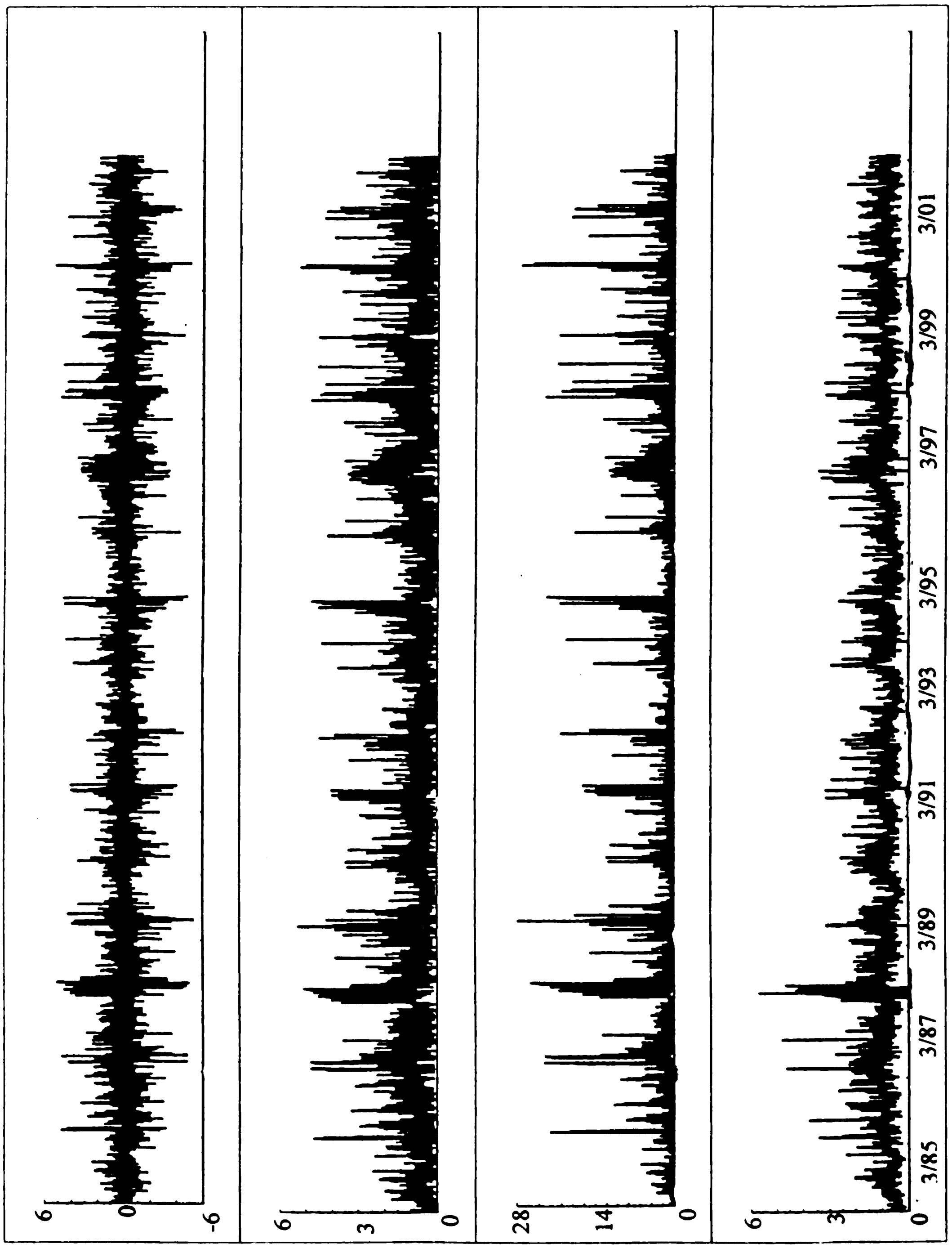


Figure 4.2 (cont'd).

c. Gold

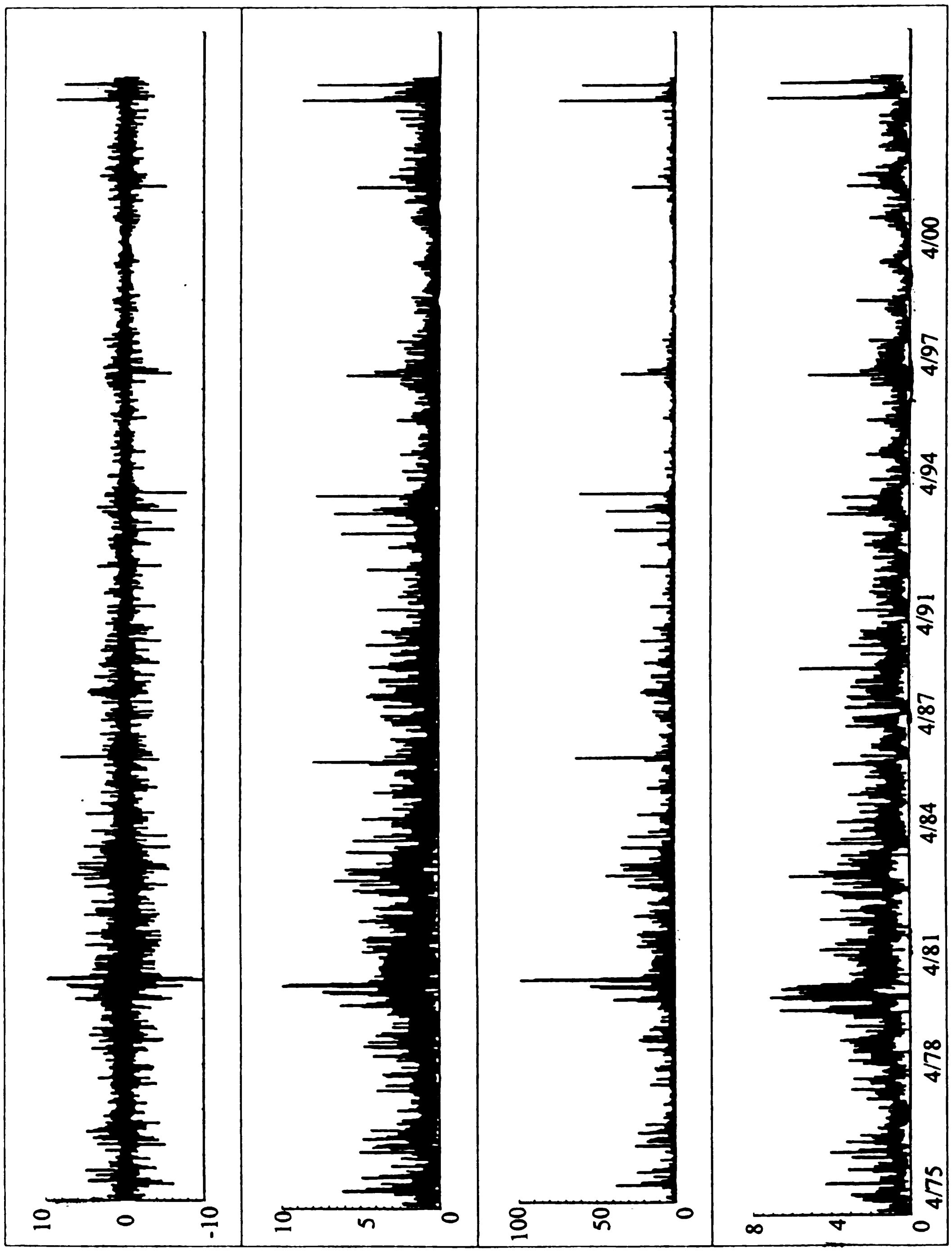


Figure 4.2 (cont'd).

d. Silver

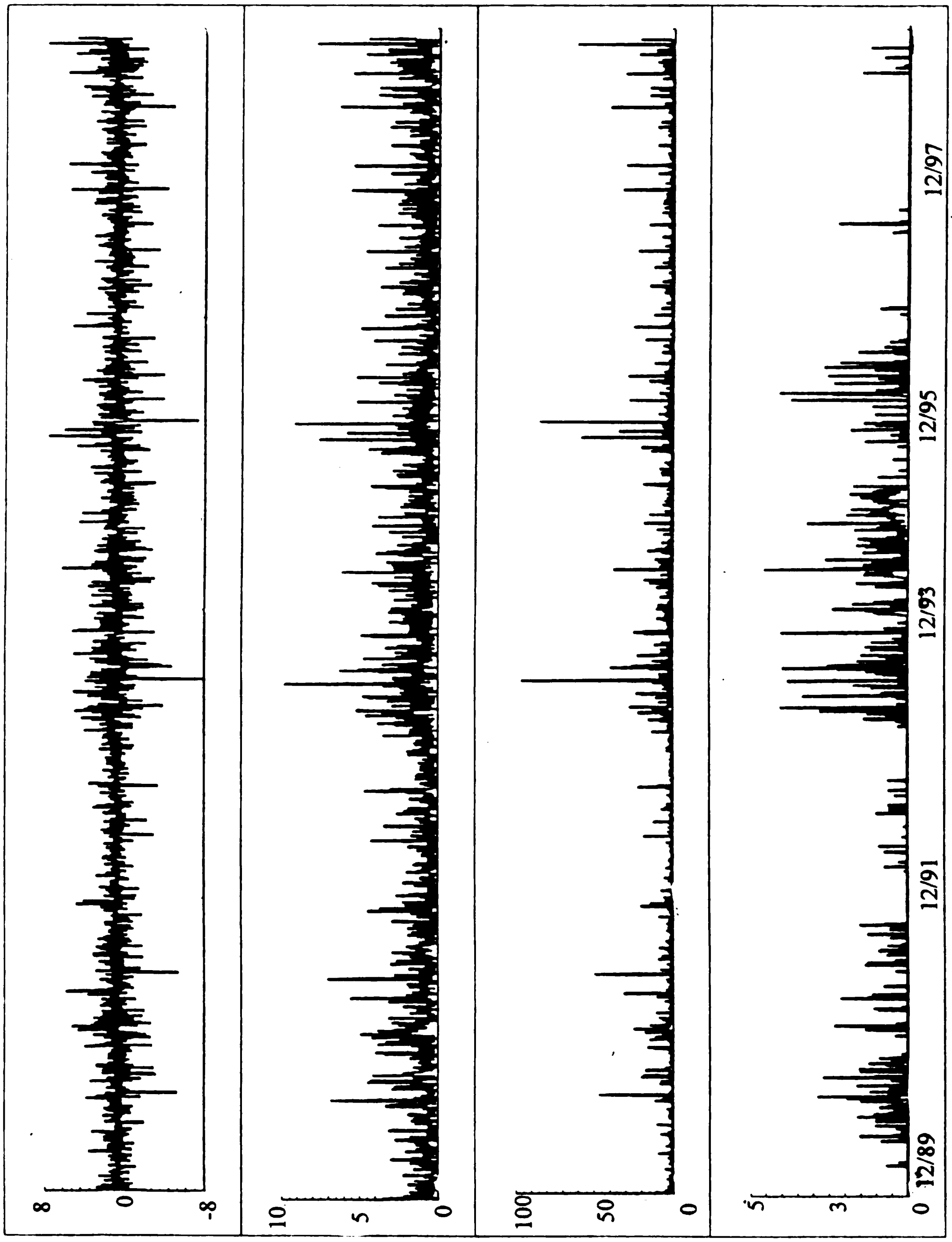


Figure 4.2 (cont'd).

e. Soybean

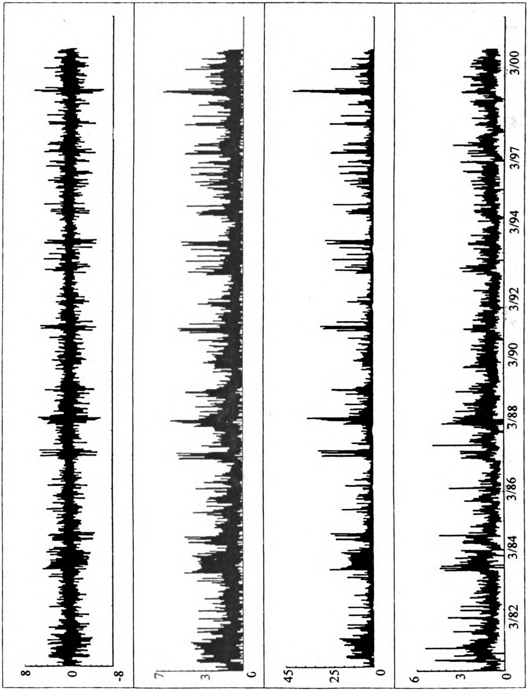


Figure 4.2 (cont'd).

f. Unleaded Gasoline

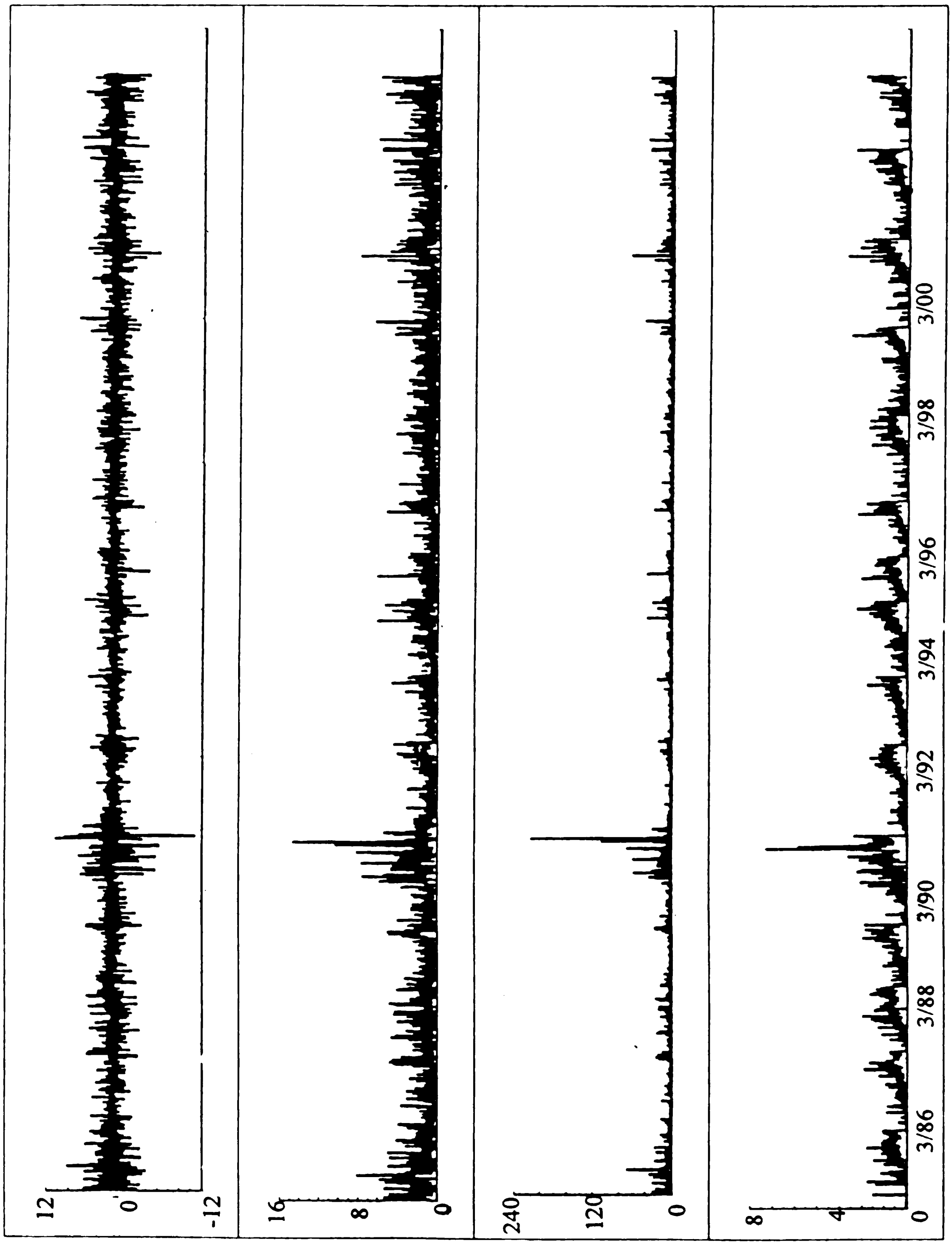


Figure 4.3: Autocorrelations for cash returns, absolute and squared returns

a. Coffee

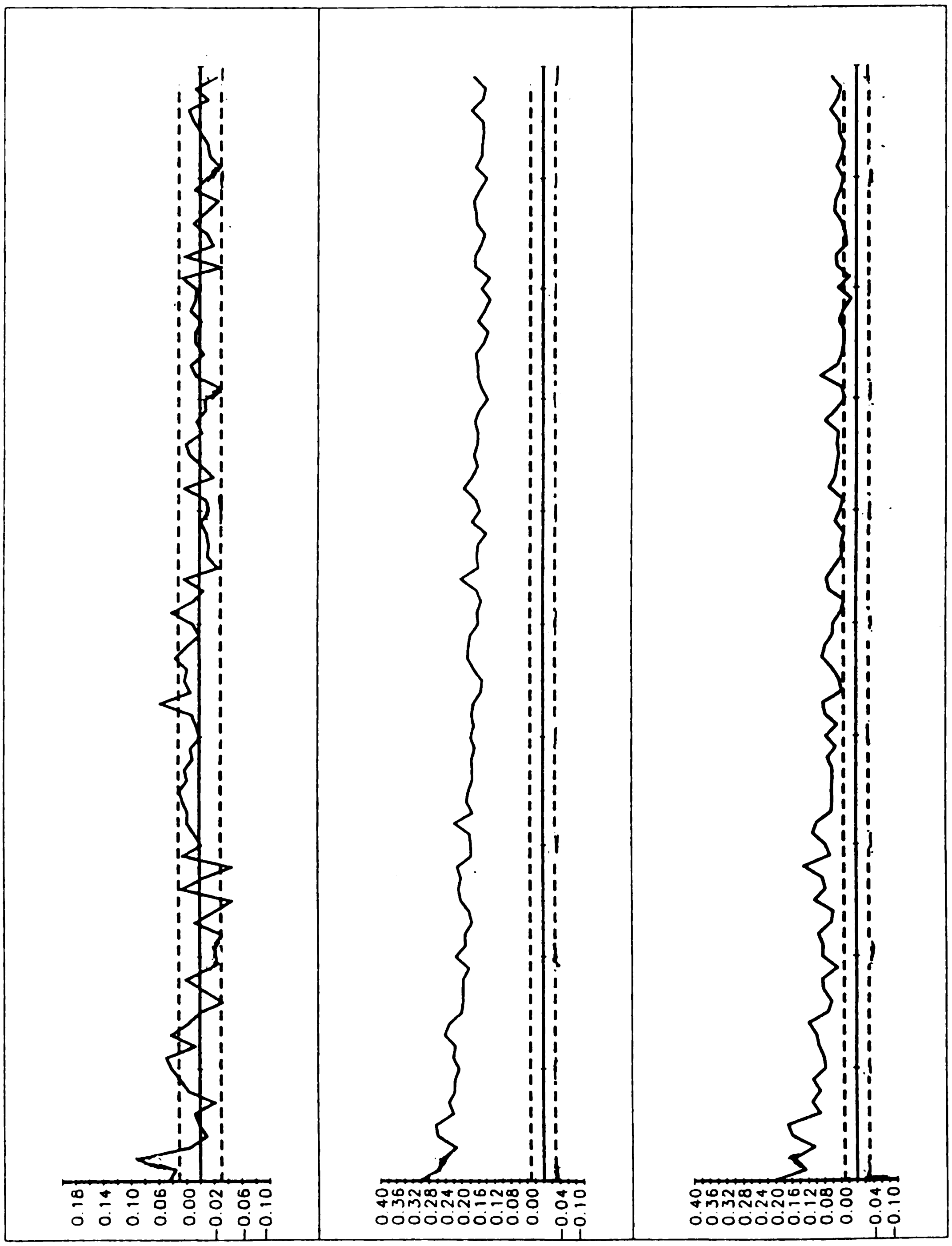


Figure 4.3 (cont'd).

b. Corn

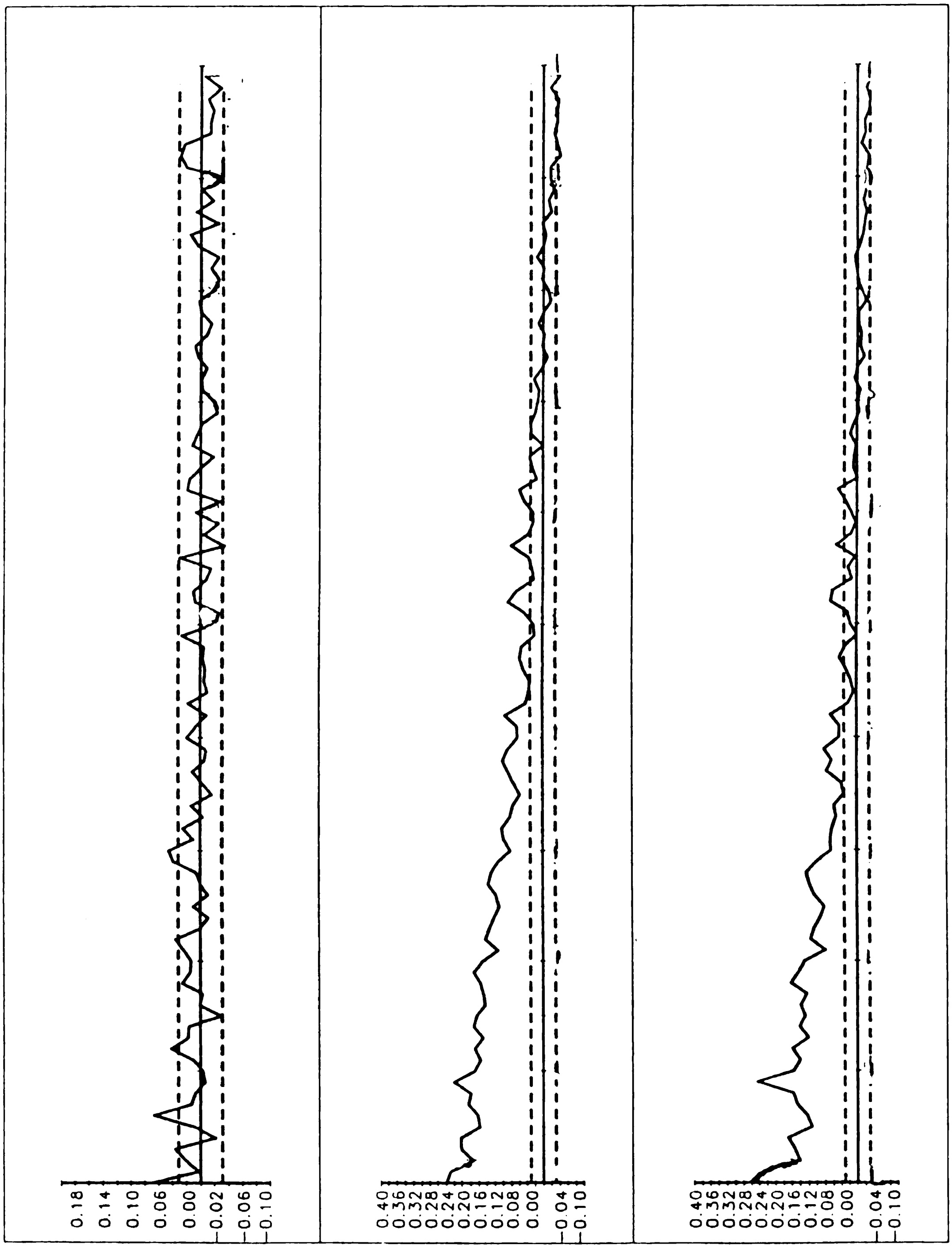


Figure 4.3 (cont'd).

c. Gold

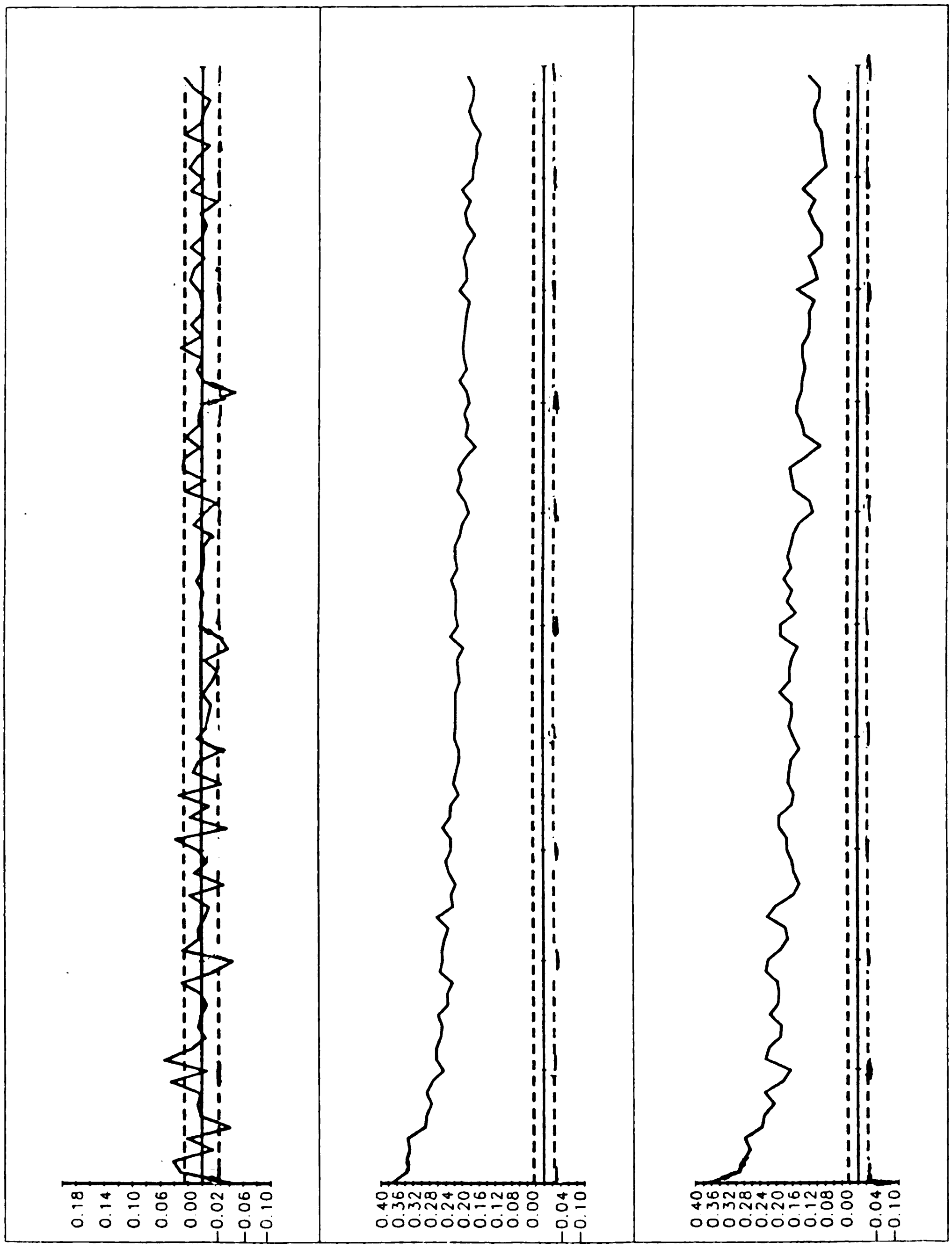


Figure 4.3 (cont'd).

d. Silver

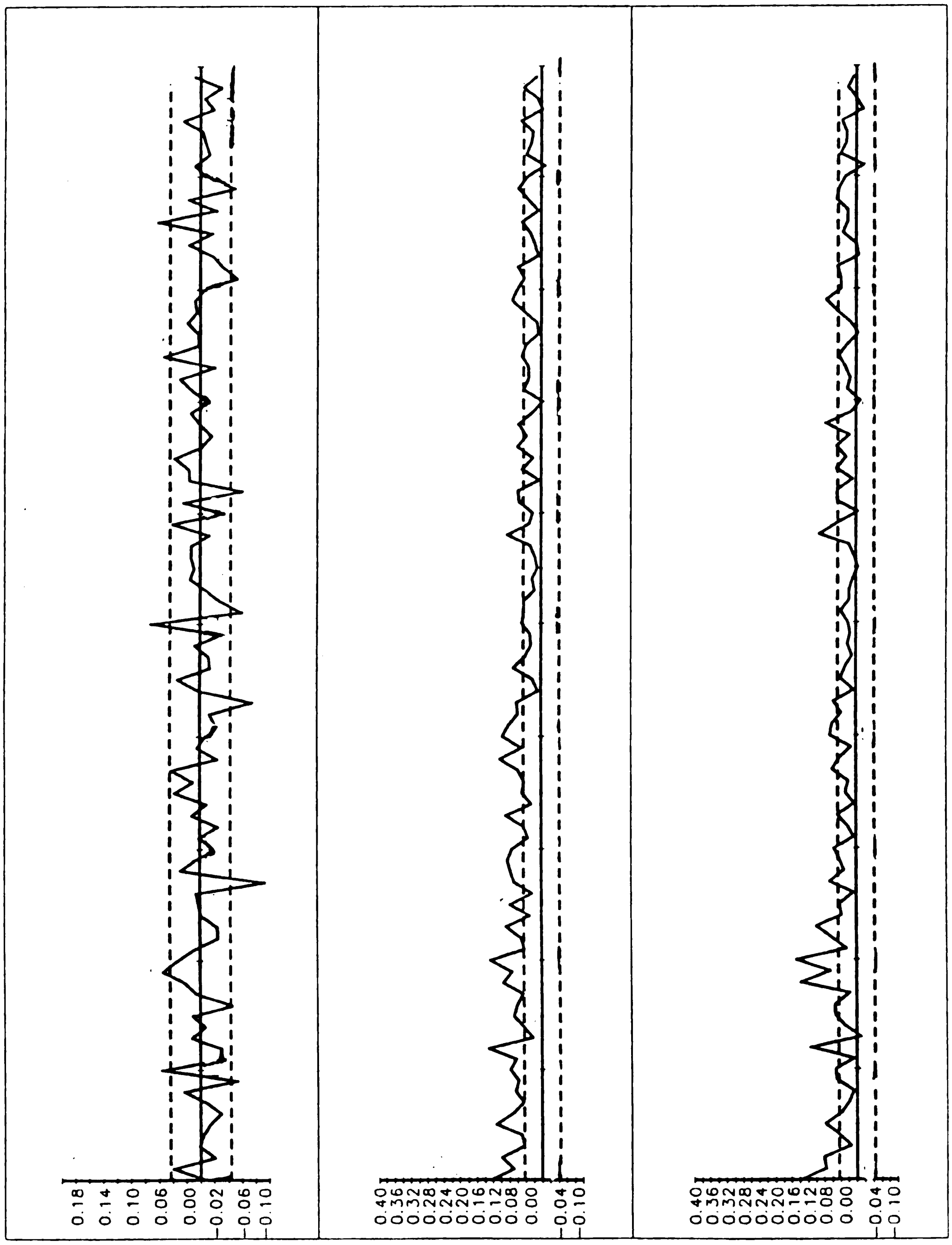


Figure 4.3 (cont'd).

e. Soybean

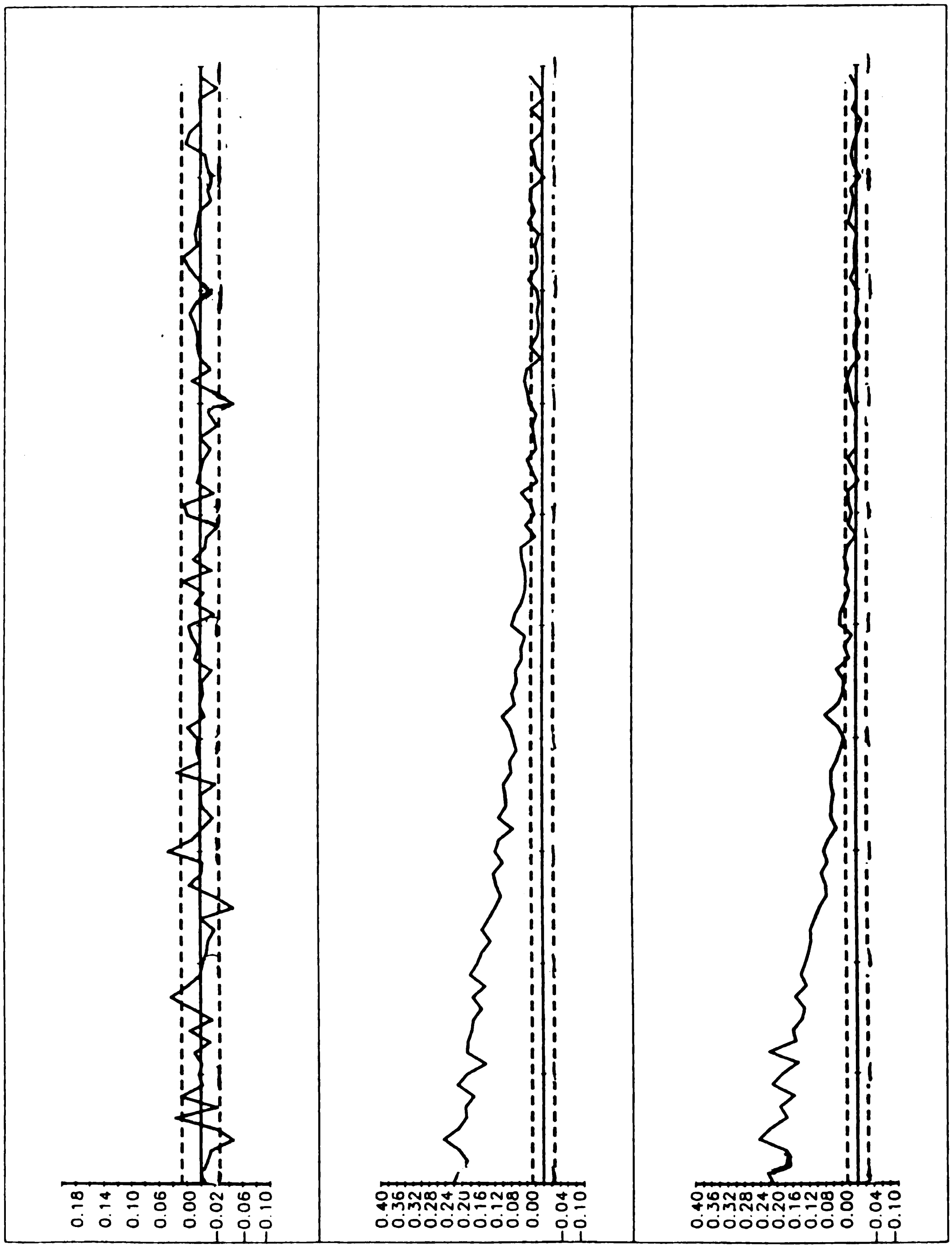


Figure 4.3 (cont'd).

f. **Unleaded Gasoline**

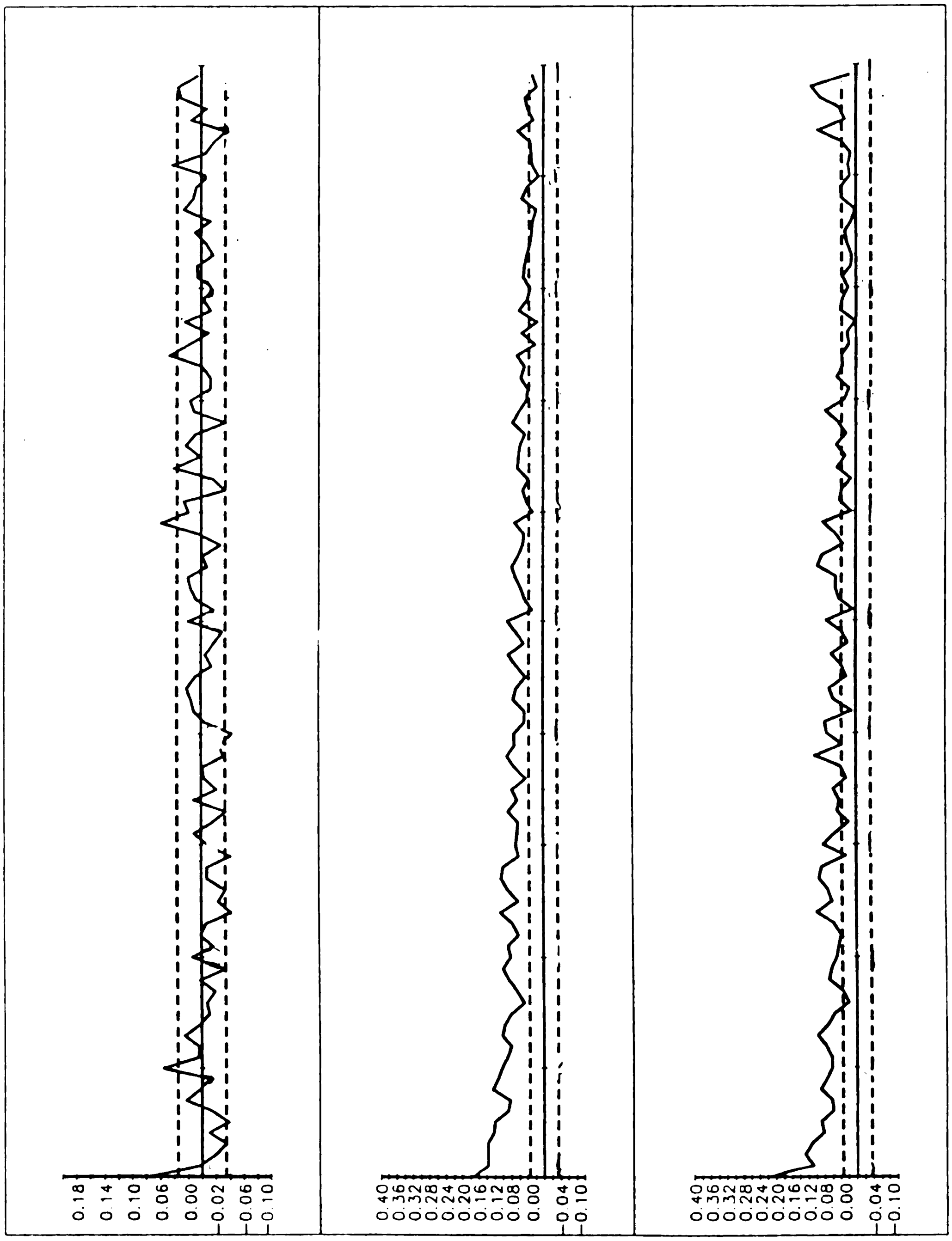


Figure 4.4: Autocorrelations for future returns, absolute and squared returns, and intraday range
a. Coffee

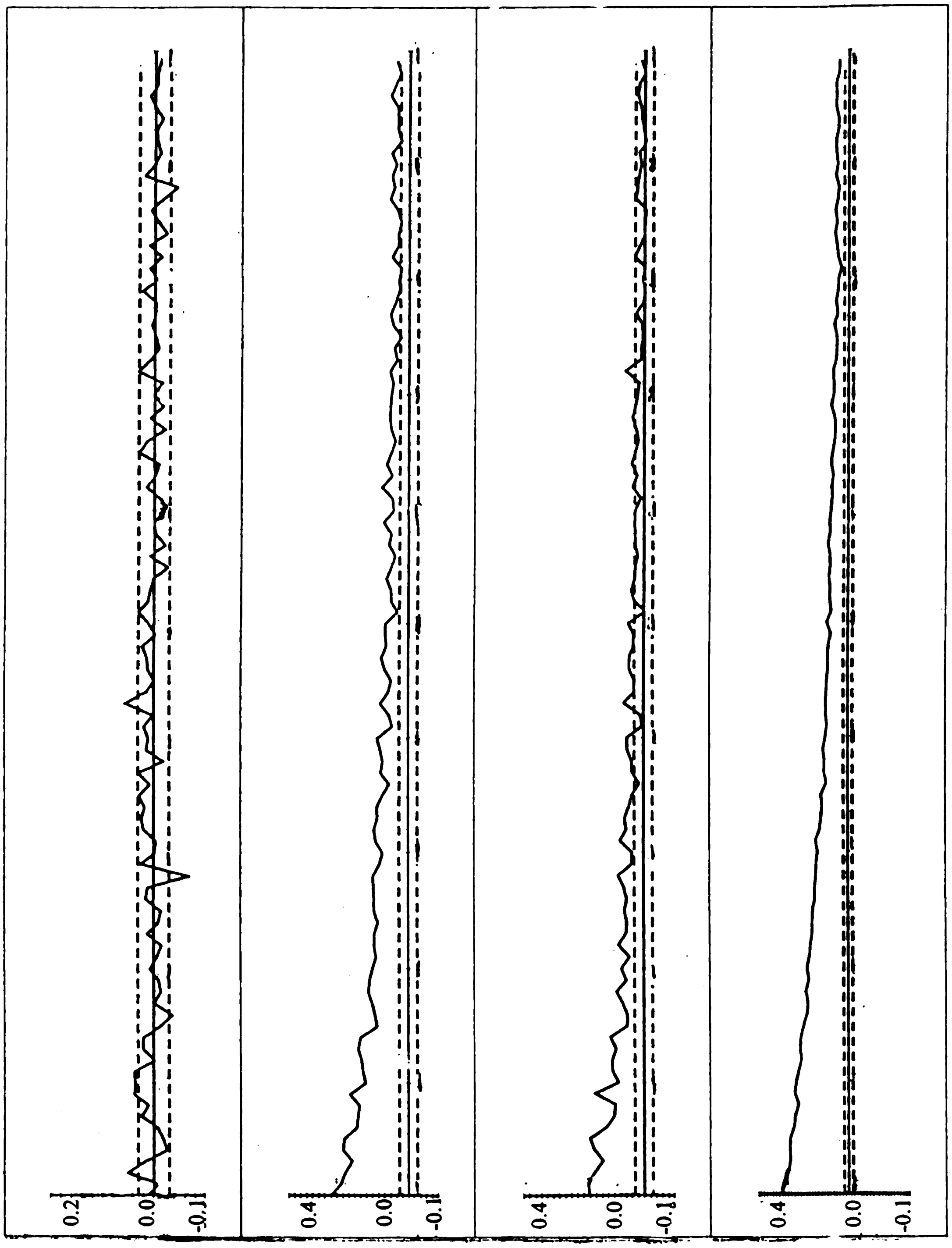


Figure 4.4 (cont'd).

b. Corn

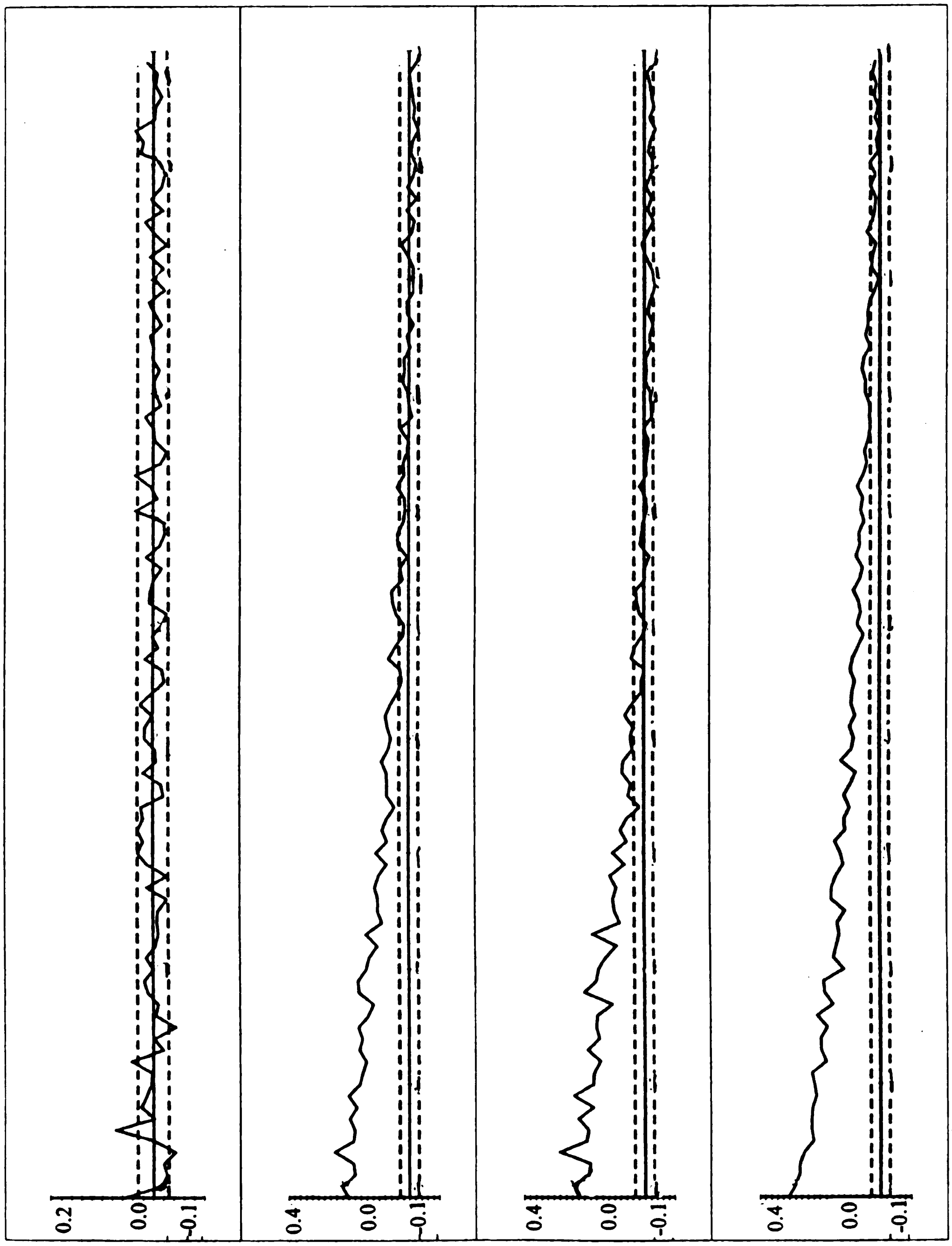


Figure 4.4 (cont'd).

c. Gold

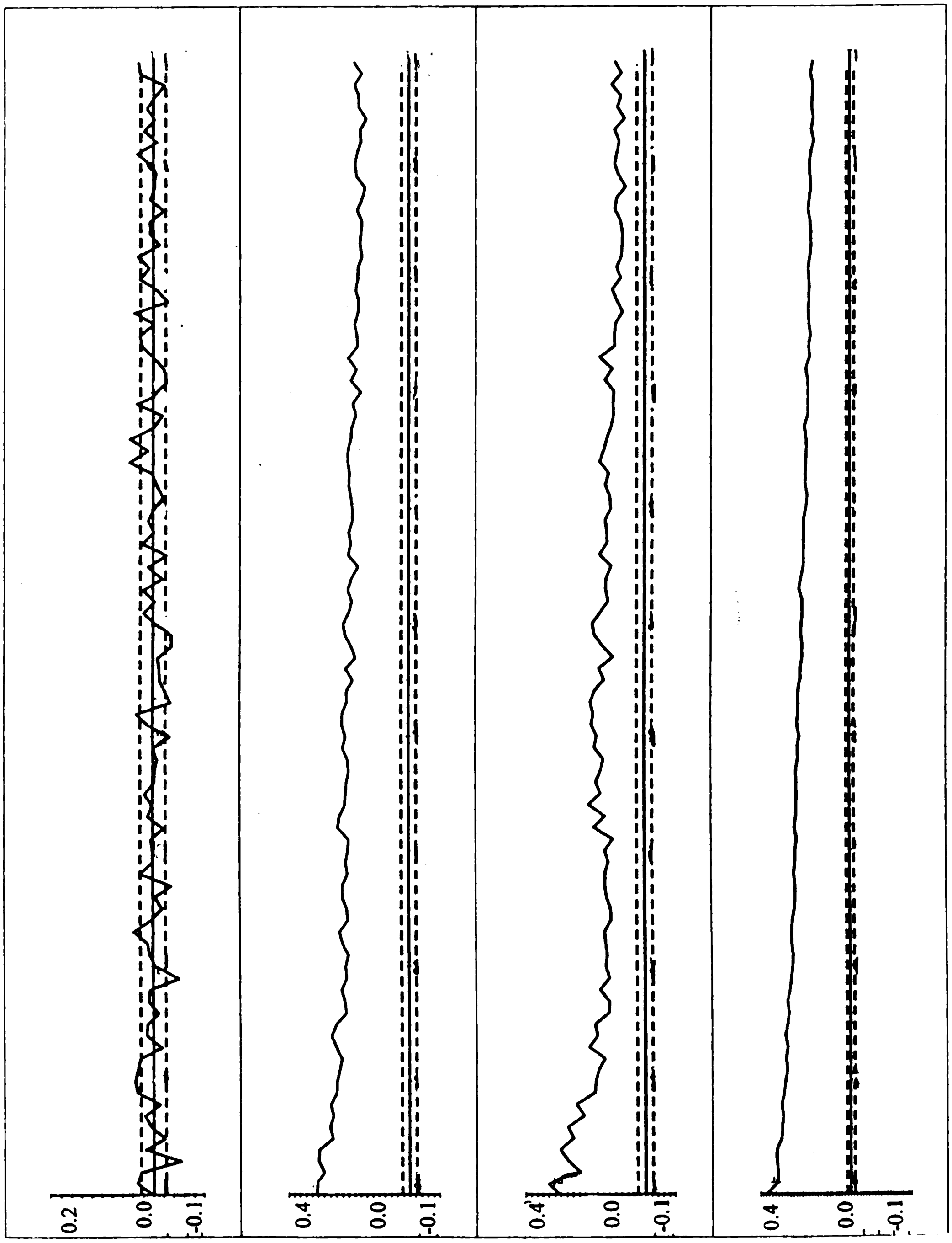


Figure 4.4 (cont'd).

d. Silver

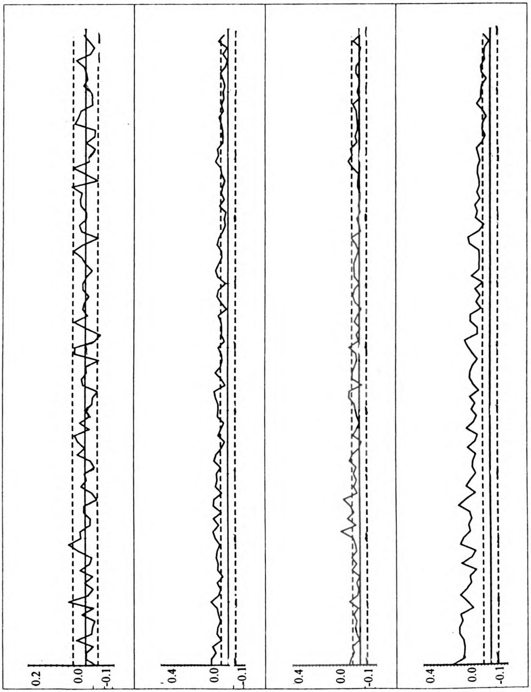


Figure 4.4 (cont'd).

e. Soybean

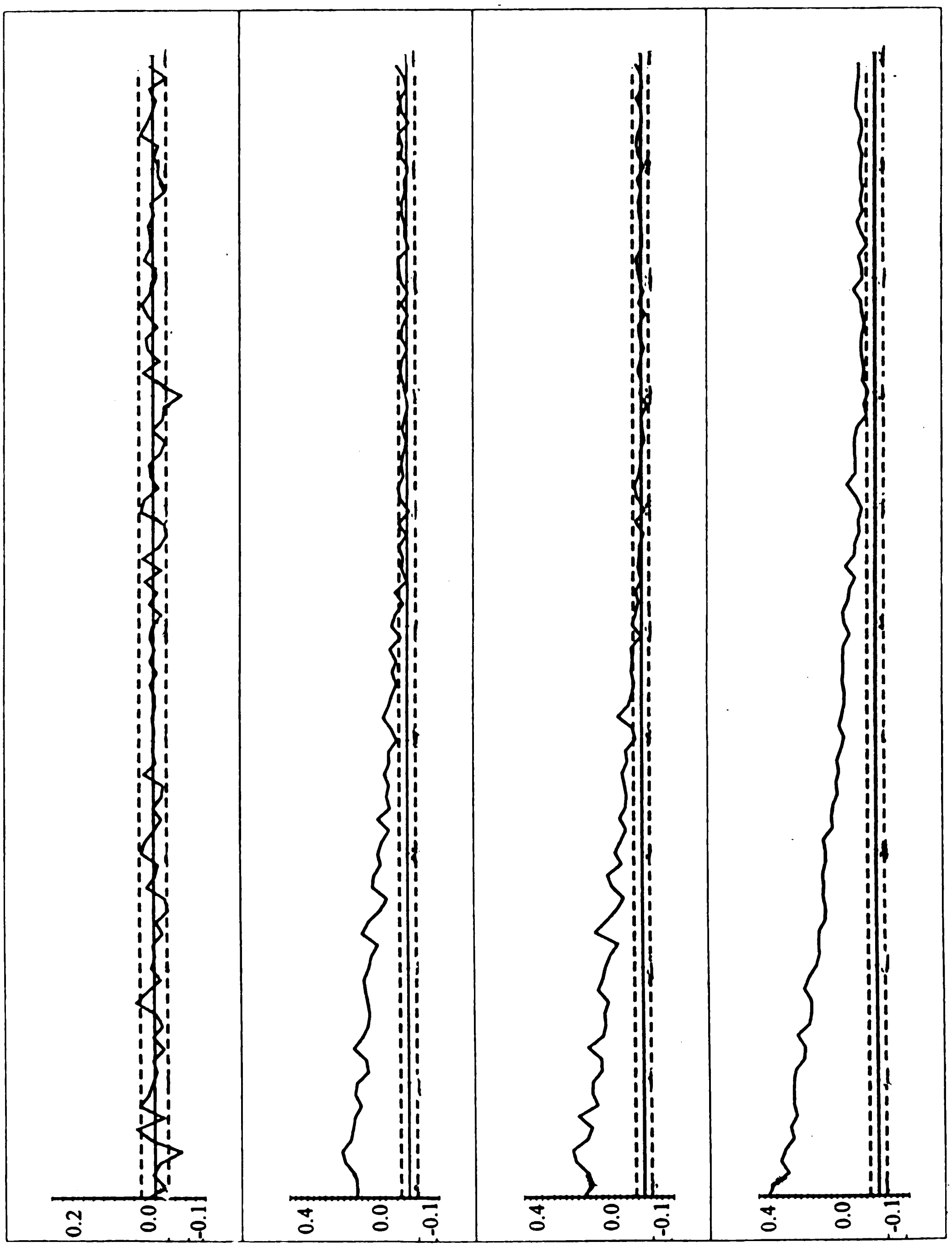


Figure 4.4 (cont'd).

f. Unleaded Gasoline

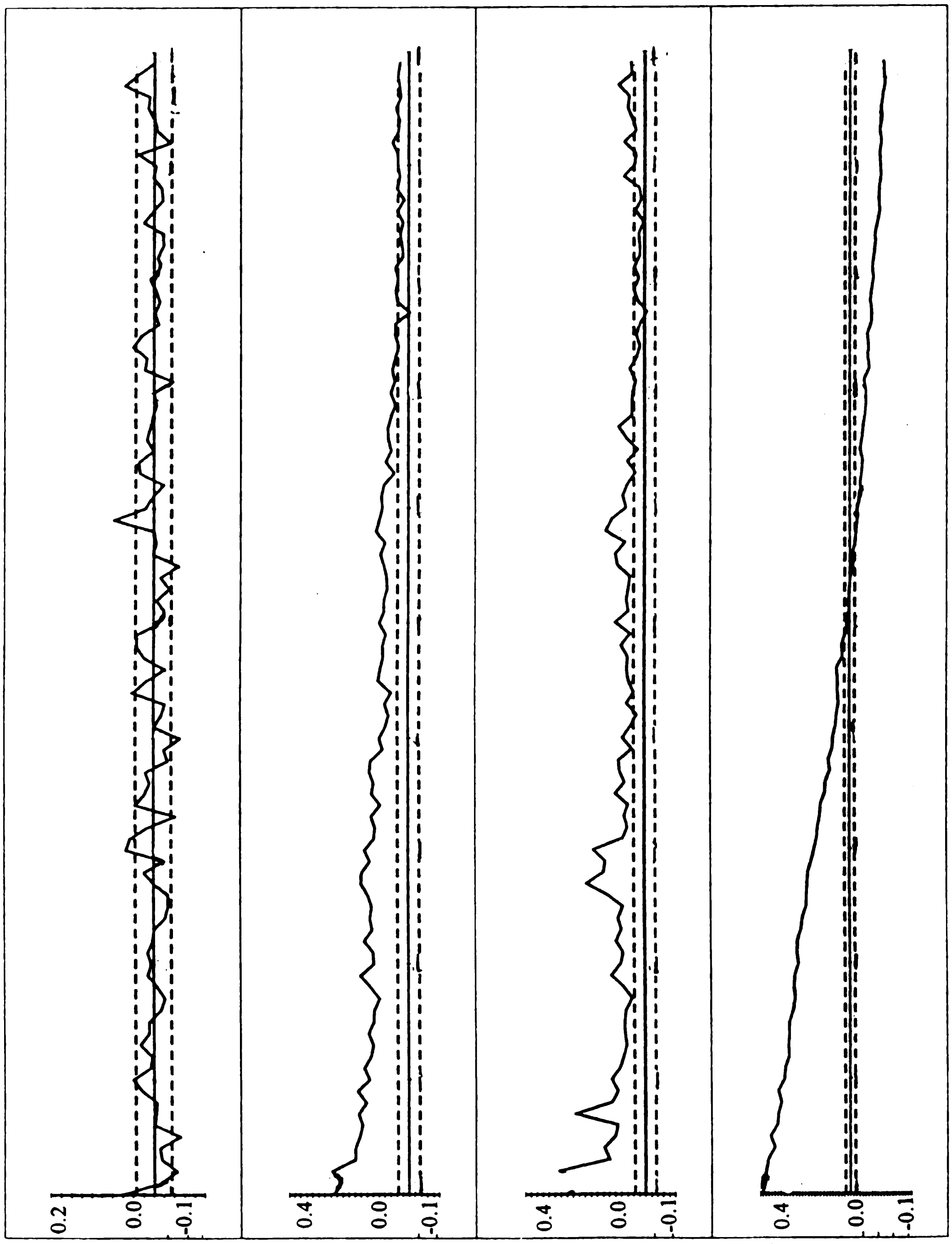


Figure 4.5: Future returns and estimated conditional variances

a. Coffee

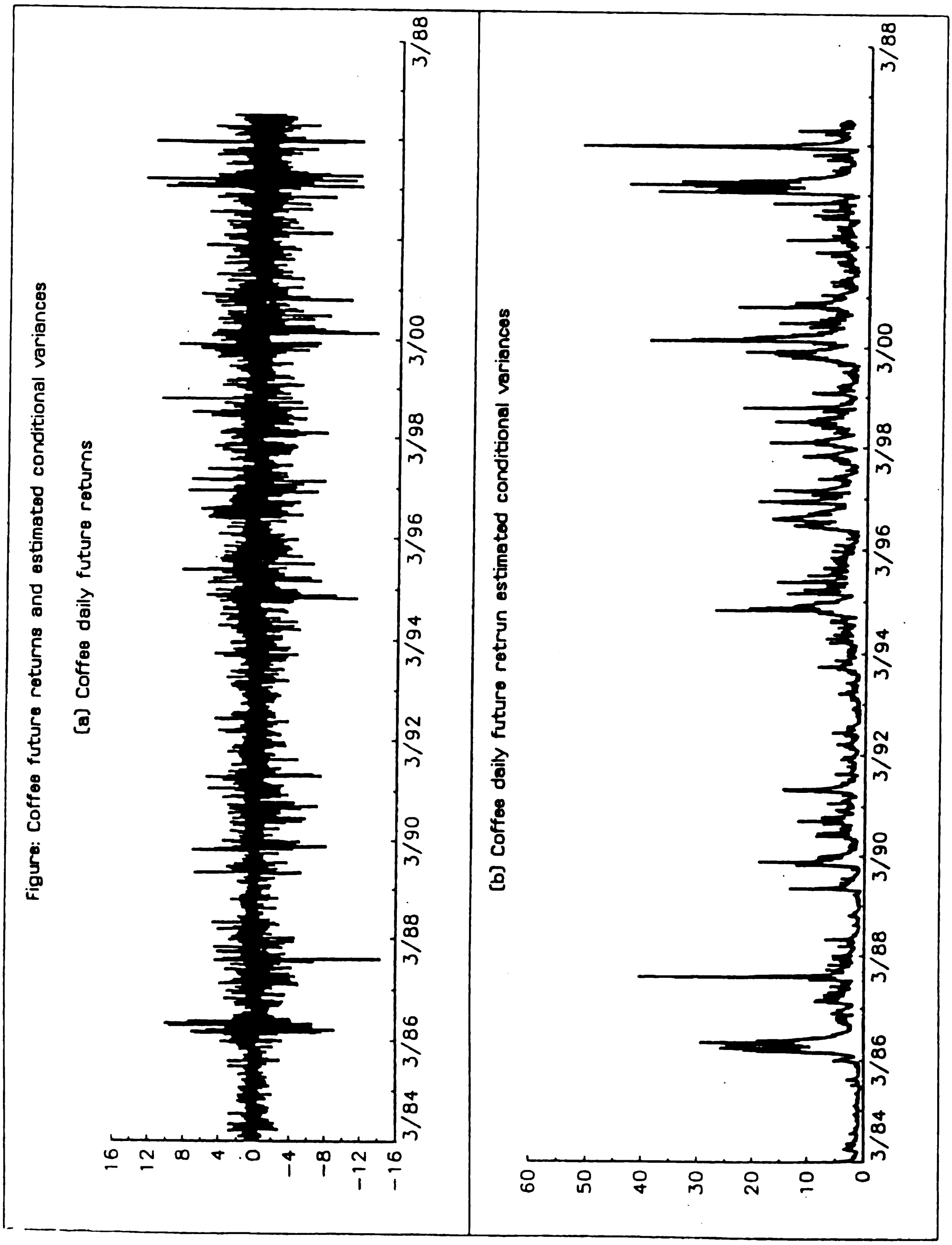


Figure 4.5 (cont'd).

b. Corn

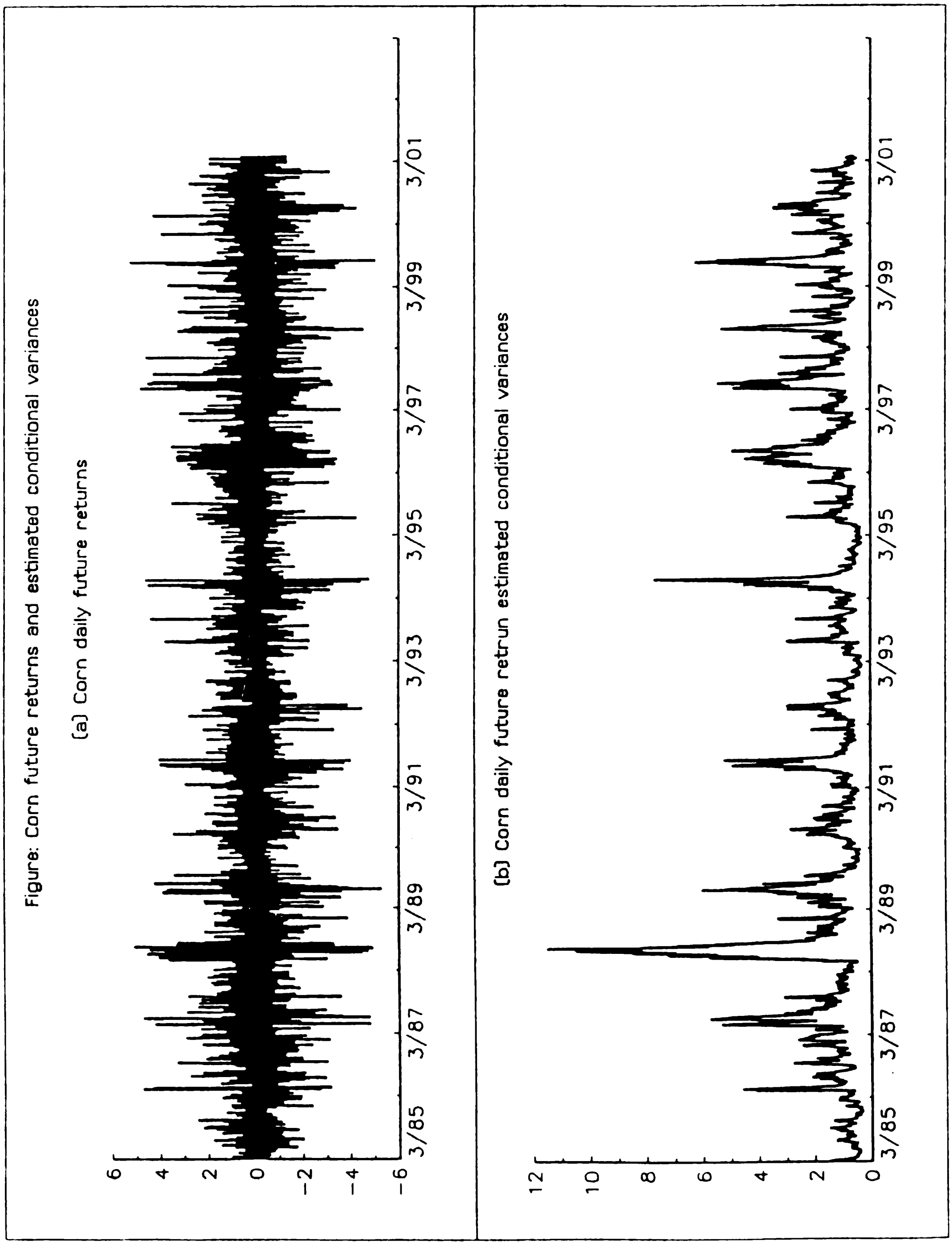


Figure 4.5 (cont'd).

c. Gold

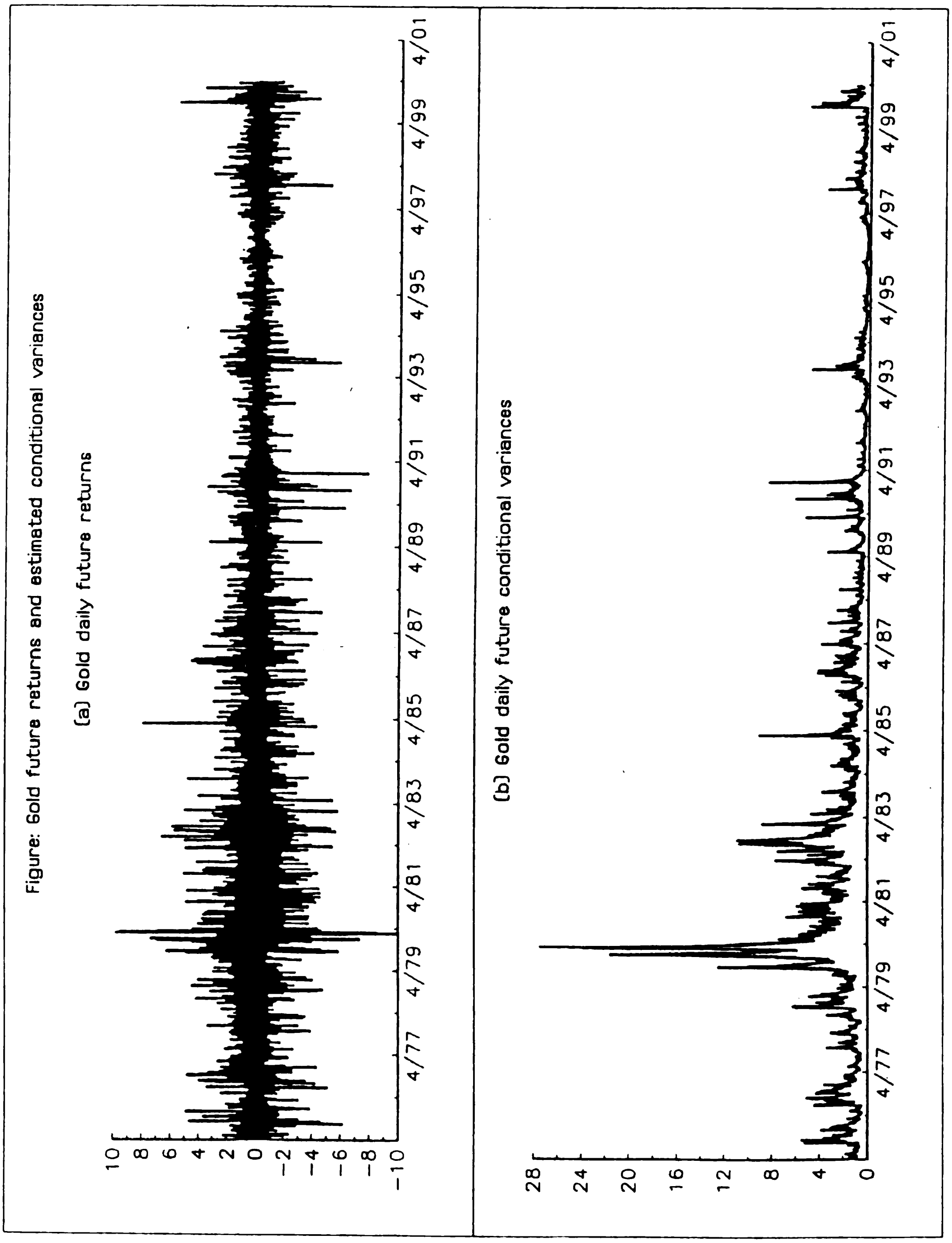


Figure 4.5 (cont'd).

d. Silver

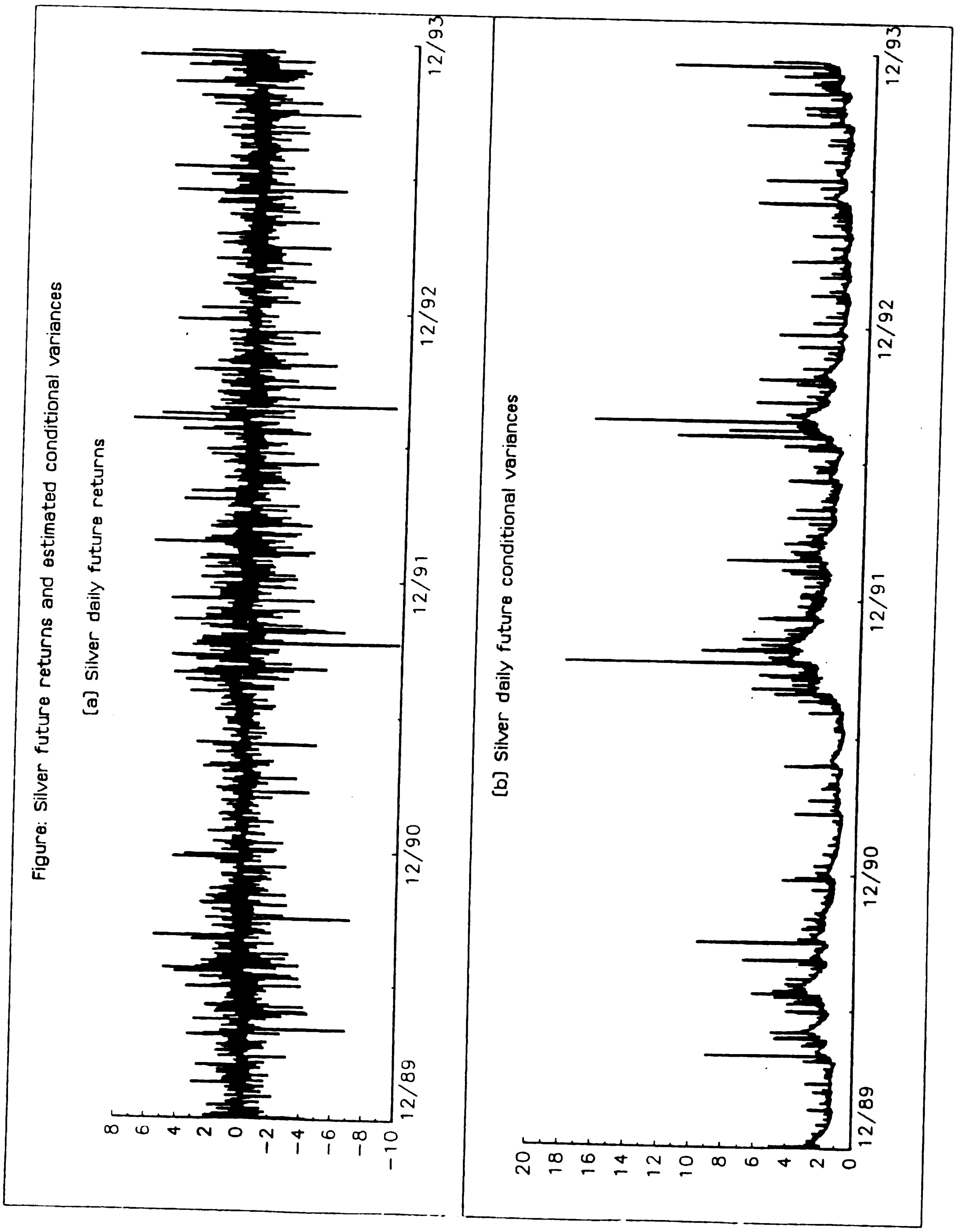


Figure 4.5 (cont'd).

e. Soybean

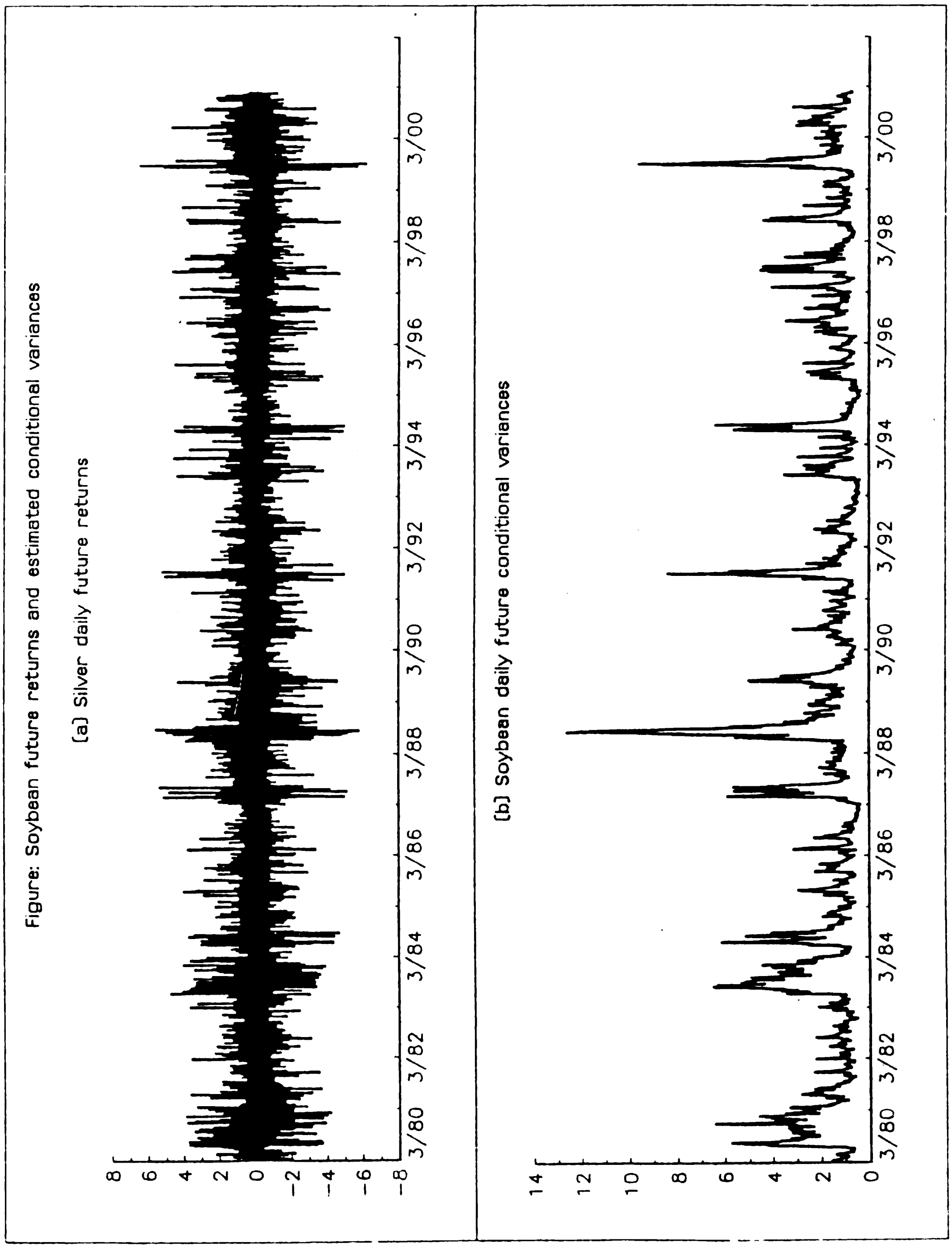


Figure 4.5 (cont'd).

f. Unleaded Gasoline

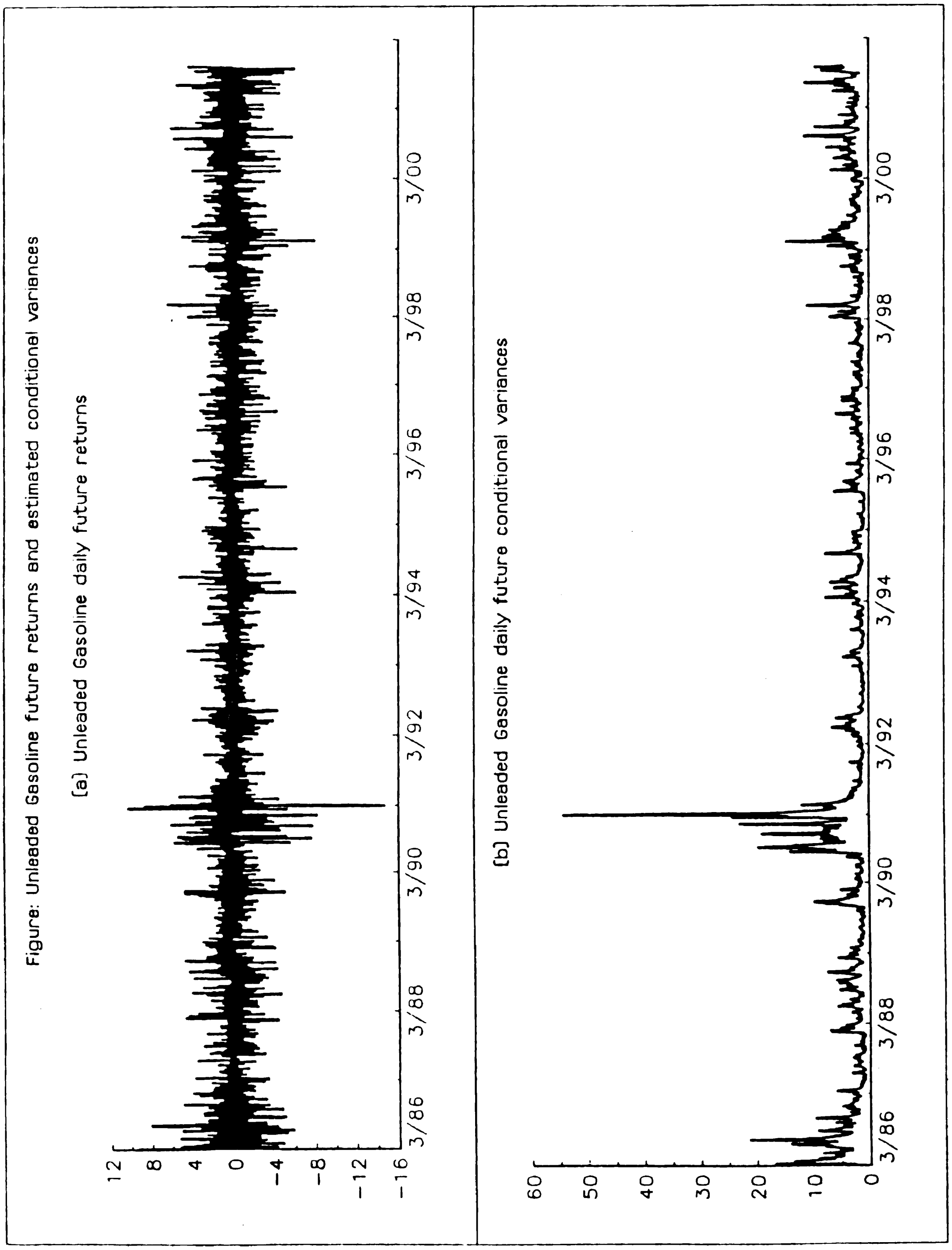


Table 4.1: Summary statistics for commodity future and cash returns

	mean	med	min	max	var.	skew.	kurt.
coffee	-0.020	0.007	-14.247	12.739	4.453	-0.275	7.289
	-0.020	0.000	-14.458	21.328	4.544	0.008	12.950
corn	-0.022	0.000	-5.264	5.213	1.416	0.016	5.098
	-0.008	0.000	-7.486	7.903	2.168	-0.334	6.068
gold	-0.018	-0.025	-9.909	9.745	1.580	-0.046	10.056
	0.008	0.000	-7.750	9.291	1.591	0.116	9.928
silver	-0.015	0.000	-9.776	7.801	2.082	-0.241	7.709
	0.009	0.000	-9.432	5.827	1.805	-0.280	7.128
soybean	-0.002	0.000	-6.172	6.433	1.591	-0.070	5.201
	-0.004	0.000	-11.490	7.867	1.936	-0.446	7.049
u. gas.	0.039	0.034	-14.618	10.285	2.754	-0.202	7.768
	-0.018	0.000	-18.251	12.573	6.189	-0.242	6.882

Table 4.2: Summary statistics for commodity future absolute and squared returns and intraday range

	mean	med	min	max	var.	skew.	kurt.
coffee	1.480	1.022	0.000	14.247	2.263	2.368	12.213
	4.453	1.045	0.000	202.979	124.949	7.693	88.993
	1.542	1.271	0.000	9.866	1.303	1.838	9.038
corn	0.871	0.657	0.000	5.264	0.657	1.855	7.447
	1.416	0.432	0.000	27.714	8.215	4.326	26.218
	0.938	0.830	0.000	5.676	0.262	1.857	9.886
gold	0.813	0.503	0.000	9.909	0.919	2.842	15.267
	1.580	0.253	0.000	98.190	22.619	8.755	119.255
	0.749	0.556	0.000	7.162	0.542	2.469	13.416
silver	0.996	0.677	0.000	9.776	1.090	2.440	12.565
	2.081	0.458	0.000	95.568	29.124	7.934	98.294
	0.168	0.000	0.000	4.535	0.213	4.300	27.064
soybean	0.922	0.678	0.000	6.433	0.742	1.898	7.652
	1.592	0.460	0.000	41.385	10.656	4.537	30.726
	0.945	0.822	0.000	5.381	0.293	1.763	8.452
u. gas.	1.187	0.877	0.000	14.618	1.347	2.462	14.960
	2.755	0.769	0.000	213.693	51.209	12.336	279.598
	0.789	0.554	0.000	13.058	0.872	2.355	17.903

Table 4.3: KPSS and Phillips-Perron test results for commodity future log prices levels, returns, absolute returns, squared returns and intraday range
a. KPSS Test: Commodity Future Prices:

series	coffee	corn	gold	silver	soybean	u. gasoline
level	2.682	2.804	5.930	3.698	2.892	2.813
return	0.080	0.079	0.160	0.163	0.041	0.107
squared return	2.333	0.180	4.779	0.305	0.484	0.570
absolute return	3.633	0.245	9.202	0.414	0.798	0.653
intraday range	4.909	0.351	9.619	0.911	1.578	0.115

b. Phillips-Perron Test: Commodity Future Prices:

level	-2.096	-2.551	-1.908	-2.400	-3.026	-3.254
return	-64.127	-60.126	-79.168	-46.425	-71.684	-52.950
squared return	-53.998	-51.024	-59.158	-41.983	-59.146	-43.791
absolute return	-49.740	-51.875	-57.835	-40.487	-60.718	-43.756
intraday range	-40.084	-46.548	-42.264	-35.396	-49.315	-28.550

c. KPSS Test: Commodity Cash Prices

level	2.401	1.856	6.459	4.643	1.663	1.238
return	0.089	0.054	0.270	0.145	0.044	0.063
squared return	6.055	0.278	4.952	0.295	0.375	0.412
absolute return	11.567	0.335	8.945	0.356	0.658	1.099

d. Phillips-Perron: Commodity Cash Prices

level	-1.447	-2.065	-1.956	-2.019	-2.862	-3.004
return	-61.993	-59.290	-82.966	-45.274	-73.017	-51.983
squared return	-47.393	-49.861	-53.496	-39.515	-57.993	-47.105
absolute return	-52.613	-48.496	-53.590	-39.029	-58.757	-44.989

Table 4.4: Estimated $MA-GARCH$ Models for the commodity future returns

	coffee	corn	gold	silver	soybean	u. gasoline
μ	-0.044	-0.018	0.036	-0.045	-0.028	0.028
	(0.025)	(0.016)	(0.010)	(-0.030)	(0.014)	(0.024)
θ	0.040	0.048	.	.	.	.
	(0.017)	(0.018)	.	.	.	.
ω	0.042	0.033	0.002	0.019	0.035	0.055
	(0.015)	(0.009)	(0.001)	(0.010)	(0.008)	(0.019)
α	0.110	0.096	0.053	0.026	0.085	0.097
	(0.016)	(0.012)	(0.009)	(0.007)	(0.009)	(0.017)
β	0.888	0.882	0.949	0.956	0.893	0.883
	(0.016)	(0.015)	(0.008)	(0.009)	(0.011)	(0.021)
$\ln(\ell)$	-8530.280	-6075.413	-8935.093	-3505.472	-8205.102	-5703.621
Skewness	-0.122	-0.318	-0.304	-0.166	0.045	-0.113
Kurtosis	4.900	7.026	7.183	7.118	4.381	4.355
Q_{20}	22.544	26.091	28.905	28.300	21.135	28.822
Q^2_{20}	30.579	19.145	29.009	11.033	36.399	17.683
T	4206	4055	6295	2002	5267	3153

Key: $\ln(\ell)$ is the maximized log likelihood. The number in parenthesis indicate the asymptotic robust QMLE standard errors of the corresponding parameter estimates. The Q_{20} and Q^2_{20} are the Ljung-Box statistics at 20 degrees of freedom based on the standardized residuals and squared standardized residuals respectively. The skewness and kurtosis are based on the standardized residuals. T is the sample size.

Table 4.5: Estimated $MA - FIGARCH$ Models for the commodity future returns

	coffee	corn	gold	silver	soybean	u. gasoline
μ	-0.037	-0.018	0.034	-0.042	-0.029	0.027
	(0.025)	(0.016)	(0.010)	(-0.030)	(0.014)	(0.024)
θ	0.040	0.050	.	.	.	0.038
	(0.018)	(0.017)	.	.	.	(0.019)
δ	0.533	0.582	0.424	0.241	0.546	0.541
	(0.085)	(0.121)	(0.050)	(0.040)	(0.106)	(0.089)
ω	0.062	0.036	0.015	0.198	0.041	0.136
	(0.024)	(0.012)	(0.006)	(0.046)	(0.014)	(0.032)
β	0.684	0.653	0.691	0.579	0.650	0.451
	(0.070)	(0.097)	(0.073)	(0.024)	(0.092)	(0.095)
ϕ	0.326	0.162	0.388	0.420	0.168	.
	(0.062)	(0.052)	(0.066)	(0.024)	(0.045)	.
$\ln(\ell)$	-8514.112	-6080.833	-8907.757	-3512.494	-8209.061	-5706.865
Skewness	-0.148	-0.121	-0.318	-0.112	0.064	-0.130
Kurtosis	4.669	4.201	7.026	7.290	4.531	4.317
Q_{20}	24.182	27.984	26.091	27.913	21.998	23.087
Q^2_{20}	30.796	31.805	19.145	9.629	36.046	12.738
T	4206	4055	6295	2002	5267	3153
$W_{\delta=0}$	39.351	23.206	73.116	36.644	26.358	.

Key: $W_{\delta=0}$ stands for the robust Wald test statistics testing the null of a $GARCH(1,1)$ model against a $FIGARCH(1,\delta,1)$ model. The rest of the table is same as Table 4.4.

Table 4.6: Estimated $MA - GARCH$ Models for the commodity cash returns

	coffee	corn	gold	silver	soybean	u. gasoline
μ	-0.066	0.019	0.008	-0.025	-0.010	0.036
	(0.039)	(0.019)	(0.010)	(-0.027)	(0.015)	(0.040)
δ	0.151	.	.	.	.	.
	(0.030)	.	.	.	.	.
θ	-0.097	.	-0.060	.	.	0.080
	(0.038)	.	(0.016)	.	.	(0.019)
ω	0.022	0.041	0.009	0.030	0.028	0.111
	(0.013)	(0.011)	(0.009)	(0.018)	(0.007)	(0.044)
α	0.094	0.107	0.081	0.041	0.084	0.076
	(0.015)	(0.013)	(0.044)	(0.015)	(0.010)	(0.017)
β	0.909	0.878	0.920	0.943	0.903	0.907
	(0.014)	(0.014)	(0.045)	(0.022)	(0.011)	(0.022)
$\ln(\ell)$	-7970.439	-6890.949	-9006.606	-3338.934	-8598.366	-7096.422
Skewness	-0.381	-0.399	-0.023	-0.167	0.182	-0.088
Kurtosis	13.492	4.800	11.233	5.997	4.490	4.352
Q_{20}	30.029	32.118	50.366	26.312	26.110	22.398
Q_{20}^2	18.066	19.05	23.071	25.830	23.917	17.398
T	4206	4055	6295	2002	5267	3153

Key: d is the long memory parameter in the $ARFIMA(0, d, 1)$ model that is fitted to conditional mean of coffee cash returns. The rest of the table is same as Table 4.4.

Table 4.7: Estimated $MA - FIGARCH$ Models for the commodity cash returns

	coffee	corn	gold	silver	soybean	u. gasoline
μ	-0.074	0.019	-0.009	-0.021	-0.012	-0.031
	(0.034)	(0.019)	(0.009)	(-0.027)	(0.015)	(0.040)
d	0.074	.	.	.	.	.
	(0.017)	.	.	.	.	.
θ	0.074	0.030	-0.063	.	.	0.084
	(0.022)	(0.019)	(0.014)	.	.	(0.019)
δ	0.367	0.499	0.342	0.268	0.668	0.438
	(0.047)	(0.144)	(0.034)	(0.046)	(0.158)	(0.097)
ω	0.172	0.110	0.042	0.137	0.036	0.216
	(0.077)	(0.028)	(0.022)	(0.034)	(0.010)	(0.102)
β	0.235	0.396	0.500	0.570	0.738	0.602
	(0.069)	(0.164)	(0.127)	(0.025)	(0.107)	(0.127)
ϕ	.	.	0.313	0.429	0.151	0.280
	.	.	(0.130)	(0.025)	(0.059)	(0.084)
$\ln(\ell)$	-8013.027	-6887.237	-8824.516	-3335.056	-8604.379	-7098.089
Skewness	-0.841	-0.380	-0.034	-0.111	-0.156	-0.089
Kurtosis	14.736	4.739	8.786	5.778	4.527	4.445
Q_{20}	30.775	28.622	53.507	29.054	25.693	22.824
Q^2_{20}	22.495	32.006	7.564	20.588	27.344	11.748
T	4206	4055	6295	2002	5267	3153
$W_{\delta=0}$	.	.	99.924	34.201	17.879	41.593

Key: $W_{\delta=0}$ stands for the robust Wald test statistics testing the null of a $GARCH(1,1)$ model against a $FIGARCH(1,\delta,1)$ model. The rest of the table is same as Table 4.4.

Table 4.8: GPH estimation results the cash returns, squared and absolute returns
a. Cash Returns

m	coffee	corn	gold	silver	soybean	u. gasoline
$T^{0.55}$	-0.002	0.132	-0.004	-0.071	-0.135	-0.154
	(-0.037)	(2.023)	(-0.077)	(-1.313)	(-2.220)	(-2.206)
$T^{0.65}$	0.078	0.081	0.082	-0.071	-0.037	-0.041
	(1.838)	(1.876)	(2.193)	(-1.313)	(-0.943)	(-0.887)
$T^{0.75}$	0.047	0.078	0.010	-0.027	-0.080	-0.057
	(1.669)	(2.732)	(0.430)	(-0.717)	(-3.087)	(-1.832)

b. Cash squared Returns:

$T^{0.55}$	0.276	0.478	0.496	0.385	0.345	0.358
	(6.467)	(7.296)	(8.536)	(4.845)	(5.663)	(5.121)
$T^{0.65}$	0.276	0.372	0.399	0.170	0.474	0.248
	(6.467)	(8.630)	(10.668)	(3.125)	11.971	(5.299)
$T^{0.75}$	0.247	0.210	0.355	0.146	0.337	0.211
	(8.795)	(7.393)	(14.689)	(3.925)	13.064	(6.730)

c. Cash absolute returns

$T^{0.55}$	0.455	0.519	0.496	0.438	0.435	0.394
	(7.020)	(7.927)	(8.542)	(5.512)	(7.143)	(5.633)
$T^{0.65}$	0.373	0.416	0.421	0.264	0.463	0.298
	(8.755)	(9.650)	(11.260)	(4.848)	(11.681)	(6.371)
$T^{0.75}$	0.277	0.281	0.370	0.164	0.334	0.249
	(9.875)	(9.860)	(15.351)	(4.423)	(12.945)	(7.944)

Key: m stands for the number of periodogram ordinates used in the GPH estimator. The values in parentheses are the t statistics for testing the null of $H_0 : \delta = 0$ versus the alternative of $H_1 : \delta > 0$. The t values are computed by using the theoretical variance of $\pi^2/24m$.

Table 4.9: GPH estimation results the future returns, squared and absolute returns and intraday range

a. Future Returns

m	coffee	corn	gold	silver	soybean	u. gaso-line
$T^{0.55}$	-0.023	0.039	-0.008	-0.055	-0.040	0.078
	(-0.357)	(0.599)	(-0.133)	(-0.692)	(-0.659)	(1.115)
$T^{0.65}$	0.045	0.096	0.078	-0.075	-0.020	0.029
	(1.044)	(2.219)	(2.087)	(-1.375)	(-0.508)	(0.629)
$T^{0.75}$	-0.008	-0.021	0.009	-0.057	-0.033	-0.053
	(-0.279)	(-0.721)	(0.360)	(-1.548)	(-1.268)	(-1.680)

b. Future squared Returns:

$T^{0.55}$	0.219	0.429	0.444	0.307	0.413	0.437
	(3.377)	(6.549)	(7.655)	(3.857)	(6.783)	(6.246)
$T^{0.65}$	0.327	0.419	0.370	0.170	0.382	0.347
	(7.673)	(9.717)	(9.901)	(3.124)	9.638	(7.409)
$T^{0.75}$	0.271	0.354	0.415	0.084	0.365	0.262
	(9.652)	(12.452)	(17.191)	(2.260)	14.152	(8.376)

c. Future absolute returns

$T^{0.55}$	0.375	0.400	0.464	0.339	0.442	0.519
	(5.796)	(6.110)	(7.993)	(4.268)	(7.257)	(7.421)
$T^{0.65}$	0.401	0.367	0.403	0.211	0.441	0.316
	(9.411)	(8.503)	(10.779)	(3.873)	(11.137)	(6.756)
$T^{0.75}$	0.314	0.336	0.350	0.162	0.366	0.285
	(11.170)	(11.807)	(14.487)	(4.365)	(14.194)	(9.104)

d. Future intraday ranges

$T^{0.55}$	0.468	0.421	0.483	0.415	0.476	0.558
	(7.218)	(6.429)	(8.324)	(5.219)	(7.827)	(7.979)
$T^{0.65}$	0.515	0.480	0.490	0.370	0.532	0.531
	(12.079)	(11.123)	(13.115)	(6.802)	(13.440)	(11.352)
$T^{0.75}$	0.415	0.374	0.409	0.239	0.395	0.501
	(14.785)	(13.156)	(16.959)	(6.448)	(15.299)	(16.014)

Key: Same as table (4.8).

Table 4.10: Local Whittle Estimates of long memory parameter for commodity cash and future returns and volatility proxies

a. Cash Series

Series	coffee	corn	gold	silver	soybean	u. gasoline
return	0.081	0.082	0.047	-0.072	-0.009	-0.180
	(0.051)	(0.051)	(0.038)	(0.076)	(0.052)	(0.053)
squared return	0.431	0.564	0.440	0.394	0.422	0.384
	(0.044)	(0.060)	(0.035)	(0.071)	(0.048)	(0.051)
absolute return	0.552	0.596	0.494	0.710	0.600	0.562
	(0.039)	(0.057)	(0.032)	(0.084)	(0.044)	(0.050)

b. Future Series

Series	coffee	corn	gold	silver	soybean	u. gasoline
return	0.094	-0.018	0.048	-0.043	-0.045	0.057
	(0.051)	(0.045)	(0.038)	(0.075)	(0.043)	(0.055)
squared return	0.379	0.599	0.349	0.323	0.472	0.408
	(0.043)	(0.064)	(0.031)	(0.067)	(0.049)	(0.054)
absolute return	0.552	0.538	0.503	0.473	0.598	0.583
	(0.041)	(0.057)	(0.032)	(0.074)	(0.044)	(0.052)
Intraday range	0.562	0.491	0.567	0.452	0.644	0.774
	(0.043)	(0.052)	(0.036)	(0.065)	(0.040)	(0.078)

Key: The values in parentheses are the robust standard errors.

CHAPTER 5

On the long memory properties of Emerging Capital Markets: Evidence from Istanbul Stock Exchange

5.1 Introduction

The presence of long memory components in stock returns has important implications for many of the paradigms of financial economics. If stock returns display long-term dependence, then they exhibit significant autocorrelation between observations widely separated in time. Since the series realizations are not independent over time, realizations from the remote past can help predict future returns, hence giving rise to the possibility of consistent speculative profits. This is in contrast to the martingale or random walk type behavior that many theoretical financial asset pricing models usually assume. Therefore, optimal consumption/savings and portfolio decisions may become sensitive to the investment horizon. The presence of long memory in asset returns contradicts the weak form market efficiency hypothesis, which states that, conditioning on past returns, future asset returns are unpredictable. A finding of long memory in asset returns calls into question linear modelling and invites the development of nonlinear pricing models at the theoretical level to account for long memory behavior. Mandlebrot (1971) observes that in the presence of long memory, the arrival of new market information can not be fully arbitraged away and martingale models of asset prices can not be obtained from arbitrage. If the underlying continuous stochastic processes of asset returns exhibit long memory, then the pricing derivatives by martingale models as well as statistical inference concerning asset pricing models based on standard testing procedures (Yajima, 1985) may not be appropriate.

Due to the theoretical and empirical importance of the issue, there is an extensive literature on analyzing the long memory properties of financial asset returns in major financial markets. Greene and Fielitz (1977), by using the R/S statistic of Hurst (1951), test long-term dependence in the daily returns of 200 individual stocks on the New York Stock Exchange from December 23, 1963, to November 29, 1968, and report evidence of persistence. Aydogan and Booth (1988) used also the original R/S analysis to test for long memory in common stock returns. Lo (1991), by using a modified version of the R/S statistic which controls the possible short term dependencies in the data, found no evidence in favor of long memory of the monthly and daily returns on Center for Research in Security Prices (CRSP) stock indexes. Ding, Granger, and Engle (1993) examined the long memory properties of several transformations of the absolute value of daily returns on the Standard and Poor's (S&P) 500, and obtained considerable evidence of long memory in the squared and absolute returns. Crato (1994), used the exact maximum likelihood method of Sowell (1992), and found no evidence of long memory for the stock return series of G-7 countries. By using both the modified R/S method of Lo (1991), and the Geweke and Porter-Hudak (1983) (GPH) method, Cheung and Lai (1995) found no evidence of persistence in several international stock return series. Lobato and Savin (1998) test the presence of long memory in daily returns and their squares on S&P 500 series by using semi-parametric procedures. Their test results indicate no evidence for long memory in the levels of daily returns but evidence of long memory in absolute and squared returns.

Despite the extant literature that analyzes the long memory properties of major stock markets prices, there is little research done on the time series properties of Emerging Markets asset prices. Outside the world's developed economies, there is a host of emerging capital markets (ECM) in Europe, Latin America, Asia, the Middle East and Africa. As pointed out by Harvey (1995) compared to developed markets, ECMs exhibit higher expected returns as well as higher volatility. Due to low correlation with developed countries' stock markets, the unconditional portfolio risk of a world investor would be significantly reduced. These markets have attracted a great deal of attention from investors and investment funds seeking to further diversify their portfolios as these stock markets provide a new menu of opportunities for investors of the world. Despite temporary setbacks, ECMs continue to be important conduits of diversification, and a complete characterization and understanding of the dynamic behavior of stock returns in ECMs is warranted. One may think that ECMs are likely to exhibit characteristics different from those observed in developed capital markets. Barkoulas et al. (2000) recently analyzed the long memory properties of weekly Greek stock market data and obtained strong evidence of long memory in the conditional mean process, a finding contrary to the results from developed stock markets. One may expect biases due to market thinness and non-synchronous trading that is possibly more severe in the ECMs. Moreover, in contrast to developed capital markets, which are highly efficient in terms of the speed of information reaching all traders, investors in Emerging Capital Markets may tend to react slowly and gradually to new information. All these may lead one to expect ECMs stock returns behave differently and have distinct properties compared to developed capital markets.

The purpose of this chapter is to analyze the long memory properties of stock price returns in an emerging capital market; the Istanbul Stock Exchange (ISE). Specifically, the paper tries to answer the following question. Do daily and weekly ISE index returns have the *long memory* property, with index returns being approximately uncorrelated, and with very persistent autocorrelation in squared and absolute returns? To my knowledge, no study has analyzed the long memory dynamics of Istanbul Stock Exchange market returns.

The ISE, the only stock exchange in Turkey, was formally inaugurated in late 1985. The number of companies traded on the exchange increased from 80 at the end of 1986 to 262 at the end of 1998 (Yuksel 2000). The national market is the major component of the ISE. The total market capitalization of the firms traded has increased from 938 million US dollars at the end of 1986, to 56 billion US dollars at the middle of 1999. Turkey has one of the most liberal foreign exchange regimes in the world, with a fully convertible currency as well as a policy that allows foreign institutional and individual investments in securities listed on the ISE since 1989. Turkish stock and bonds markets are open to foreign investors, without any constraints on the repatriation of capital and profits. Just between the beginning of 1996 and the end of 1999 foreign investment in ISE has more than tripled. According to Yuksel (2000) about half of the floating equity in ISE is owned by foreign investors. These observations show that ISE is one of the important ECMs in the world economy and a better understanding of the dynamic properties of the ISE index returns will be useful not only for comparison purposes, but also for the international investors whose portfolios include equities from ISE.

This chapter uses the Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity (FIGARCH) model of Baillie et al. (1996). Since the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model attempts to account for volatility persistence, but has the feature that persistence decays relatively fast, we use the GARCH model as a benchmark and compare its results with the FIGARCH model, as the latter model is capable of modelling very long temporal dependencies in conditional variance of a process. In order to better asses the presence of long memory in the volatility of index returns, this chapter also models absolute returns and squared returns using Fractionally Integrated Autoregressive Moving Average (ARFIMA) model of Granger and Joyeux (1980), and Hosking (1981). Moreover, estimates of the long memory parameter for the volatilities of stock returns from semi-parametric methods are also obtained. Particularly, the GPH estimator from Geweke and Portar-Hudak (1983) and a local Whitlle estimator based on Fox and Taqque (1986) are used. The findings of the this chapter indicate presence of long memory in the volatility process of ISE 100 stock returns. Contrary to empirical evidence from some other ECMs, the conditional mean of ISE 100 daily and weekly dollar stock index returns do not posses the long memory component.

The rest of the chapter is organized as follows. Section 5.2 describes the data and examines the empirical autocorrelations of the series. Section 5.3 presents and discusses the empirical results. The last section provides the conclusion.

5.2 The Data

The data set consists of daily US dollar Turkish lira spot exchange rates and the Turkish stock index based on the closing prices of a value-weighted index comprising the top a hundred listed firms on the ISE National Market by their market capitalization. Exchange rate data is obtained from the Central Bank of the Republic of Turkey (CBRT), while ISE 100 index data is obtained from the ISE. In choosing the stocks included in the index, the stocks are ranked in a descending order according to market and daily average traded values. Those stocks that have the highest market values and daily average trading values are included in the ISE National-100 index.

The sample period spans 01/04/1988 to 09/28/2001 for a total of 3440 observations. The index used in this study is expressed in terms of US dollars in order to avoid the effect of local inflation risks. The base year for the index is adjusted so that the index at 01/04/1988 is equal to 100. Then the following formula is used to convert the index into dollar denominated base; $100 \times \frac{P_t}{S_t} S_{base}$, where P_t is the index at time t, S_t is the spot exchange rate at date t and S_{base} is the spot exchange rate at base date. The weekly index series is constructed from the daily data by taking the index corresponding to Thursday of the week. In cases where data is not available for Thursdays, Wednesday data is used.

Following the standard practice, the stock returns are defined as $R_t = 100 \times \Delta \ln(P_t)$, where P_t is the stock index at date t, absolute returns as $|R_t|$, and squared returns as R_t^2. Figure 1 gives the graphs of the daily stock index returns, absolute returns and squared returns over the sample period. It appears from the graphs that relatively volatile periods, characterized by large price changes, alternate with more tranquil periods in which the index remains more or less stable. This indicates that large index returns (both positive and negative) seem to occur in clusters and so does volatility. The volatility clustering phenomenon which is typical of asset prices and exchange rates, seems to occur in the ISE as well.

Summary statistics for the index returns are given in table 5.1. The table indicates that both daily and weekly stock returns have small negative means and medians over the sample period. One of the usual ways of getting an idea of the distribution of a time series y_t is to look at the *kurtosis* and the *skewness* and compare them with that of a normal random variable. The last two columns of table 5.1 indicate that the kurtosis of both daily and weekly returns are much larger than that of a normal random variable. This reflects the fact that the tails of the distribution of index returns are fatter than the tails of the normal distribution. This in turn means that large observations occur more often than one might expect for a normally distributed variable.

Since any symmetric distribution have skewness equal to zero, table 5.1 indicates that the distribution of daily and weekly stock index returns have some asymmetry. The negative values of skewness indicate that for the ISE stock returns over the sample period considered, the left tail of the distribution is fatter than the right tail, or large negative returns tend to occur more often than large positive ones. The analysis here indicates that daily stock return distribution is far from being normal.

To gain some insight into the dependence structure of the series, figure 5.2 displays the first 100 autocorrelations for the daily stock index, index returns, absolute returns and squared returns together with two-sided 5 percent critical values ($\pm 1.96/\sqrt{T}$ where T is the sample size). The asymptotic critical values are not strictly valid for a process with ARCH effects. Still they may be considered to be useful as guidelines. It is clear from the figure that the ISE 100 log index has autocorrelations close to unity at all selected lags and, hence, it seems to mimic the correlation properties of a random walk process. There is a small, positive but significant first order autocorrelation in the stock index returns, while higher orders are not significant at conventional levels. On the other hand, for the absolute and squared returns, the autocorrelations start off at a moderate level (about 0.32) but remain significantly positive for a substantial number of lags. Moreover, autocorrelation in the absolute returns is generally somewhat higher than the autocorrelation in the squared returns. This illustrates what has become known as the 'Taylor property' (see Taylor, 1986, pp.52-55), that is, when calculating the autocorrelations for the series R_t^δ for various values of δ, one almost invariably finds that autocorrelations are largest for $\delta = 1$. As is evident from the figure autocorrelations for absolute returns are not only larger than those of squared returns, but also much more persistent in the sense that they decay much more slowly. The autocorrelations in absolute and squared returns seem to mimic the correlation properties of a long memory processes rather than a short memory stationary process for which autocorrelations decay to zero at an exponential rate. As is evident from the figure, the autocorrelations in absolute and squared returns decay very slowly, indicating that linear association between distant observations is somewhat persistent and autocorrelations decay at a hyperbolic rate. This described behavior of autocorrelations in absolute and squared returns is consistent with the time series models with long memory or long range dependence. The above described characteristics of autocorrelations in the ISE 100 index, index returns, absolute and squared returns are in conformity with the findings from developed stock markets . For example, see Ding and Granger (1993).

5.3 Empirical Results

In light of the discussion in section 5.2, conditional variance of the ISE 100 stock index returns are modelled by the FIGARCH process which allows one to model persistence in the autocorrelations of index returns as well as volatility clustering phenomenon. The robust Wald statistic is used to check if the estimated FIGARCH model better represents the long memory property of the data compared to a GARCH specification. Results of the estimated $ARMA(P,Q) - FIGARCH(p,\delta,q)$ models for returns are represented in table 5.2. The estimate of long memory parameter, δ, for daily data is 0.538 and for the weekly returns it is 0.319. These estimates are significantly different from zero. Various tests for specification of the models were performed. In particular, a robust Wald test of a stationary $GARCH(1,1)$ model under the null hypothesis versus a $FIGARCH(1,\delta,1)$ model under the alternative hypothesis has a numerical value of 35.060, which shows a clear rejection of the null hypothesis when compared with the critical values of a χ^2di stribution with one degree of freedom. In none of the data frequencies the estimated $GARCH$ models performed better than the $FIGARCH$ models, and the sum of the estimates of α and

β in the $GARCH$ models were very close to one, indicating that the volatility process is highly persistent. In both daily and weekly returns the standardized residuals from the estimated models exhibit less skewness and kurtosis than the returns. The Box-Pierce portmanteau statistic, Q fails to reject the null hypothesis of independently and identically distributed squared standardize residuals at conventional significance levels.

The results from the $FIGARCH(1, \delta, 0)$ indicate that the conditional variance of ISE 100 index returns contain long memory. In the $FIGARCH$ model the long memory parameter corresponds to the squared error term. Hence, results from table 5.2 provide evidence that the squared stock returns exhibit long memory. To further investigate this issue, table 5.3 gives the estimates of the long memory parameter from the GPH, Conditional Sum of Squares (CSS), and the local Whittle estimation as applied to the squared and absolute returns. The results from table 5.3 indicate that both squared and absolute returns have statistically significant long memory. This result is supported from all estimation methods. Moreover, the findings also support the Taylor Effect. In general, the estimate of the long memory parameter is higher for the absolute returns than that of the squared returns. The results are in line with those of the FIGARCH estimates reported in table 5.2.

5.4 Conclusion

This chapter has investigated the volatility clustering and the long memory in an emerging capital market, namely Istanbul Stock Exchange, by utilizing the ISE National 100 daily and weekly index returns. The long memory $MA(1) - FIGARCH(1, \delta, 0)$ model is found to provide a good representation of the daily returns while a Martingale-$FIGARCH(1, \delta, 0)$ model is found to fit better for the weekly returns data. Estimates of the long memory parameter are found to be significantly different from zero, indicating that the ISE 100 index volatility is a long memory process, thus rejecting a GARCH specification.

Further analysis of squared and absolute returns supports the presence of long memory in the volatility process. In particular, autocorrelations of squared and absolute returns, and estimates from GPH, local Whittle, and CSS methods all support the findings from the FIGARCH model. Moreover, results from estimates of the long memory parameter provide evidence of the so-called Taylor Effect. The evidence of approximate Martingale behavior in the conditional mean of the ISE 100 index returns and the presence of long memory in absolute and squared returns is similar to that obtained from major capital markets in the literature. The finding of short memory in returns is in contrast to the evidence of long memory in the conditional mean of return process for some other Emerging Capital Markets. The evidence of the long memory component presented in this study may indicate that financial security prices are not immune to persistent informational asymmetries, especially over longer time spans. Following Anderson and Bollerslev (1997), if we interpret the volatility as a combination of heterogenous information arrivals then it may be argued that, despite the short memory information arrivals, the conditional variance of stock returns exhibit long memory characteristics. In this sense, the evidence of long memory is an intrinsic feature of the returns generating process. The finding of long memory both in daily and weekly frequency supports the argument that long memory is an intrinsic property of the return process rather than exogenous occasional shifts. To better understand this issue, it may be worthwhile to study dynamics of individual stock returns from Emerging Capital Markets. Moreover, use of high frequency data may also reveal important information on the long memory component of stock returns.

BIBLIOGRAPHY

[1] Anderson, T. G. and T. Bollerslev (1997), Heterogenous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns, *Journal of Finance,* **3,** 975-1005.

[2] Aydogan, K. G.G. Booth (1988), Are there long cycles in common stock returns?, *Southern Economic Journal* **55**, 141-149.

[3] Baillie, R. T., (1996), Long Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics,* **73,** 5-59.

[4] Baillie, R. T., (1998), Comment *Journal of Business & Economic Statistics,* **16,** 273-276.

[5] Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen (1996), Fractionally integrated Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* **74,** 3-30.

[6] Baillie, R.T., C.-F. Chung, and M.A. Tieslau (1996), Analyzing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics* **11,** 23-40.

[7] Baillie, R. T. , Y. W. Han, and Tae-Go Kwon (2001), Further long memory properties of inflationary shocks, forthcoming, *Southern Economic Journal.*

[8] Barkoulas, J. T., Baum, C. F., and Travlos, N. (2000), Long memory in the Greek stock market, *Applied Financial Economics,* **10,** 177-84.

[9] Beran, J. (1994), Statistics for Long-Memory Processes, Chapman & Hall

[10] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307-327.

[11] Bollerslev, T. and J. M. Wooldridge (1992), Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, *Econometric Reviews* **11,** 143-172.

[12] Bollerslev, T. and H.O.A. Mikkelsen (1996), Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73,** 151-184.

[13] Chung, C-F. and R.T. Baillie, 1993, Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models, *Empirical Economics* **18,** 791-806.

[14] Cheung, Y. and Lai, K. (1995), A search for long memory in international stock market returns, *Journal of Internationtal Money and Finance,* **14,** 597-615.

[15] Crato, N. (1994), Some international evidence regarding the stochastic behavior of stock returns, *Applied Financial Economics,* **4,** 33-9.

[16] Ding, Z., C.W.J. Granger, and R.F. Engle (1993), A long memory property of stock returns and a new model, *Journal of Empirical Finance,* **1,** 83-106.

[17] Fox, R., and Taqqu, M. S. (1986), Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics,* **14,** 517-532.

[18] Granger, C., (1980), Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics,* **14,** 227-238.

[19] Granger, C., and R. Joyeux (1980), An Introduction to Long Memory Time Models and Fractional Differencing, *Journal of Time Series Analysis,* **1,** 15-29.

[20] Green, M. T., and Fieltz B. D. (1977), Long-term Dependence in Common Stock Returns, *Journal of Financial Economics,* **4,** 339-349.

[21] Hosking, J. (1981), Fractional Differencing, *Biometrika,* **68,** 165-176.

[22] Hurst, H. (1951), Long Term Storage Capacity of Reservoirs, *Transactions of the American Society of Civil Engineers,* **116,** 770-799.

[23] Gweke, J. and Portar-Hudak, S. (1983), The estimation and application of long memory time series models, *Journal of Time Series Analysis* **4,** 221-238.

[24] Harvey, C. R. (1995), Predictable risk and returns in Emerging Markets, *The Review of Financial Studies,* **8,** 773-816.

[25] Hurvich, C. M. and Beltrao, K. I. (1994), Automatic semiparametric estimation of the parameter of a long memory time series, *Journal of Time Series Analysis,* **15,** 285-302.

[26] Hurvich, C.M., Deo, R. and Brodsky, J. (1998), The mean squared error of Gweke and Portar-Hudak's estimator of the long memory parameter of a long-memory time series, *Journal of Time Series Analysis* **19,** 19-46.

[27] Lee, S. W. and B. E. Hansen (1994), Asymptotic theory for the $GARCH(1,1)$ quasi-maximum likelihood estimator, *Econometric Theory* **10,** 29-52.

[28] Lo, A. W. (1991), Long-term memory in stock market prices, *Econometrica,* **59,** 1279-313.

[29] Lobato, I. N., and Savin, N. E. (1998), Real and spurious long-memory properties of stock-market data, (with discusssion), *Journal of Business & Economic Statistics,,* **16,** 261-283.

[30] Mandelbrot, B. B. (1971), When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models, *Review of Economics and Statistics,* **53,** 225-36.

[31] Robinson, P.M. (1990), Time series with strong dependence, Advances in econometrics, 6th world congress, Cambridge University Press, Cambridge.

[32] Robinson, P. M. (1995), Log-periodgram regression time series with long-range dependence *Annals of Statistics* **23,** 1048-72.

[33] Robinson, P.M. and F.J. Hidalgo, (1997), Time series regression with long-range dependence, *Annals of Statistics* **27,** 77-104.

[34] Samarov, A. and M.S. Taqqu (1988), On the efficency of the sample mean in long memory noise, *Journal of Time Series Analysis* **9,** 191-200.

[35] Sowell, F. (1992), Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics,* **53,** 165-188.

[36] Taylor, S. (1986), Modelling Financial Time Series, John Wiley & Sons, New York.

[37] Yajima, Y. (1985), On estimation of long memory time series models, *Australian Journal of Statistics* **27,** 303-320.

[38] Yajima, Y. (1991), Asymptotic properties of the LSE in a regression model with long-memory stationary errors, *Annals of Statistics* **19,** 158.

[39] Yuksel, S. A. (2000), Three essays on the microstructure of the Turkish stock market, PhD thesis, Department of Finance, Michigan State University, E. Lansing, MI.

Table 5.1: Summary statistics for ISE100 stock returns

Series	mean	med	min	max	variance	skewness	kurtosis
daily returns	-0.004	0.031	-13.288	13.040	2.281	-0.348	10.730
weekly returns	-0.017	0.059	-17.688	12.915	13.780	-0.261	5.143

Table 5.2: Estimated $ARMA(P, Q) - FIGARCH(p, \delta, q)$ Models for ISE 100 Index returns

	Daily Returns	Weekly Returns
μ	-0.005	0.0025
	(0.025)	(0.099)
Θ_1	0.131	.
	(0.021)	.
ω	0.173	0.319
	(0.040)	(0.135)
β	0.269	0.023
	(0.123)	(0.108)
δ	0.538	0.319
	(0.108)	(0.135)
T	3339	686
$\ln(L)$	-5808.093	-1830.700
Skewness	-0.227	-0.192
Kurtosis	5.337	4.004
$Q(10)$	27.432	23.217
$Q^2(10)$	12.490	6.490
$Q(20)$	36.683	35.799
$Q^2(20)$	21.720	15.119

Key: $\ln(L)$ is the value of the maximized Gaussian likelihood, and QMLE standard errors are presented in parentheses below corresponding parameter estimates. The $Q(10)$, $Q^2(10)$, $Q(20)$, and $Q^2(20)$ are the Ljung-Box test statistics with 10 and 20 degrees of freedom based on the standardized residuals, and squared standardized residuals respectively. The sample skewness and kurtosis are also based on the standardized residuals.

Table 5.3: GPH, CSS and local Whittle estimates of long memory parameter for the ISE100 stock squared returns and absolute returns

Ordinates	R_t^2		$\|R_t\|$	
m	Daily	Weekly	Daily	Weekly
$T^{0.5}$	0.226	0.154	0.365	0.180
	(2.685)	(1.227)	(4.336)	(1.435)
	[-9.191]	[-6.724]	[-7.540]	[-6.517]
$T^{0.6}$	0.183	0.324	0.334	0.287
	(3.289)	(3.576)	(5.979)	(3.164)
	[-14.636]	[-7.451]	[-11.938]	[-7.863]
$T^{0.7}$	0.133	0.220	0.266	0.265
	(3.573)	(3.368)	(7.157)	(4.044)
	[-23.347]	[-11.911]	[-19.762]	[-11.235]
$T^{0.8}$	0.192	0.194	0.268	0.216
	(7.759)	(4.107)	(10.856)	(4.572)
	[-32.725]	[-17.103]	[-29.629]	[-16.638]
d_{CSS}	0.258	0.209	0.250	0.202
	(0.0973)	(0.095)	(0.030)	(0.051)
$d_{Whitlle}$	0.246	0.287	0.479	0.537
	(0.050)	(0.121)	(0.049)	(0.114)

Key: m stands for the number of periodogram ordinates used in the GPH estimator. The values in parentheses are the t statistics for testing the null of $H_0 : d = 0$ versus $H_1 : d > 0$, and the values in square parentheses are the t statistics for testing the null of $H_0 : d = 1$ versus the alternative of $H_1 : d < 1$. The t statistics are computed by using the theoretical variance of $\pi^2/24m$. The d_{CSS} and $d_{Whittle}$ are the estimate of long memory parameter from CSS estimator, and local Whitlle estimator respectively. Values in the parentheses are the robust standard errors.

Figure 5.1: ISE National 100 Daily stock indices, index returns, absolute and squared returns

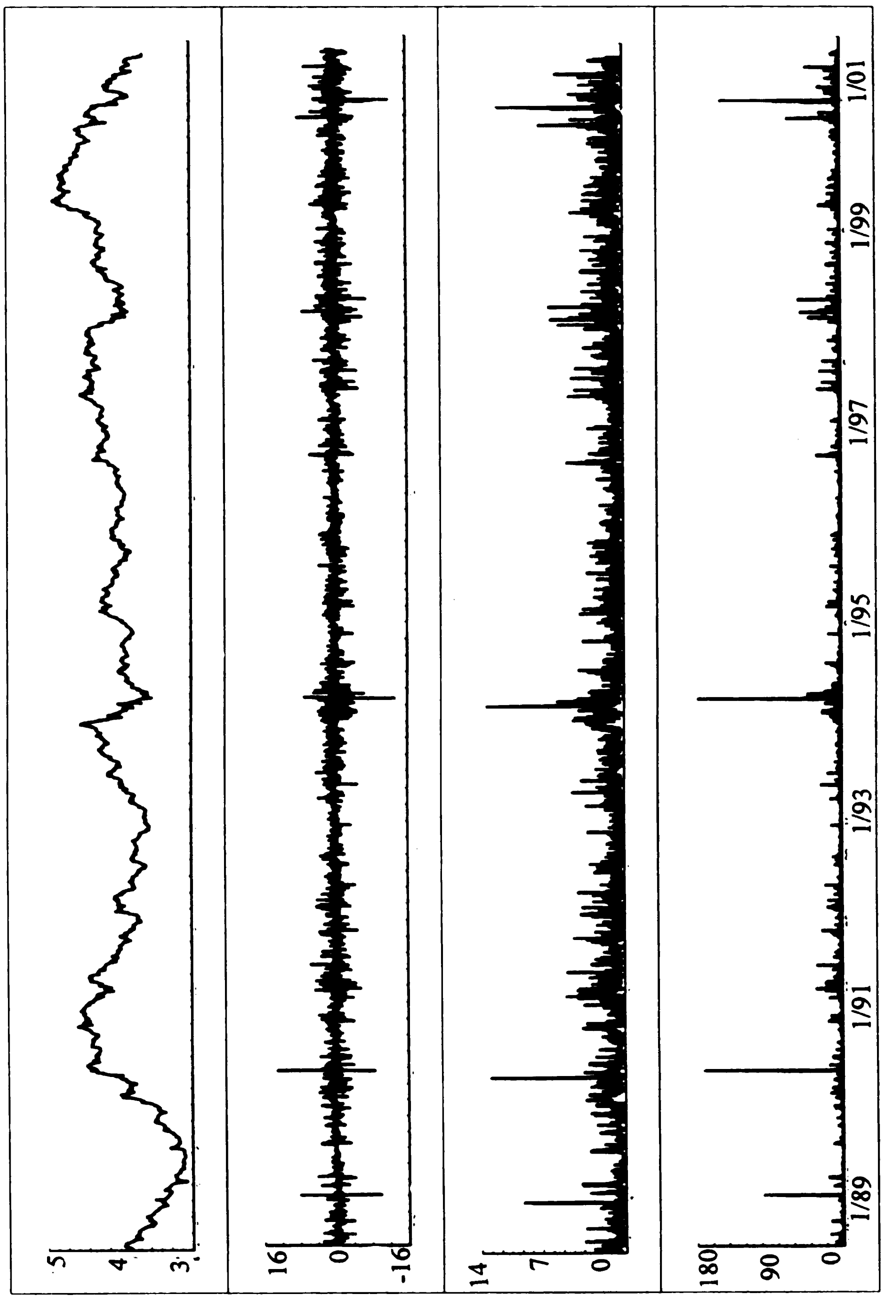
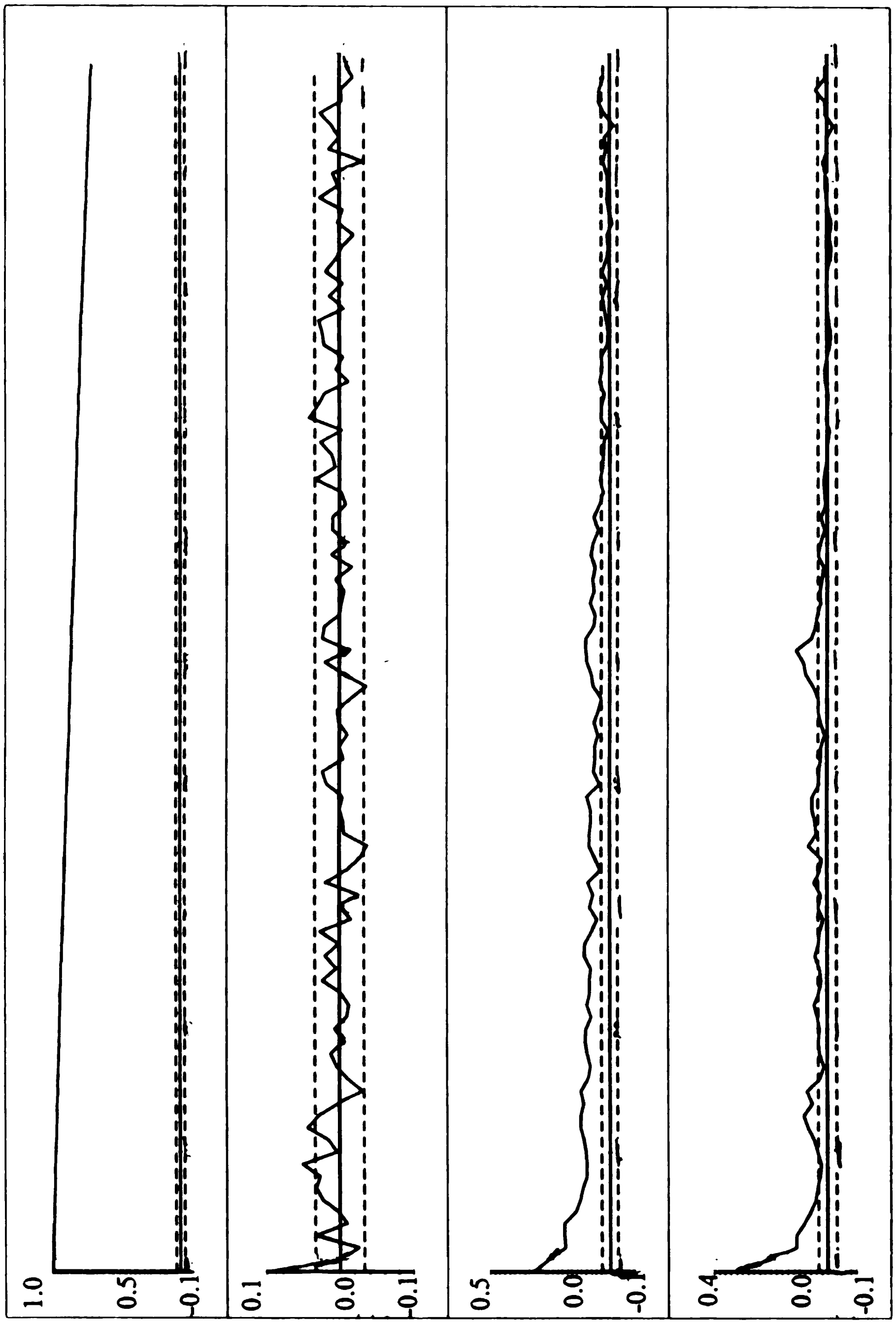


Figure 5.2: Correlograms of ISE 100 stock index returns



CHAPTER 6

Revisiting the nonlinearity and persistence in real exchange rates: evidence from a new unit root test and an ESTAR specification

6.1 Introduction

As discussed in chapter 3, there is a growing strand of research on nonlinear behavior of real exchange rates. The findings of chapter 3 and the discussion of the empirical and theoretical literature there indicated that in the presence of transaction costs real exchange rates are expected to adjust to equilibrium in a nonlinear fashion. It is also shown that the power of the standard unit root and stationarity tests is based on the parametric specification of the STAR model. When the parametric specification is one that indicates that the generated data has a unit root in the middle regime while the root(s) in the outer regime(s) becomes closer to unity, (hence the generated data is locally non stationary but globally remains stationary) the Augmented Dickey-Fuller (ADF) (Dickey and Fuller 1984) and the Phillips-Perron

(PP) (Phillips and Perron 1988) tests lack power in detecting the non-linear mean reversion. The formal testing of the conjecture that the real exchange rate can be mean reverting once the nonlinearity is controlled for remains a challenge for empirical researchers. As discussed in chapters 1 and 3, the linearity tests and the estimation of STAR models require the time series under consideration to be stationary. As the simulation experiments in chapter 3 indicated, if the true data generating process is a linear random walk, the linearity tests may spuriously indicate the presence of nonlinearity. This finding implies that the distribution of the linearity tests possibly differs for a non stationary process hence use of asymptotic χ^2 critical values may not be appropriate. This issue deserves further analysis which is beyond the scope of this chapter. To avoid this problem, the first difference of real exchange rates are used in chapter 3. This chapter, develops a unit root test that is specifically designed to test the random walk with or without drift against a globally mean reverting ESTAR process.

Some recent studies also considered the issues pertaining to stationarity and non-linearity within the context of STAR models and real exchange rates. Taylor et al. (2001) show empirically the stationarity of real exchange rates from multivariate tests before proceeding to their ESTAR model estimation. Killian and Taylor (2001) use simulations to assess the level of their test of random walk against an ESTAR alternative. These approaches are not totally satisfactory. Indeed, the Multivariate ADF (MADF) and the Johansen Likelihood Ratio (JLR) tests of Taylor and Sarno (1998) are not designed specifically to test unit root against mean reverting STAR alternatives. Taylor et al. (2001) show by simulation that these tests have better power properties compared to univariate ADF test when the true data generating process is a mean reverting ESTAR model. The MADF test assumes that all the series have a unit root under the null hypothesis hence the test has the tendency to reject the null when even only one of the series is stationary. This problem was also pointed out in Taylor and Sarno (1998). To avoid the pitfall of the MADF test, the JLR test assumes that at least one of the series has a unit root under the null hypothesis. The rejection of this null implies that all the series are stationary only if we assume that each of the series is a realization of an $I(0)$ or $I(1)$ process. Otherwise, the rejection of the null hypothesis in the JLR test will mean that at least one of the series is not a unit root process. Hence, it will not be informative about the other series. Moreover, the testing procedures in Taylor et al. (2001) departs from the original PPP criterion by calling for further economic information about the other real exchange rates in the testing step, but has the drawback that this additional information is left aside in the univariate estimation of ESTAR models for the real exchange rate. Killian and Taylor (2001) approach is relevant provided that the rejection of their null of the unit root guarantees the stationarity of their nonlinear ESTAR representation under the alternative, which in fact needs to be shown.

This chapter departs from chapter 3 in that it develops a unit root test, namely a sup Wald test, ($sup\,Wald$), that has power against nonlinear mean reversion. Two null hypotheses are considered; *random walk without drift* and *random walk with drift* against mean reverting ESTAR alternative. The distribution of the test statistics are derived and are conjectured to be nuisance parameter free. We apply the tests to G-7 countries' real exchange rates against the US dollar for the floating period. Findings from the new tests support the nonlinear mean reversion of real exchange rates. The empirical power and size of the tests are studied through simulations and are compared with those of the standard unit root tests. The simulations indicate that $sup\,Wald$ tests have good size and power properties and perform better than the standard unit root tests. This chapter also studies the dynamic adjustment mechanism of real exchange rates to a shock by utilizing generalized impulse response functions. The results from the estimated ESTAR models, the generalized impulse response functions and the distributions of generalized impulse responses in the outer regimes reveal the nonlinear and persistent behavior of the real exchange rates in this study.

The rest of the chapter is organized as follows; the next section discusses the foundations of nonlinear behavior of real exchange rates, and conditions for stationarity in the ESTAR model. Section 6.3 introduces the $sup\,Wald$ test and gives the asymptotic distribution of the tests. The empirical size and power of the tests are discussed in section 6.4. Section 6.5 gives and discusses the empirical findings. Section 6.6 concludes the chapter. The proofs of the propositions are given in the appendix to the chapter.

6.2 Foundations of nonlinear adjustment of real exchange rates and ESTAR model

6.2.1 Motivation for a nonlinear adjustment in real exchange rates

Similar to chapter 3 we chose to study the nonlinear dynamics in real exchange rates by using ESTAR model that is discussed in chapter 1. As discussed in chapter 3, the nonlinear behavior of real exchange rate may result from transaction costs. Dumas (1992), and Sercu et al. (1995) study a two-country model with trading costs. The models in these papers predict that the presence of trading costs leads to the existence of a region of no trade in which the real exchange rate may follow a random walk as arbitrage does not take place. Outside the region, international arbitrage takes place and brings the real exchange rate back to the nearest threshold level which corresponds to the marginal cost of shipping. As a result, the exchange rate is expected to behave discontinuously. Since in the real world, there are several goods and transaction costs differ for each good, it is intuitive to think that the shifts will be gradual rather than abrupt. Hence, a Smooth Transition Autoregressive model should better represent the shifts in the real exchange rates than the Threshold Autoregressive models (TAR).

The presence of transaction costs alone could not account for many of the observed very large movements in real exchange rates, either in terms of day-to-day volatility or in terms of periods of substantial and persistent overvaluation or undervaluation of real exchange rates. An example for this would be the overvaluation of the U.S. dollar in the 1980s. Killian and Taylor (2001) propose a complementary explanation that is based on the presence of heterogenous foreign exchange traders; noise traders and rational speculators (or arbitrageurs). Noise traders' demand for foreign exchange is affected by beliefs that are not fully justified by news about the fundamentals. Arbitragers on the other hand, form fully rational expectations about the return on holding foreign exchange and they sell foreign exchange when noise traders push prices up and buy when noise traders depress prices, thereby making a profit in the process. In this model, the unpredictability of noise traders' future opinions creates risk to arbitrageurs that prevents complete arbitrage. The arbitrage is limited by three types of risk; the future realizations of fundamental may turn out to be higher than expected, because of the unpredictable swings in the demand of noise traders a foreign exchange that is overpriced today may be even more overpriced tomorrow, and lastly the equilibrium value of the exchange rate can not be observed directly and hence arbitrageurs will have difficulty in detecting the deviations from fundamentals. Assuming that agents assign less probability to levels of exchange rate corresponding to large deviations from the fundamental level than the values close to the fundamental (this is because larger deviations are increasingly implausible from a theoretical point of view), few rational traders will be inclined to take a strong position when the exchange rate is close to the fundamental value. Therefore, closer to the unobserved equilibrium the exchange rate is driven mainly by noise traders. As the exchange rate moves away from the unobserved equilibrium, a consensus will gradually be reached among the rational traders that the exchange rate is misaligned, inducing them to take stronger positions against the prevailing exchange rate and ensuring the ultimate mean reversion of the exchange rate toward the unobserved true economic fundamental. As argued by Killian and Taylor (2001) this nonlinearity may be described by a STAR model, in which the strength of mean reversion is an increasing function of past deviations from the equilibrium.

Differently from chapter 3, we postulate an ESTAR model of the form for the real exchange rates;

$$q_t = \phi(L)\Delta q_t + [\mu + \rho q_{t-1}](1 - F(z_t;\gamma, c)) + [\mu^* + \rho^* q_{t-1}]F(z_t;\gamma, c) + u_t \qquad (6.1)$$

where $\phi(L) = \phi_1 L + \phi_2 L^2 + \cdots + \phi_{p-1}L^{p-1}$, $F(.)$ is the exponential transition function given in chapter 1 and 3, $z_t = q_{t-d}$ for $d \in 1, 2, \cdots, \bar{d}$. As discussed in chapter 3, the exponential form of the transition function makes good economic sense in this application because it implies symmetric adjustment of the real exchange rate above and below equilibrium (or positive and negative deviations from PPP). The transition parameter γ determines the speed of transition between the two extreme regimes, with lower values of γ implying slower transition. The middle regime corresponds to $q_{t-d} = c$, when $F = 0$ and (6.1) becomes a linear model;

$$q_t = \phi(L)\Delta q_t + \ \mu + \rho q_{t-1} + u_t.$$

The outer regime corresponds, for a given γ, to $\lim_{[q_{t-d}-c]\to\pm\infty} F(q_{t-d};\gamma, c)$, where (6.1) becomes a different AR(p) model;

$$q_t = \phi(L)\Delta q_t + \mu * + \rho * q_{t-1} + u_t,$$

with a correspondingly different speed of mean reversion so long as $\rho \neq \rho*$. In any empirical application of STAR models, it is necessary to determine the dimension $\bar{d}$ and the number of lagged values of the real exchange rate influencing the transition function, that is, the delay parameter d. In general, applied practice with ESTAR models has favored restricting $\bar{d}$ to be a singleton (see e.g. Teräsvirta, 1994; Taylor, Peel and Sarno, 2001; and Killian and Taylor, 2001). Granger and Teräsvirta (1993) and Teräsvirta (1994) suggest a series of nested tests for determining the appropriate delay parameter. In the present application to monthly real exchange rate data, similar to Taylor, Peel, and Sarno (2001), we found that the model that worked best for each country (in terms of goodness of fit, statistical significance of parameters, and adequate diagnostics) set the delay parameter to 1. The finding of the delay parameter being 1 seems reasonably intuitive since it allows the effects of deviations from equilibrium to affect the nonlinear dynamics with a shorter lag rather than larger lags. This is because, there is no compelling reason why there should be very long lags before the real exchange rate begins to adjust in response to a shock.

6.2.2 Stationarity of ESTAR model

Since, this chapter aims to test the random walk against a stationary ESTAR alternative, we need to determine under which conditions the ESTAR model given in (6.1) is a globally stationary process. For this end, consider the $ESTAR(p)$ model given in the following equation.

$$y_t = \pi' x_t(1 - F(z_t; \gamma, c)) + \pi^{*\prime} x_t F(z_t; \gamma, c) + u_t \tag{6.2}$$

where $x_t = (1, y_{t-1}, \cdots, y_{t-p})'$, $F(z_t; \gamma, c) = 1 - \exp(-\gamma(z_t - c)^2)$, $z_t = y_{t-d}$ for $d = 1, 2, \cdots, p_{pmax}$. As for the disturbances, we have the following assumption.

Assumption 1: *Assume that $u_t \sim iid$, with $E(u_t) = 0$, $E|u_t| < \infty$ and independent of y_0. The distribution of u_t is absolutely continuous and its density is positive everywhere.*

Note that Assumption 1 is satisfied for $u_t \sim iid(0, \sigma^2)$. As discussed in Tøstheim (1990) the stationarity properties of the ESTAR model given in (6.2) are dictated by what happens in the limit when z_t goes to infinity. As z_t goes to infinity (both positive and negative infinity) $F(\pm\infty;\gamma,c)$ converges to 1. Therefore, as z_t goes to infinity, y_t becomes a two-regime self exciting threshold model;

$$y_t = \pi' x_t(1 - I(|z_t|> \ c)) + \pi^{*\prime} x_t I(|z_t|> \ c) + u_t \tag{6.3}$$

The stationarity properties of general threshold models are not known. Chan et al. (1985) give necessary and sufficient conditions for a multiple regime $TAR(1)$ model with d=1. At an intuitive level, we can expect that the process for y_t given by (6.2) be globally stationary when the roots of the autoregressive polynomial in the outer regime lie outside the unit circle. In other words, the largest root in absolute value of the characteristic polynomial in the outer regime, $1 - \pi_1^*\xi - \pi_2^*\xi^2 - \cdots - \pi_p^*\xi^p = 0$ be less than 1. This means that the smallest root in the middle regime, $1 - \pi_1\xi - \pi_2\xi^2 - \cdots - \pi_p\xi^p = 0$ may be equal to one (having a unit root in the inner regime) while the process stays globally stationary.

In order to gain some insight into the stationarity of the data generated from an ESTAR process with parameter specification that satisfy the conditions stated in the last paragraph, a simulation experiment is conducted. The data, y_t, for $t = 1, \cdots, T$ from the ESTAR model, $y_t = \rho y_{t-1}(1 - F(y_{t-1}\gamma, c)) + \rho * y_{t-1}F(y_{t-1}, \gamma, c)) + u_t$, with $\rho = 1$, $\rho{*}= \ 0.8$, $\gamma = 3, 5, 10, 20$, and $u_t \sim iidN(0,1)$ are generated. The threshold parameter, c is kept at 0. The data is generated N=10,000 times and in each replication, first 100 simulated data points are discarded. The sample sizes of $T = 300$, 500, 1000 are used. Letting $y_{t,i}$ be the value of y_t in simulation replication i for $t = 1, \cdots, T$; and $i = 1, \cdots, N$. The j-step ahead covariance across replications, $\hat{\delta}_{t,j} = \frac{1}{N}\sum_{i=1}^{N} y_{t,i}y_{t-j,i}$, for $t = j+1, \cdots, T$ and $j = 1, 2, 3, \cdots, J = 10$, are estimated and graphed against time t for each j. The purpose of this simulation is to see whether $\hat{\delta}_{t,j}$ does or does not depend on t. For a covariance stationary process we should expect that $\hat{\delta}_{t,j}$ stay approximately constant, over time t. Since the estimated

$\hat{\delta}_{t,j}$s for any given j do not differ across the different specifications of γ and sample size T, the results from $\gamma = 10$ for $j = 2, 5, 7, 9$ and $T = 1000$ are given in panels of figure (6.1). As it can be see from the graphs, $\hat{\delta}_{t,j}$s stay almost constant over time for any given j. This indicates that the data generated from ESTAR model has on average covariances that do not depend on time, implying covariance stationarity.

6.3 Testing Unit root against stationary ESTAR alternatives

Following Micheal et al. (1997) we can re-write the ESTAR model given in (1.1) as follows;

$$y_t = \phi(L)\Delta y_t + [\mu + \rho y_{t-1}](1 - F(z_t; \gamma, c)) + [\mu^* + \rho^* y_{t-1}]F(z_t; \gamma, c) + u_t, \qquad (6.4)$$

where $\phi(L) = \phi_1 L + \phi_2 L^2 + \cdots + \phi_{p-1}L^{p-1}$. We can re-parameterize the transition function by first letting $\lambda = \sqrt{\gamma}c$. This parameterization will be useful in proving the asymptotic behavior of the unit root tests. Note that we can write $F(.)$ as $F(z_t; \lambda, c) = 1 - \exp\left(-(\frac{\lambda}{c}z_t - \lambda)^2\right)$. In model (6.4) we can test $H_0^a : \mu = \mu* = 0 \,\text{and}\, \rho = \rho* = 1$, *random walk without drift*, and $H_0^b : \mu = \mu* \,\text{and}\, \rho = \rho* = 1$, *random walk with drift* against the alternative $H_1 : y_t$ follows a stationary ESTAR process. Under the null hypotheses we assume that the roots of $1 - \alpha_1\xi - \alpha_2\xi^2 - \cdots - \alpha_p\xi^{p-1} = 0$, where $\alpha_1 = (1 + \phi)$, $\alpha_i = \phi_i$ for i odd and $\alpha_i = \phi_i - \phi_{i-1}$ for i even, lie outside the unit circle. Under both null hypotheses the parameters λ and c are not identified. Thus it is impossible to obtain consistent estimates of λ and c under both null hypotheses. The proposed unit root test is the Wald test which test the parameter restrictions given in the above null hypotheses. The unrestricted model is given by equation (6.4). The restricted model is given by

$$y_t = \phi(L)\Delta y_t + y_{t-1} + u_t, \qquad (6.5)$$

$$y_t = \phi(L)\Delta y_t + \mu + y_{t-1} + u_t$$

under H_0^a and H_0^b respectively.

As noted by Leybourne et al. (1998) the ESTAR model given in (6.4) is linear in autoregressive parameters for given λ and c. Hence, for given λ and c we can estimate the unrestricted and restricted models by OLS. Denoting the vector of residuals from the unrestricted model by $\hat{u}$ and the vector of residuals from the restricted model by $\tilde{u}$, we can write the Wald test in terms of the residual sum of squares under homoscedasticity as;

$$W_T(\lambda, c) = T\frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}} \tag{6.6}$$

Proposition 1: *Let $d = \bar{d} = 1$ be fixed. Let $\lambda > 0$ and $\tilde{c} = c/\sqrt{T} > 0$ be fixed. Suppose $(\lambda, \tilde{c})$ belongs to Λ where Λ is a compact set of R^{+2}. Under H_0^a, the Wald test satisfies*

$$W_T(\lambda, c)_{T\to\infty} \xrightarrow{L} \zeta(\varphi) \tag{6.7}$$

poinwise in (λ, c), where $\varphi = (\lambda, \tilde{c}, \delta)$, $\delta = \sigma/(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_{p-1})$ and $\zeta(\varphi)$ is a function of Brownian motions given in the proof of the proposition. Under the alternative the statistic diverges.

Since, under the null hypothesis γ and c are not identified we can make any assumptions about them. The assumption $c = \sqrt{T}\tilde{c}$ is reminiscent of the assumption made in the structural change literature where the break point is hypothesized to be equal to τT where τ is in $(0, 1)$. Under H_0^a, $y_t/\sqrt{T}$ converges to a Brownian motion $\delta B(r)$ with $r = t/T$. Note that since $z_t = y_{t-d}$ the the behavior of the transition function in the limit will be characterized by the behavior of y_t as T goes to infinity. If we assume that γ and c are fixed, then the transition function,

$$F(z_t; \gamma, c) = 1 - \exp\left(-(\sqrt{\gamma}z_t - c\sqrt{\gamma})^2\right)$$

$$= 1 - \exp\left(-T\left(\sqrt{\gamma}\frac{z_t}{\sqrt{T}} - \frac{c\sqrt{\gamma}}{\sqrt{T}}\right)^2\right)$$
$$\longrightarrow 1$$

as $T \to \infty$. This means that for fixed γ and c the process becomes linear asymptotically and hence the test statistic will lose its power in detecting nonlinear stationarity of the time series under consideration. On the other hand if we assume that $(\lambda, \tilde{c})$ are fixed, then we have;

$$F(z_t;\gamma,c) = 1 - \exp\left[-\left(\frac{\sqrt{T}\lambda}{c}\frac{z_t}{\sqrt{T}} - \lambda\right)^2\right] \xrightarrow{L} 1 - \exp\left[-\left(\frac{\lambda}{\tilde{c}}\delta B(r) - \lambda\right)^2\right] \text{as}\, T \to \infty.$$

The following proposition gives the distribution of the Wald test under the null hypothesis of H_0^b. As noted in Hamilton (1994) the distribution of ADF and PP tests differ under "random walk without drift" and under "random walk with drift". In a similar fashion, proposition 2 shows that the distribution of the Wald test is different from the distribution one obtains under H_0^a.

Proposition 2: *Let $d = \bar{d} = 1$, and $\tilde{c} = \frac{c}{T}$, and λ be fixed. Suppose $(\lambda, \tilde{c})$ belongs to Λ, is a compact set of R^{+2}. Under the null hypothesis H_0^b the asymptotic distribution of Wald test given in equation (6.7) is a $\chi^2(\varphi)$ variate with φ is given in the proof of the proposition. Under the alternative the statistic diverges.*

Note that under H_0^b, when (γ, c) are fixed,

$$F(z_t;\gamma,c) = 1 - \exp\left(-\gamma T^2\left(\frac{z_t}{T} - \frac{c}{T}\right)^2\right) \xrightarrow{L} 1 \text{ as } T \to \infty.$$

When we assume that $(\lambda, \tilde{c})$ are fixed, then

$$F(z_t;\gamma,c) = 1 - \exp\left(-\left(\frac{\lambda}{c}T\frac{z_t}{T} - \lambda\right)^2\right) = 1 - \exp\left(-\left(\frac{\lambda}{\tilde{c}T}T\frac{z_t}{T} - \lambda\right)^2\right)$$
$$\xrightarrow{L} 1 - \exp\left(-\left(\frac{\lambda}{\tilde{c}}\mu - \lambda\right)^2\right) \text{as}\, T \to \infty.$$

The proofs of propositions 1 and 2 are given in the appendix.

Note that the limiting distribution of the Wald test under both null hypotheses depends on the unknown parameters (λ, c). As these parameters are not identified under the null hypotheses, the choice of (λ, c) is arbitrary. Hence the limiting distribution of the test statistic is not nuisance parameter free. One way to get away from this problem and gain power is to use the same testing strategy as in testing linearity against self exciting threshold autoregressive model (SETAR) (see for instance Hansen (1997, and Caner and Hansen 2001)), namely taking the supremum of the test statistic with respect to the nuisance parameters. The sup Wald test then will be given by:

$$supW \equiv sup_{(\lambda,c)\in\Omega\times C} W_T(\lambda, c), \tag{6.8}$$

where $\Omega = \left[\underline{\ell}, \overline{\ell}\right]$ and $C = [\underline{c}, \overline{c}]$ are such that $0 < \underline{\ell} < \lambda < \overline{\ell}$, and $0 < \underline{c} < \frac{\bar{c}}{\gamma} < \overline{c}$. Since the test will have power for any λ, any fixed Ω can be chosen. Obviously the test will have power even if we choose one single value for λ, but the use of a range of values will increase the power of the test. One important issue is not to make the interval too wide as a very large λ may make the transition function F to be flat. As for the choice of C, we can follow the same approach taken in the SETAR literature (see for instance Hansen 1997,and Caner and Hansen 2001) and select the c corresponding to the ordered values of $|z_t|$ and discard 15% of the highest and smallest values. This will guarantee that the boundaries $\underline{c}$ and $\overline{c}$ do not depend on any unknown parameter. We conjecture that the distribution of *sup Wald* tests will be nonstandard in the sense that it is going to be the supremum of a number of random functions, but nuisance parameter free. Unfortunately, for a rigorous proof of this conjecture, we need a uniform convergence in $\Lambda = \Omega \times C$ which we haven't been able to prove. To our best knowledge, there is no result in the econometrics literature that we can use to prove our conjecture. If we had a uniform convergence of Wald tests discussed above the proof of our conjecture for the *sup Wald* tests would be trivial in the sense that our conjecture would follow by continuous mapping theorem. In the rest of the chapter we assume that our conjecture is true and following Caner and Hansen (2001) we compute critical values by simulation.

6.4 Empirical Critical Values and size and power properties of the *supWald* tests

To compute the empirical critical values we have generated data from (6.5). When fitting (6.5) to real exchange rates μ was found to be statistically indistinguishable from zero for most of the real exchange rates and it was around 0.05 for some of the rates. Hence data is generated with $\mu = 0$ and with $\mu = 0.05$ in computing. the critical values. In generating the data, disturbances, u_t in (6.5), are drawn from $iid\,\aleph\,(0,1)$. Table 6.1, reports the empirical critical values from 20,000 replications of sample size 312 since 312 corresponds to the sample size in this study. The two dimensional grid search in γ and c was performed for the following sets of values: $\gamma \in (0.25, 0.5, 0.75, 1, 1.25, \cdots, 15)$ and $c \in [\underline{c}\,\bar{c}]$ with $\underline{c}$ and $\bar{c}$ such that 15% of the smallest and highest values of $|y_{t-1}|$ are excluded from the grid. In addition to the standard version, heteroscedasticity- robust versions of the tests are also computed.

In order to analyze the size and power properties of the proposed tests, a finite sample study is performed. The empirical critical values reported in table 6.1 are used in the simulation experiments. Therefore, the power is actually a size-corrected power. In computing the size of the tests, the data is generated under the null hypotheses of H_0^a and H_0^b with $\mu = 0$ and $\mu = 0.05$. The disturbances are drawn from $i.i.d.\aleph(0,1)$. The standard error is normalized to unity in all of the experiments. Table (6.2) reports the empirical rejection frequencies from 5,000 Monte Carlo replications with $T = 312$. For comparison purposes, the empirical size of the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) statistics are also reported. The empirical size of the *supWald* test is quite accurate and comparable with the size of ADF and

PP. The heteroscedasticity robust versions, $supWh$ and $supWh_{\mu}$, seem to be slightly more conservative than the standard versions.

The power of the tests is examined by generating 5,000 series under the alternative (6.1) for various parameters values. Throughout the experiments, ρ is kept fixed at unity, while the autoregressive parameter in the outer regime, $\rho*$, was varied to see the effect of having an autoregressive root in the outer regime that changes from values in the stationary range to values closer to unity. This parameterization is consistent with the fitting of ESTAR models to the data as we will see in the next section. The data is generated under $\mu = \mu* = 0$ and $\mu \neq \mu*$. Since the results did not vary significantly, only $\mu = \mu* = 0$ and $\mu = 0.05$, $\mu* = -0.05$ are reported in table 6.3. The smoothness parameter, γ, was varied to see the influence of the change in the curvature of the transition function on the power of the tests. The values reported are closer to the smoothness parameter estimates obtained in the empirical section. Since, it did not have any significant effect on the power of the tests, the threshold parameter, c, is set at 0.05. Again for comparison purposes, the power of ADF and PP tests are also reported. As can be observed from the table, as the autoregressive parameter, $\rho*$, in the outer regime approaches unity, the power of all tests declines. However, the fall in the power of ADF and PP is more than that of the $supWald$ tests. For instance, the power of ADF and PP tests is about 40 percent, while that of $supW_{\mu}$ is about 83 percent in the case given in panel d of the table. In cases where $\rho* = 0.95$ the $supWald$ tests outperform the ADF and PP tests. Moreover, the power of $supWald$ tests in general increases with γ.

6.5 Empirical Results

6.5.1 The data

The data set comprises monthly observations on consumer price indices for the US, the UK, Canada, Germany, Italy, Japan, and Switzerland, and end-of-period spot exchange rates for the UK pound (BP), German mark (GM), Canadian dollar (CD), Italian lira (IL), Japanese yen (JY), and Swiss franc (SF) against the US dollar. The data covers the sample period from 1973:01 to 1998:12, and is taken from the International Monetary Fund's (IMF) *International Financial Statistics* data compact discs. Real exchange rate series are constructed with these data in logarithmic form as in chapter 3. The data is centered around sample mean.

6.5.2 Unit root test results

Table (6.4) gives the results from standard unit root tests, namely ADF (Dickey and Fuller, 1981) , and PP (Phillips and Perron, 1988), stationarity test of Kwiatkowski Phillips, Schmidt, and Shin (1992) (KPSS) together with the results of $sup Wald$ tests applied to real exchange rates. The PP and ADF tests reject the unit root null for only BP and IL only at 10 percent level. For all other series, ADF and PP tests indicate the presence of a unit root at the 10 percent significance level. ADF and PP fails to reject the null hypothesis of a unit root for all of the real exchange rates at the 5 percent level. KPSS rejects the null of stationarity in all real exchange rates.

Since we have seen that ADF and PP tests lose power when the autoregressive parameter in the outer regime becomes closer to unity, we can argue that these results can not constitute a strong evidence for non-stationarity of real exchange rates. According to the $sup Wald$ tests reported in table 6.4 the random walk hypothesis is rejected strongly for all of the real exchange rates in favor of a globally stationary ESTAR model. Note that except IL and JY for none of the real exchange rates in our sample we were able to obtain constant term estimates in the fitted ESTAR model. Therefore, we did not test the null hypothesis of random walk with a drift, (H_0^b). For the JY and the IL $sup\,Wald$ tests reject the null of random walk with drift at the 5 and 10 percent levels, respectively. Given the results from the $sup\,Wald$ tests we can argue that real exchange rates in our sample are globally stationary, although they may exhibit random walk behavior locally. This result indicates that once a threshold type of nonlinearity is taken into consideration, real exchange rates are stationary. After empirically showing that real exchange rates are stationary, the next task is to model the nonlinear behavior of real exchange rate under the alternative of a globally stationary ESTAR model.

6.5.3 ESTAR model estimation and persistence of real exchange rates

While the results of $sup\,Wald$ tests impart some idea of the mean reverting nature of real exchange rates, a sensible way to gain a full insight into the mean-reverting properties of real exchange rates is to model this behavior by the nonlinear model that is assumed under the alternative hypothesis, and also to look at the propagation mechanism with which the adjustment process takes place after a shock to the level of real exchange rates. Thus, table 6.5 reports the estimated ESTAR models of the form given in (6.1). The estimation of the ESTAR model given in (6.1) was performed using the constrained maximum likelihood method. The CML library in Gauss with the Newton-Raphson optimization algorithm is used in estimation. The constraints, $\gamma > 0$ and $c \in [\underline{c}, \bar{c}]$, with $\underline{c}$ and $\bar{c}$ such that 15% of the observations in absolute value are below $\underline{c}$ and 15% are above $\bar{c}$, are imposed. Following, Leyboune et al. (1998) the objective function is concentrated so that optimization is carried out for γ and c only. For details, see Leyboune et al. (1998) or chapter 1 of this dissertation. The starting values are obtained from a two-dimensional grid search over γ and c. Following the suggestion of Teräsvirta (1998), the transition function is re-parameterized as follows:

$$F(z_t;\gamma,c) = 1 - \exp\left(\frac{\gamma}{s.e.(z_t)}(z_t - c)^2\right),$$

where $s.e.(z_t)$ is the sample standard deviation of the transition variable, so as to make γ approximately scale-free. The grid for γ was set arbitrarily to $0.1, 0.2, \cdots, 20$, while the grid for c is set as explained above.

For each of the estimated ESTAR models, we could not reject the hypothesis of no remaining nonlinearity of ESTAR form for values of d ranging from 2 to 12 on the basis of the $p-$values of Lagrange multiplier (LM) tests (table 6.5 reports only the p-values corresponding to the maximal value of the LM statistic, $pNLES_{max}$). Neither could we reject the hypothesis of remaining nonlinearity of LSTAR variety with values of delay parameter in the range of 1 to 12 ($pNLLS_{max}$ in the table). This procedure suggests setting $d = 1$. The residual diagnostic statistics are satisfactory in all cases (Eithrehim and Teräsvirta, 1996). The estimated transition parameter in each case appears to be strongly significantly different from zero both on the basis of the individual $t-ratios$ as well as in terms of the empirical marginal significance levels reported in the square brackets. Since under the null hypothesis that $\gamma = 0$, each of the real exchange rate series follow a unit root process, the usual $t - ratios$ should be interpreted with caution. In the presence of a unit root under the null hypothesis we can not assume that the distribution of $t - ratio$ will be given by student's t distribution. Following Taylor, Peel, and Sarno (2001), the empirical $p-$values are computed by Monte Carlo methods assuming that the true data generating process for the logarithm of the real exchange rate series was a random walk with the parameters of the data generating process calibrated using the actual real exchange rate over the sample period. The empirical $p-$ values are based on 5,000 simulations of length 412, initialized at 0, from which the first 100 data points were discarded in each case.

At each replication ESTAR of the form reported in table (6.5) was estimated. The percentage of replications for which a $t-ratio$ for the estimated transition parameters was greater in absolute value than that reported in table (6.5) was obtained was then reported as the empirical $p-$value in each case. Note that since this test can also be considered to be a unit root test against a nonlinear mean reverting alternative, the results also support the findings from $sup\,Wald$ tests reported in the previous section. As can be seen from panels of figure 6.1, the estimated models fit the data very well and real exchange rate visit both inner and outer regimes in each case. The graph of the transition function against time reveals that BP, DG, GM, and SF (European zone except IL) series tend to stay closer to the outer regime until 1985 and stay closer to inner the regime between 1986 and 1993 and then again tend to stay closer to the outer regime after the early 1990s. On the other hand, CD, IL, and JY tend to stay closer to the outer regime for most of the time during our sample period.

The ESTAR estimates reported in table 6.5 indicate that the autoregressive parameter in the inner regime is, for all series, either unity or above unity, implying a unit root behavior in the inner regime. This is consistent with the theoretical foundations given above in the sense that whenever the deviation from the equilibrium is small real exchange rates behave as a random walk. On the other hand, the autoregressive estimate for the outer regime is, although less than unity for all series, close to unity, implying near unit root behavior in the real exchange rates even globally. This finding is consistent with the findings of chapter 3 in that it implies that deviations from equilibrium should persist for a long time. This finding also motivates the need to evaluate estimated models on the basis of impulse response functions as the estimated parameters indicate that the real exchange rates may reveal persistent deviations from equilibrium. To this end, the panels of figure 6.2 give the estimated generalized impulse response functions (GIRF). The GIRFs are calculated as in chapter 3. For a linear univariate model, the impulse response function is equivalent to a plot of the coefficients of the moving average representation (see e.g. Hamilton, 1994, p. 318). As discussed in chapter 1 estimating the impulse response function for a nonlinear model raises special problems both of interpretation and of computation, (see also, Koop, Peseran, and Potter, 1996). In particular, with nonlinear models, the shape of the impulse response function is not independent with respect to either the history of the time series at the moment the shock occurs, the size of the shock considered, or the distribution of future exogenous innovations. In this sense, impulse response functions are themselves random variables. As discussed in chapter 1, the distribution of impulse responses can be utilized to gain insight about the persistence of shocks in STAR models. It is intuitive to think that if a time series process is stationary and ergodic, the effects of all shocks eventually converge to zero for all possible histories of the process. Hence the distribution of impulse responses collapses to a spike at 0 as the horizon approaches to infinity. In contrast, for non-stationary time series the dispersion of the distribution of impulse responses is positive for all horizons. Koop Peseran and Potter (1996) suggest use of dispersion of the distribution of generalized impulse responses at the finite horizons as a tool in obtaining information about the persistence of shocks.

In this chapter we compute history- and shock-specific generalized impulse responses for all observations in the sample period as discussed in chapters 1 and 3. The values of the normalized initial shock equal to $\iota/\hat{\sigma}_u = 1, 5, 10, 20, 40$, where $\hat{\sigma}_u$ denotes the estimated standard deviation of the residuals from the ESTAR model. For each combination of history and initial shock, we compute generalized impulse responses for horizons $k = 1, 2, \cdots, N$ with $N = 120$. The conditional expectation in (1.42) are estimated as the means over 5,000 realizations of q_{t+k} with and without using the selected initial shock to obtain q_t and using randomly sampled residuals of the estimated ESTAR models elsewhere. All generalized impulse responses are initialized such that they equal $\iota/\hat{\sigma}_u$ at $k = 0$.

The estimated generalized impulse responses that correspond to the histories associated with the average value of the transition function, are graphed in the panels of figure 6.2 for each of the real exchange rates. These impulse response functions very clearly illustrate the nonlinear nature of the adjustment, with the impulse response functions for larger shocks decaying much faster than those for smaller shocks. Careful analysis of the panels of figure 6.2 indicate that shocks to the level of real exchange rates are although decays for all shocks, in all cases the speed with which the impulse responses decays and becomes half of the original normalized value of the initial shock changes with the magnitude of the initial shock. For even moderate size shocks it takes several months for the shocks to revert back to half of the initial magnitude. Since, impulse response functions are random variables that depend on the shock and the initial history of the series considered, the distribution of impulse responses for those histories corresponding to the value of the transition function being in the upper 95 quartile are given in the panels of figure 6.3. Note that these impulse responses correspond practically to periods where the real exchange rate is in the outer regime. Therefore we expect that the real exchange rate to be mean reverting and hence the distribution of generalized impulse responses accumulate around zero at finite horizons. The panels of figure 6.3 illustrate clearly that as the horizon increases the distribution of generalized impulse responses tend to pile up around zero. However, in none of the cases, the distribution of generalized impulse responses do not form a spike around zero even for horizons of 120 months which correspond to 10 years after an initial shock occurs. These results support the findings in chapter 3 and lead us to reach a similar conclusion in that despite the evidence of mean reverting nonlinearity in real exchange rates, they are very persistent in terms of the response to shocks.

6.6 Conclusion

The high persistence of the deviations from PPP is well documented in the literature. This chapter explored the nonlinear mean reversion of deviations from PPP within the context of an exponential smooth transition autoregressive model. The chapter proposes $sup\,Wald$ tests to test the random walk hypothesis against globally stationary ESTAR alternatives. Results from standard unit root tests and the KPSS test indicate non-stationarity of real exchange rates while results from $sup\,Wald$ test revealed stationarity of real exchange rates once nonlinearities are controlled for. The Monte Carlo experiments on the power of $sup\,Wald$ and standard unit root tests indicated that for parametric specifications that are closer to the fitted ESTAR models in the data, $sup\,Wald$ tests have better power properties than the standard unit root tests. Estimation, and further analysis of real exchange rates by generalized impulse response functions, indicated the nonlinearity and persistence of deviations from the PPP. Although, the larger deviations tend to decay more rapidly, the half-life estimates seem to be consistent with the studies that do not take nonlinearity into consideration, see for instance, Rogoff (1996).

6.7 Appendix: Proof of propositions 1 and 2

For the sake of completeness, in the following we first re-produce the definition of a *regular* transformation and the theorem 3.1 of Park and Phillips (1999).

Definition 6.1: (Definition 3.1 of Park and Phillips, 1999) *A transformation T is said to be regular if and only if,*

(a)it is continuous in a neighborhood of infinity, and

(b) on every compact set Π, there exist $\underline{T}_\epsilon$, $\bar{T}_\epsilon$ and $\delta_\epsilon > 0$ for each $\epsilon > 0$ satisfying

$$\underline{T}_\epsilon(x) \leq T(y) \leq \bar{T}_\epsilon(x)$$

for all x, $y \in C$ such that $|x-y| < \delta_\epsilon$, and $\int_\Pi \left(\bar{T}_\epsilon - \underline{T}_\epsilon\right)(x)dx \rightarrow 0$, as $\epsilon \rightarrow 0$.

According to Park and Phillips (1999) the class of regular transformations includes all continuous functions on a compact support. For that reason, the exponential function is a regular function for any given value of λ and c. Since in the proofs we assume that the parameter space for $(\lambda, \tilde{c})$ is compact the exponential function indexed by the parameters (λ, c) satisfies the regularity conditions given in definition 3.2 of Park and Phillips. Moreover, since any regular transformation is closed under addition, subtraction, and multiplication the transformations obtained by addition, subtraction and multiplication of the exponential function is regular. For details, see Park and Phillips (1999) pages 8-10.

Definition 6.2 (Definition 3.1 of Park and Phillips 1999) *We say that for the function $T(x, \omega)$ (defined on a compact set of parameter space, Π) is regular if*

(a) T is regular for all $\pi \in \Pi$

(b) for all $x \in R$, $T(x, .)$ is equicontinuous in a neighborhood of x.

Since the exponential function is continuous for all x and (γ, c) it should satisfy the regularity conditions stated above.

Theorem: (Theorem 3.1 of Park and Phillips, 1999) *Under certain regularity conditions on the disturbances of the time series process given in (6.2) (u_t being* *a Martingale difference sequence is enough) and under a regular transformation T on a compact set Π*

$$\frac{1}{n}\sum_{t=1}^{n} T\left(\frac{y_t}{\sqrt{n}}, \pi\right) \rightarrow_{a.s.} \int_0^1 T\left(B(r), \pi\right) dr,$$

uniformly in $\pi \in \Pi$. Moreover, if $T(.,\pi)$ is regular, then $\frac{1}{n}\sum_{t=1}^{n} T\left(\frac{y_t}{\sqrt{n}}, \pi\right) u_t \rightarrow_d \int_0^1 T\left(B(r), \pi\right) dB(r)$ as $n \rightarrow \infty$.

The proofs of propositions use these results frequently.

Proof of Proposition 1: The proof of the proposition follows the similar steps given in Hamilton (1994, chapter 17) and uses theorem 3.1 of Park and Phillips (1999). Letting $v_t = y_t - y_{t-1}$, the model in (6.2) can be written as

$$y_t = x_t'\beta + u_t \tag{6.9}$$

where

$$x_t = (v_{t-1}, \cdots, v_{t-p+1}, (1 - F_t), y_{t-1}(1 - F_t), F_t, y_{t-1}F_t)',$$

$$\beta = (\phi_1, \cdots, \phi_{p-1}, \mu, \rho, \mu*, \rho*)',$$

$u_t \sim iid(0, \sigma_u^2)$ and for notational simplicity the dependence on t of transition function is denoted by F_t. Note that x_t depends on λ and $\tilde{c}$ which we have assumed to be fixed. Given the representation in (6.9), the deviation of OLS estimates ($\hat{\beta}$) from the true value (β) is

$$\hat{\beta} - \beta = \left[\sum x_t x_t'\right]^{-1} \sum x_t u_t \tag{6.10}$$

These can be written as follows:

$$\sum x_t x_t' = \begin{bmatrix} A_{11} & A_{21}' \\ A_{21} & A_{22} \end{bmatrix} \tag{6.11}$$

where;

$$A_{11} = \begin{bmatrix} \sum v_{t-1}^2 & \sum v_{t-1}v_{t-2} & \cdots & \sum v_{t-1}v_{t-p+1} \\ \sum v_{t-2}v_{t-1} & \sum v_{t_2}^2 & \cdots & \sum v_{t-2}v_{t-p+1} \\ \vdots & \vdots & \cdots & \vdots \\ \sum v_{t-p+1}v_{t-1} & \sum v_{t-p+1}v_{t-2} & \cdots & \sum v_{t-p+1}^2 \end{bmatrix}$$

$$A_{21} = \begin{bmatrix} \sum (1-F_t)v_{t-1} & \cdots & \sum (1-F_t)v_{t-p+1} \\ \sum y_{t-1}(1-F_t)v_{t-1} & \cdots & \sum y_{t-1}(1-F_t)v_{t-p+1} \\ \sum F_t v_{t-1} & \cdots & \sum F_t v_{t-p+1} \\ \sum y_{t-1}F_t v_{t-1} & \cdots & \sum y_{t-1}F_t v_{t-p+1} \end{bmatrix}$$

and A_{22} is a symmetric matrix given by;

$$A_{22} = \begin{bmatrix} \sum (1-F_t)^2 & & & \\ \sum y_{t-1}(1-F_t)^2 & \sum y_{t-1}^2(1-F_t)^2 & & \\ \sum F_t(1-F_t) & \sum F_t y_{t-1}(1-F_t) & \sum F_t^2 & \\ \sum y_{t-1}F_t(1-F_t) & \sum y_{t-1}^2 F_t(1-F_t) & \sum y_{t-1}F_t^2 & \sum y_{t-1}^2 F_t^2 \end{bmatrix}$$

The vector in the second expression of (6.10) is;

$$\sum x_t u_t = \begin{bmatrix} \sum v_{t-1}u_t \\ \sum v_{t-2}u_t \\ \vdots \\ \sum v_{t-p+1}u_t \\ \sum (1-F_t)u_t \\ \sum y_{t-1}(1-F_t)u_t \\ \sum F_t u_t \\ \sum y_{t-1}F_t u_t \end{bmatrix} \tag{6.12}$$

Under H_0^a, since the true process is a random walk without drift, following Hamilton (1994) we can use the following $(p-1+4) \times (p-1+4)$ diagonal scaling matrix (Υ_T) with diagonal elements $\left(\sqrt{T}, \cdots, \sqrt{T}, \sqrt{T}, T\sqrt{T}, T\right)$.

Premultiplying (6.10) by Υ_T, we can obtain;

$$\Upsilon_T\left(\hat{\beta} - \beta\right) = \left[\Upsilon_T^{-1}\left[\sum x_t x_t'\right]\Upsilon_T^{-1}\right]^{-1}\{\Upsilon_T^{-1}\left[\sum x_t u_t\right]\} \tag{6.13}$$

Now consider the matrix $\left[\Upsilon_T^{-1}\left[\sum x_t x_t'\right]\Upsilon_T^{-1}\right]$. Elements in the upper left $(p-1) \times (p-1)$ block of $\sum x_t x_t'$ (i.e. elements of A_{11}) are divided by T. The first and third row of A_{21} (similarly, first and third column of A'_{21}) are divided by T. The second and fourth row of A_{21} are divided by $T^{3/2}$. On the other hand, those entries that has not y_{t-1} in the sub-matrix A_{22} are divided by T, those that has y_{t-1} are divided by $T^{3/2}$, and those entries with $y_{t-1}^2$are divided by T^2. By the Law of Large Numbers,

$$\frac{1}{T}\sum v_{t-i}v_{t-j} \xrightarrow{P} E\left[v_{t-i}v_{t-j}\right] = \zeta_{|}i-j|.$$

Note that under H_0^a, y_t is a random walk without drift and $y_t/\sqrt{T}$ converges to $\delta B(r)$, $r = t/T$, where $B(.)$ is a standard Brownian motion. Note also that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\langle Tr\rangle} u_t$ converges to $\sigma B(r)$, where $\langle Tr\rangle$i s the largest integer that is less than or equal to Tr. Since the continuous transformations of the exponential transition function $F(z;\lambda,\tilde{c}) = 1 - exp\left[-\left(\frac{\lambda}{\tilde{c}}z-\lambda\right)^2\right]$ are themselves continuous in $z$and in $(\lambda,\tilde{c}) \in \Lambda$ they are regular in the sense of the definition given in Park and Phillips (1999). Therefore we can apply their theorem 3.1 to the remaining terms of the (6.13). For this purpose denote;

$$\tilde{F}(r) = 1 - \exp\left[-\left(\frac{\lambda}{\tilde{c}}\delta B(r) - \lambda\right)^2\right]$$

where $B(r)$ is a standard Brownian motion on $[0,1]$. By theorem 3.1 of Park and Phillips (1999),

$$\frac{1}{T}\sum F_t v_{t-i} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum (1-F_t)v_{t-i} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{\sqrt{T}}(1-F_t)v_{t-1} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{\sqrt{T}}F_t v_{t-1} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{\sqrt{T}}F_t(1-F_t) \xrightarrow{P} \delta\int_0^1 \left(B(r)\tilde{F}(r)\left(1-\tilde{F}(r)\right)\right)dr$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{\sqrt{T}}(1-F_t)^2 \xrightarrow{P} \delta\int_0^1 B(r)\left(1-\tilde{F}(r)\right)^2 dr$$

$$\frac{1}{T}\sum \frac{y_{t-1}^2}{T}F_t(1-F_t) \xrightarrow{P} \delta^2\int_0^1 B(r)^2\tilde{F}(r)\left(1-\tilde{F}(r)\right)dr$$

$$\frac{1}{T}\sum \frac{y_{t-1}^2}{T}(1-F_t)^2 \xrightarrow{P} \delta^2 \int_0^1 B(r)^2 \left(1-\tilde{F}(r)\right)^2 dr$$
$$\frac{1}{T}\sum \frac{y_{t-1}}{T}F_t^2 \xrightarrow{P} \delta^2 \int_0^1 B(r)^2 \tilde{F}(r)^2 dr,$$

pointwise in $(\lambda, \tilde{c}) \in \Lambda$. The convergence here is pointwise rather than uniform as the theorem 3.1 of Park and Phillips (1999) applies here for fixed values of λ and $\tilde{c}$. Ideally, we would like to have a uniform convergence in Λ which is very difficult to prove. To our knowledge, there does not exist a result that extends Park and Phillps's theorem 3.1 to the case where convergence is uniform in Λ. Applying Theorem 3.1 of Park and Phillips to the rest of the terms;

$$\frac{1}{T}\sum F_t(1-F_t) \xrightarrow{P} \int_0^1 \left(\tilde{F}(r)\left(1-\tilde{F}(r)\right)\right) dr$$
$$\frac{1}{T}\sum F_t^2 \xrightarrow{P} \int_0^1 \tilde{F}(r)^2 dr$$
$$\frac{1}{T}\sum (1-F_t)^2 \xrightarrow{P} \int_0^1 \left(1-\tilde{F}(r)\right)^2 dr$$

uniformly in $(\lambda, \tilde{c}) \in \Lambda$.

Hence, we have shown that

$$\left[\Upsilon_T^{-1}\left[\sum x_t x_t'\right]\Upsilon_T^{-1}\right] \xrightarrow{L} \begin{bmatrix} V & 0 \\ 0 & Q \end{bmatrix} \tag{6.14}$$

where

$$V = \begin{bmatrix} \zeta_0 & \zeta_1 & \cdots & \zeta_{p-2} \\ \zeta_1 & \zeta_0 & \cdots & \zeta_{p-3} \\ \vdots & \vdots & & \cdots \\ \zeta_{p-2} & \zeta_{p-3} & \cdots & \zeta_0 \end{bmatrix}$$
$$Q = \begin{bmatrix} Q_{11} & Q_{21}' \\ Q_{21} & Q_{22} \end{bmatrix}$$

with

$$Q_{11} = \begin{bmatrix} \int_0^1 \left(1-\tilde{F}(r)\right)^2 dr & \delta \int_0^1 B(r)\left(1-\tilde{F}(r)\right)^2 dr \\ \delta \int_0^1 B(r)\left(1-\tilde{F}(r)\right)^2 dr & \delta^2 \int_0^1 B(r)^2\left(1-\tilde{F}(r)\right)^2 dr \end{bmatrix}$$

$$Q_{21} = \begin{bmatrix} \int_0^1 \tilde{F}(r)\left(1-\tilde{F}(r)\right)dr & \delta\int_0^1 B(r)\tilde{F}(r)\left(1-\tilde{F}(r)\right)dr \\ \delta\int_0^1 B(r)\tilde{F}(r)\left(1-\tilde{F}(r)\right)dr & \delta^2\int_0^1 B(r)^2\tilde{F}(r)\left(1-\tilde{F}(r)\right)dr \end{bmatrix}$$

$$Q_{22} = \begin{bmatrix} \int_0^1 \tilde{F}(r)^2 dr & \delta\int_0^1 B(r)\tilde{F}(r)^2 dr \\ \delta\int_0^1 B(r)\tilde{F}(r)^2 dr & \delta^2\int_0^1 B(r)^2\tilde{F}(r)^2 dr \end{bmatrix}$$

Now consider the vector, $\Upsilon_T^{-1}\left[\sum x_t u_t\right]$, in (6.13). Following Hamilton (1994, pages 520-21) this term can be decomposed into two parts. Using the result from Hamilton (1994), the first $(p-1)$ elements of this vector satisfy the usual central limit theorem and hence;

$$\begin{bmatrix} \frac{1}{\sqrt{T}}\sum v_{t-1}u_t \\ \frac{1}{\sqrt{T}}\sum v_{t-2}u_t \\ \vdots \\ \frac{1}{\sqrt{T}}\sum v_{t-p+1}u_t \end{bmatrix} \xrightarrow{L} h_1 \sim \aleph\left(0, \sigma^2 V\right) \tag{6.15}$$

The asymptotic behavior of the last four elements can be obtained by using the results in Hamilton and Park and Phillips (1999). For any given $(\lambda,\ \tilde{c})$ we have;

$$\begin{bmatrix} \frac{1}{\sqrt{T}}\sum\left(1-F_t\right)u_t \\ \frac{1}{T}\sum y_{t-1}\left(1-F_t\right)u_t \\ \frac{1}{\sqrt{T}}\sum F_t u_t \\ \frac{1}{T}\sum y_{t-1}F_t u_t \end{bmatrix} \xrightarrow{L} h_2 \sim \begin{bmatrix} \sigma\int_0^1\left(1-\tilde{F}(r)\right)dB(r) \\ \sigma\delta\int_0^1 B(r)\left(1-\tilde{F}(r)\right)dB(r) \\ \sigma\int_0^1 \tilde{F}(r)dB(r) \\ \sigma\delta\int_0^1 B(r)\tilde{F}(r)dB(r) \end{bmatrix} \tag{6.16}$$

Substituting (6.15) through (6.16) into (6.13) results in

$$\Upsilon_T\left(\hat{\beta}-\beta\right) \xrightarrow{L} \begin{bmatrix} V & 0 \\ 0 & Q \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} V^{-1}h_1 \\ Q^{-1}h_2 \end{bmatrix} \tag{6.17}$$

The null hypothesis $H_0^a : \mu = \mu* = 0,\ \rho = \rho* = 1$ can be represented by $R\beta = q$, where $R = \begin{bmatrix} 0 & I_4 \end{bmatrix}$, $q = \begin{pmatrix} 0, & 1, & 0, & 1 \end{pmatrix}'$, with 0 being a $4 \times (p-1)$ zero matrix and I_4b eing the 4×4 identity matrix. The Wald test is then

$$W_T = \left(\hat{\beta} - \beta\right)' R' \left[\hat{\sigma}^2 R \left(\sum x_t x_t'\right)^{-1} R'\right]^{-1} R\left(\hat{\beta} - \beta\right) \tag{6.18}$$

Define $\tilde{\Upsilon}_T$ be the following (4×4) matrix:

$$\tilde{\Upsilon}_T \equiv \begin{bmatrix} \sqrt{T} & 0 & 0 & 0 \\ 0 & T & 0 & 0 \\ 0 & 0 & \sqrt{T} & 0 \\ 0 & 0 & 0 & T \end{bmatrix}. \tag{6.19}$$

Notice that (6.18) can be written

$$W_T = \left(\hat{\beta} - \beta\right)' R'\tilde{\Upsilon}_T \left[\hat{\sigma}^2 \tilde{\Upsilon}_T R \left(\sum x_t x_t'\right)^{-1} R'\tilde{\Upsilon}_T\right]^{-1} \tilde{\Upsilon}_T R\left(\hat{\beta} - \beta\right) \tag{6.20}$$

Observe that the matrix $\tilde{\Upsilon}_T$ has the property that

$$\tilde{\Upsilon}_T R = R\Upsilon_T$$

for $R = \begin{bmatrix} 0 & I_4 \end{bmatrix}$ and Υ_T the $(p+3) \times (p+3)$ diagonal scaling matrix given above. From (6.17),

$$R\Upsilon_T\left(\hat{\beta} - \beta\right) \xrightarrow{L} Q^{-1}h_2.$$

Therefore, (6.20) implies that

$$\begin{aligned} W_T = & \left(\hat{\beta} - \beta\right)' (R\Upsilon_T)' \left[\hat{\sigma}^2 R\Upsilon_T \left(\sum x_t x_t'\right)^{-1} \Upsilon_T R\right]^{-1} \Upsilon_T R\left(\hat{\beta} - \beta\right) \\ & \xrightarrow{L} \left(Q^{-1}h_2\right)' \left[\sigma^2 Q^{-1}\right]^{-1} \left(Q^{-1}h_2\right) \\ & = h_2 Q^{-1} h_2/\sigma^2 \equiv \zeta(\varphi) \end{aligned} \tag{6.21}$$

Note that under the alternative hypothesis y_t follows a stationary ESTAR process for $\rho*$ is strictly less than 1. Under the alternative parameters β will converge asymptotically in $\sqrt{T}$ to their pseudo true values that are functions of γ and c. Hence, the test statistic should diverge.

Proof of Proposition 2: Note that under H_0^b since the process is a random walk with drift (i.e. $y_t = \mu + y_{t-1} + u_t$) we need to use the following diagonal matrix with the diagonal elements $\left(\sqrt{T}, \cdots, \sqrt{T}, \sqrt{T}, T^{3/2}, \sqrt{T}, T^{3/2}\right)$. Note also that under H_0^b y_t/T converges to μ as $T \to \infty$. Since under the null the OLS estimate of μ is consistent we can act as if we know μ. Denote

$$\tilde{F}(\mu) = 1 - \exp\left(-\left(\frac{\lambda}{\tilde{c}}\mu - \lambda\right)^2\right).$$

Using the theorem 3.1 in Park and Phillips (1999) and proceeding as in the proof of the proposition 1 we can show that;

$$\frac{1}{T}\sum F_t(1-F_t) \xrightarrow{P} \int \left(\tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right)\right) d\tilde{F}(\mu)$$

$$\frac{1}{T}\sum F_t^2 \xrightarrow{P} \int \tilde{F}(\mu)^2 d\tilde{F}(\mu)$$

$$\frac{1}{T}\sum (1-F_t)^2 \xrightarrow{P} \int \left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu)$$

uniformly in $(\lambda, \tilde{c}) \in \Lambda$. The rest of the terms converges in probability pointwise in $(\lambda, \tilde{c}) \in \Lambda$. That is,

$$\frac{1}{T}\sum F_t v_{t-i} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum (1-F_t) v_{t-i} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{\sqrt{T}}(1-F_t) v_{t-1} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{T} F_t v_{t-1} \xrightarrow{P} 0$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{T} F_t(1-F_t) \xrightarrow{P} \int \frac{\mu}{2}\tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right) d\tilde{F}(\mu)$$

$$\frac{1}{T}\sum \frac{y_{t-1}}{T}(1-F_t)^2 \xrightarrow{P} \int \frac{\mu}{2}\left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu)$$

$$\frac{1}{T}\sum \frac{y_{t-1}^2}{T^2} F_t(1-F_t) \xrightarrow{P} \int \frac{\mu^2}{3}\tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right) d\tilde{F}(\mu)$$

$$\frac{1}{T}\sum \frac{y_{t-1}^2}{T^2}(1-F_t)^2 \xrightarrow{P} \int \frac{\mu^2}{3}\left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu)$$

$$\frac{1}{T}\sum \frac{y_{t-1}^2}{T^2}F_t^2 \xrightarrow{P} \int \frac{\mu^2}{3}\tilde{F}(\mu)^2 d\tilde{F}(\mu).$$

pointwise in $(\lambda, \tilde{c}) \in \Lambda$. In the above, integration is over the support of μ. Applying the similar steps in the proof of proposition 1 we can obtain:

$$\left[\Upsilon_T^{-1}\left[\sum x_t x_t'\right]\Upsilon_T^{-1}\right] \xrightarrow{L} \begin{bmatrix} V & 0 \\ 0 & Q \end{bmatrix} \tag{6.22}$$

where now, V is the same as above and Q becomes:

$$Q = \begin{bmatrix} Q_{11} & Q_{21}' \\ Q_{21} & Q_{22} \end{bmatrix}$$

with

$$Q_{11} = \begin{bmatrix} \int \left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu) & \int \frac{\mu}{2}\left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu) \\ \int \frac{\mu}{2}\left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu) & \int \frac{\mu^2}{3}\left(1-\tilde{F}(\mu)\right)^2 d\tilde{F}(\mu) \end{bmatrix}$$

$$Q_{21} = \begin{bmatrix} \int \tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right) d\tilde{F}(\mu) & \int \frac{\mu}{2}\tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right) d\tilde{F}(\mu) \\ \int \tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right) d\tilde{F}(\mu) & \int \frac{\mu^2}{3}\tilde{F}(\mu)\left(1-\tilde{F}(\mu)\right) d\tilde{F}(\mu) \end{bmatrix}$$

$$Q_{22} = \begin{bmatrix} \int \tilde{F}(\mu)^2 d\tilde{F}(\mu) & \int \frac{\mu}{2}\tilde{F}(r)^2 dr \\ \int \frac{\mu}{2}\tilde{F}(\mu)^2 d\tilde{F}(\mu) & \int \frac{\mu^2}{3}\tilde{F}(\mu)^2 d\tilde{F}(\mu) \end{bmatrix}$$

The limiting distribution of the first $(p-1)\times(p-1)$ elements of the vector, $\Upsilon_T^{-1}\left[\sum x_t u_t\right]$, is given in (6.15). The last four elements of this vector follows asymptotically,

$$\begin{bmatrix} \frac{1}{\sqrt{T}}\sum(1-F_t)u_t \\ \frac{1}{T^{3/2}}\sum y_{t-1}(1-F_t)u_t \\ \frac{1}{\sqrt{T}}\sum F_t u_t \\ \frac{1}{T^{3/2}}\sum y_{t-1}F_t u_t \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \frac{1}{\sqrt{T}}\sum u_t \\ \frac{1}{\sqrt{T}}\sum\left(1-\tilde{F}(\mu)\right)u_t \\ \frac{1}{T^{3/2}}\sum \mu(t-1)\tilde{F}(\mu)u_t \\ \frac{1}{\sqrt{T}}\sum \tilde{F}(\mu)u_t \\ \frac{1}{T^{3/2}}\sum \mu(t-1)\tilde{F}(\mu)u_t \end{bmatrix} \xrightarrow{L} \frac{1}{\sigma^2}\aleph\left(0,\sigma^2 Q\right) \tag{6.23}$$

Combining each component of 6.13, it follows that

$$\Upsilon_T\left(\hat{\beta}-\beta\right) \xrightarrow{L} \begin{bmatrix} V^{-1}h_1, \\ \aleph\left(0,\sigma^2 Q^{-1}\right) \end{bmatrix} \tag{6.24}$$

Under the null H_0^b consider the following selection matrix;

$$R = \begin{bmatrix} 0 & R_4 \end{bmatrix}$$

where 0 is a $(4) \times (p-1)$ zero matrix, and

$$R_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and define $\tilde{\Upsilon}_T$ now to be

$$\tilde{\Upsilon}_T \equiv \begin{bmatrix} \sqrt{T} & 0 & 0 & 0 \\ 0 & T^{3/2} & 0 & 0 \\ 0 & 0 & \sqrt{T} & 0 \\ 0 & 0 & 0 & T^{3/2} \end{bmatrix}. \tag{6.25}$$

Proceeding in a similar fashion to the proof of the proposition 1 we can show that the asymptotic distribution of W_T is

$$W_T - \rightarrow \frac{1}{\sigma^2}\aleph\left(0,\sigma^2 Q^{-1}\right)' Q^{-1}\aleph\left(0,\sigma^2 Q^{-1}\right) \longrightarrow \chi^2(\varphi) \tag{6.26}$$

By the same argument given in the proof of proposition 1, under the alternative the Wald test should diverge. This completes the proof.

BIBLIOGRAPHY

[1] Caner, M. and B. E. Hansen (2001), Threhold autoregression with a unit root, *Econometrica* **69** 1555-1596.

[2] Chan, K.S., J.D. Petrucelli, H. Tong, and S.W. Woolford (1985), A multiple threshold AR(1) model, *Journal of Applied Probability* **22**, 267–279.

[3] Dickey, D. and W. Fuller, (1981), Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* **49**, 1057-1072.

[4] Dumas, B. (1992), Dynamic equilibrium and the real exchange rate in a spatially separated world, *Review of Financial Studies* **5**, 153–180.

[5] Eitrheim Ø. and T. Teräsvirta (1996), Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–76.

[6] Granger, C.W.J. and T. Teräsvirta (1993), Modelling Nonlinear Economic Relationships, Oxford: Oxford University Press.

[7] Hamilton, J. (1994), Time Series Analysis, Princeton, New Jersey: Princeton University Press.

[8] Hansen, B. E. (1997), Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics* **1,** 119-131.

[9] Killian, L. and M. Taylor (2001), Why is it difficult to beat the random walk forecast of exchange rates? Mansucrpipt, Department of Economics, University of Michigan.

[10] Koop, G., M. H. Pesaran and S. M. Potter (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119-147.

[11] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* **54,** 159-178.

[12] Micheal, P., R.A. Nobay, and D.A. Peel (1997), Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation, *Journal of Political Economy* **105**, 862–879.

[13] O'Connel, P.G.J. (1998), Market frictions and real exchange rates, *Journal of International Money and Finance* **17**, 71–95.

[14] Park, J. Y. and P. C. B. Phillips (1999), Nonlinear regressions with integrated time series, working paper, Department of Economics, Yale University.

[15] Phillips, P. and C.B.P. Perron (1988), Testing for a unit root in time series regression, *Biometrika* **75,** 335-346.

[16] Rogoff (1996), The purchasing power parity puzzle, *Journal of Economic Literature* **34,** 647-668.

[17] Sercu, P., R. Uppal, and C. Van Hulle (1995), The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity, *Journal of Finance* **10**, 1309–19.

[18] Taylor, M. P. and L. Sarno (1998), The behavior of real exchange rates during the post-Bretton Woods period, *Journal of International Economics* **46,** 281-312.

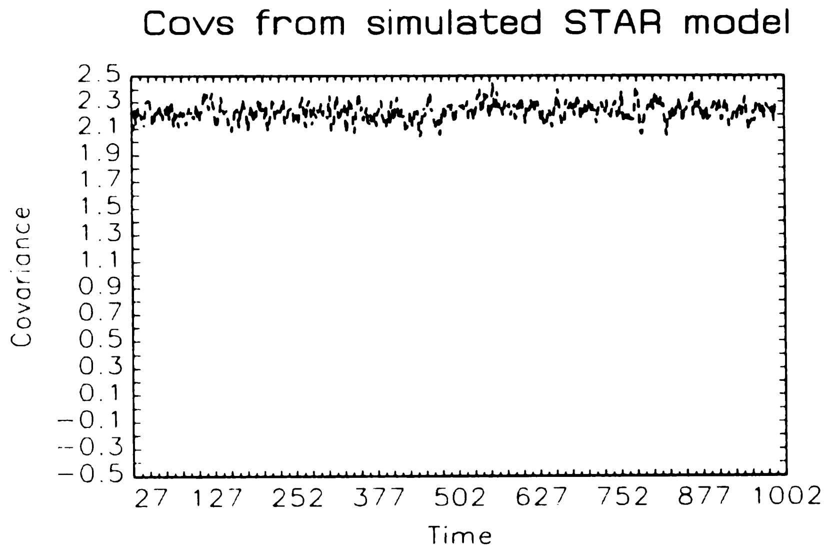
[19] Taylor, M.P., D.A. Peel, and L. Sarno (2001), Non-linear in real exchange rates: towards a solution of the purchasing power parity puzzles, Working Paper, Centre for Economic Policy Research, London, UK.

[20] Teräsvirta, T. (1994), Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89,** 208–218.

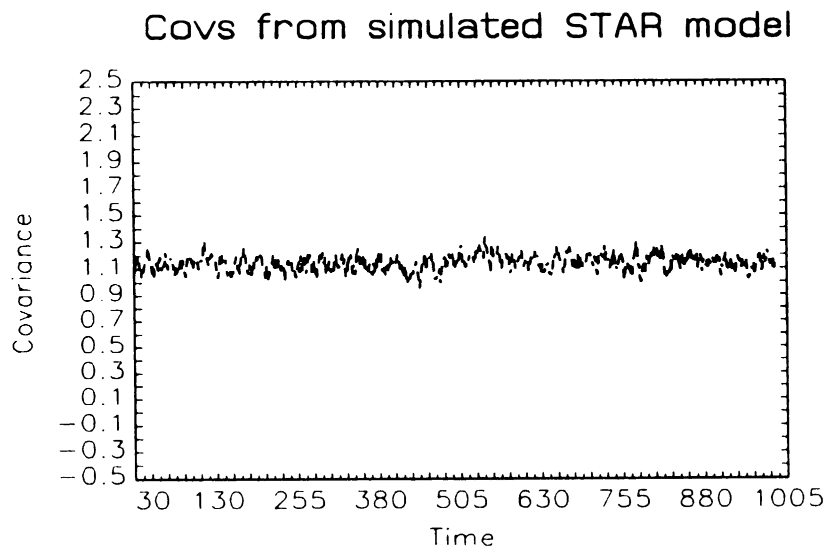
[21] Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (editors), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

[22] Tjøstheim, D. (1990), Nonlinear time series and Markov Chains, *Advances in applied probability* **22,** 587-611.

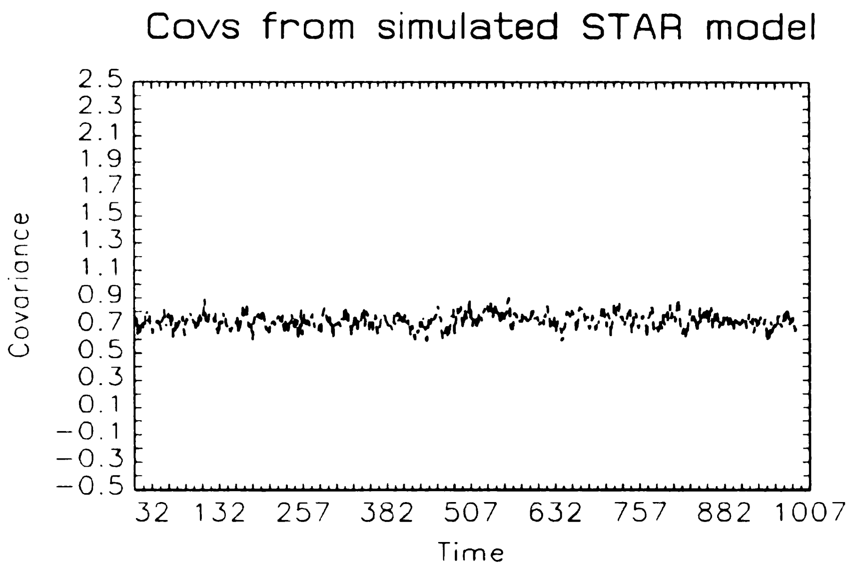
Figure 6.1: Estimated j-step ahead covariances from the simulated ESTAR model
(a) $\hat{\delta}_{t,2}$



(b) $\hat{\delta}_{t,5}$



(c) $\hat{\delta}_{t,7}$



(d) $\hat{\delta}_{t,9}$

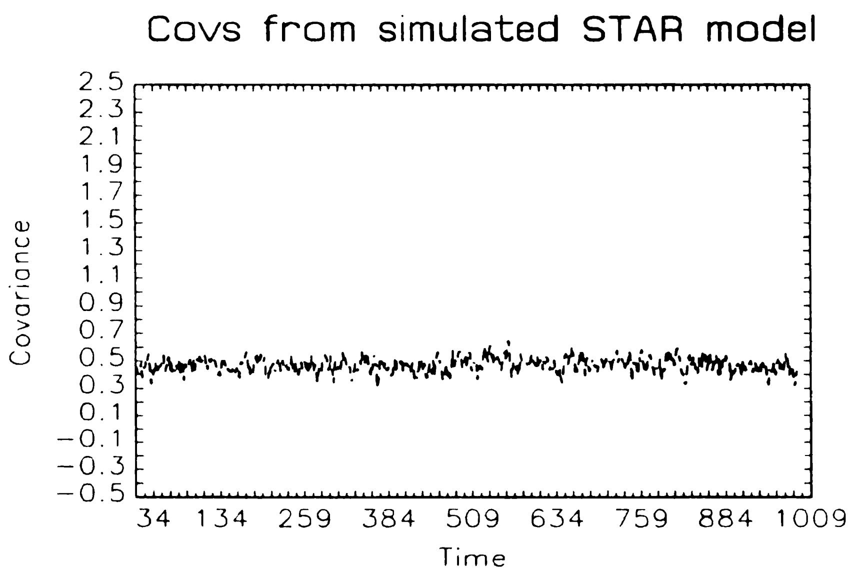


Figure 6.2: Real exchange rate series and fitted values, residuals, and estimated transition function versus time and transition variable
(a)BP

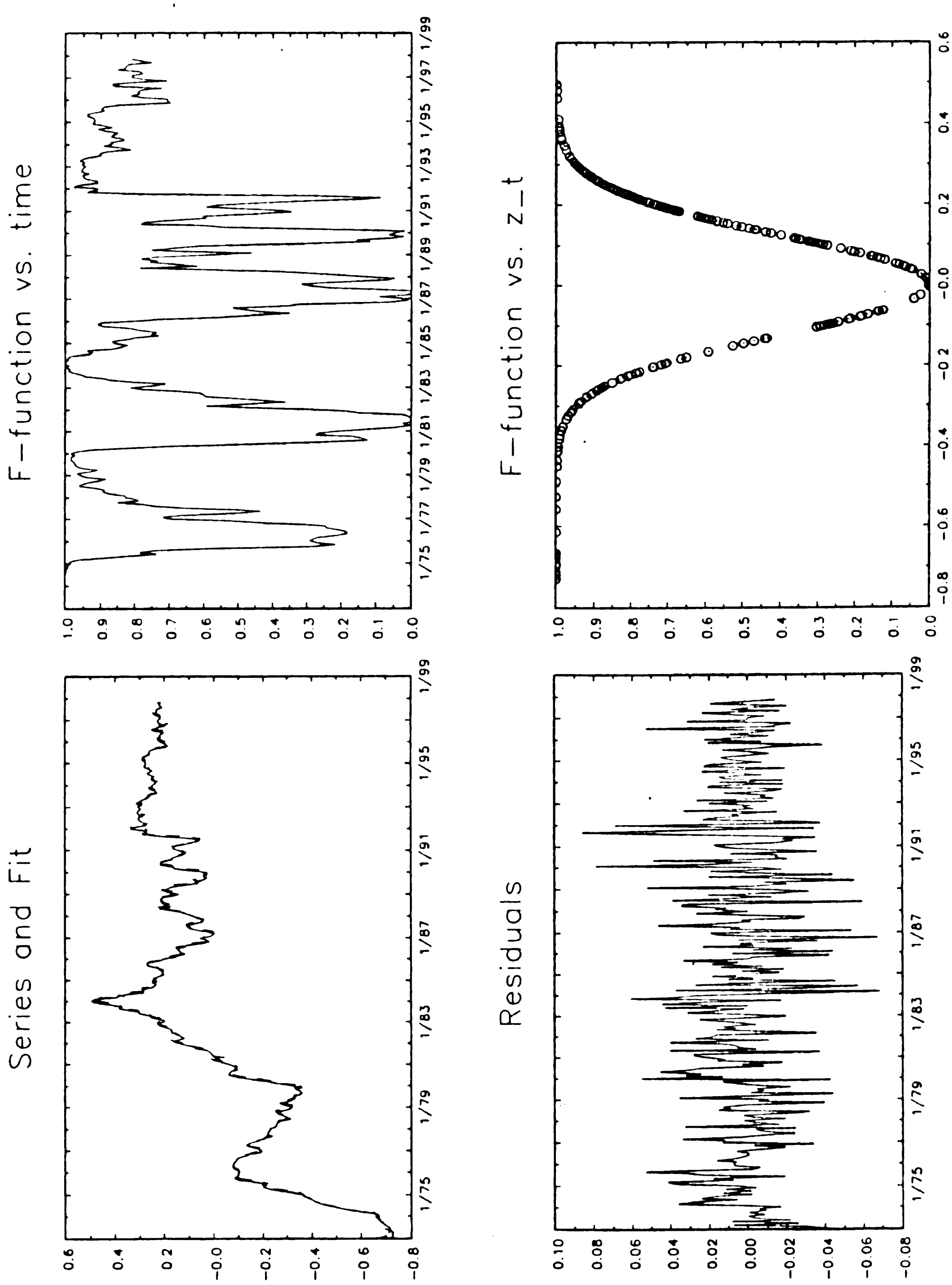


Figure 6.2 (cont'd).

(b) CD

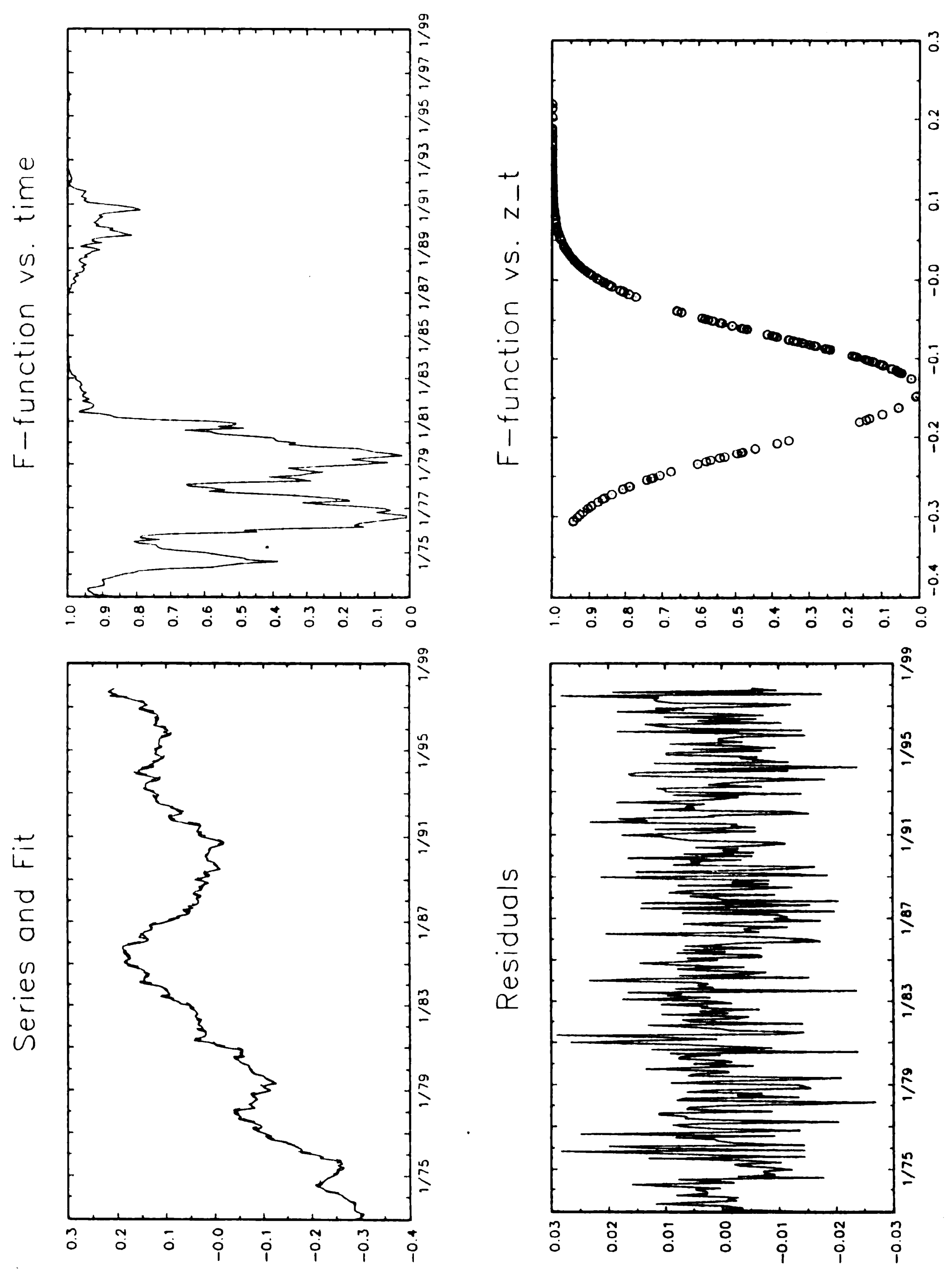


Figure 6.2 (cont'd).

(c) DG

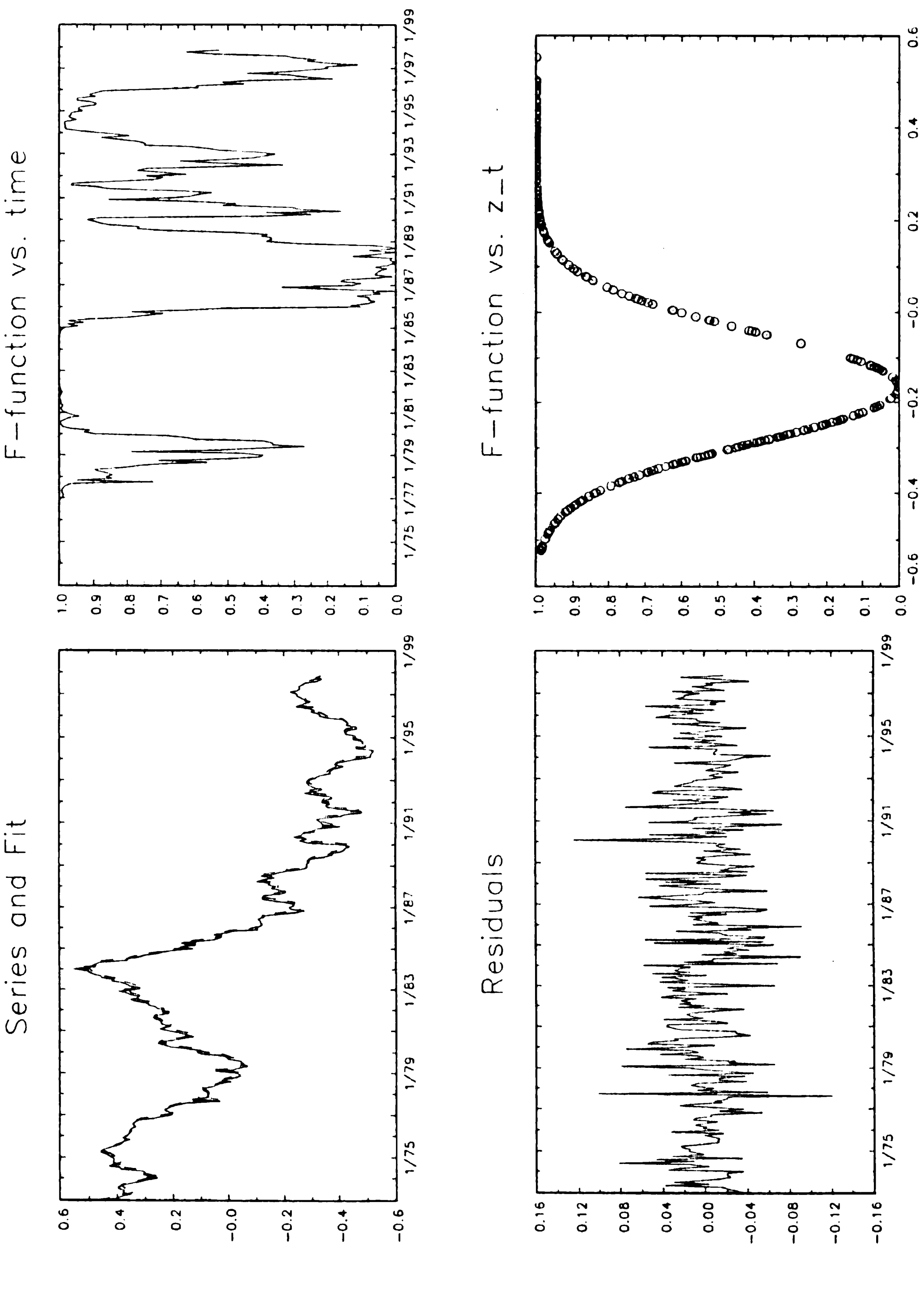


Figure 6.2 (cont'd).

(d) GM

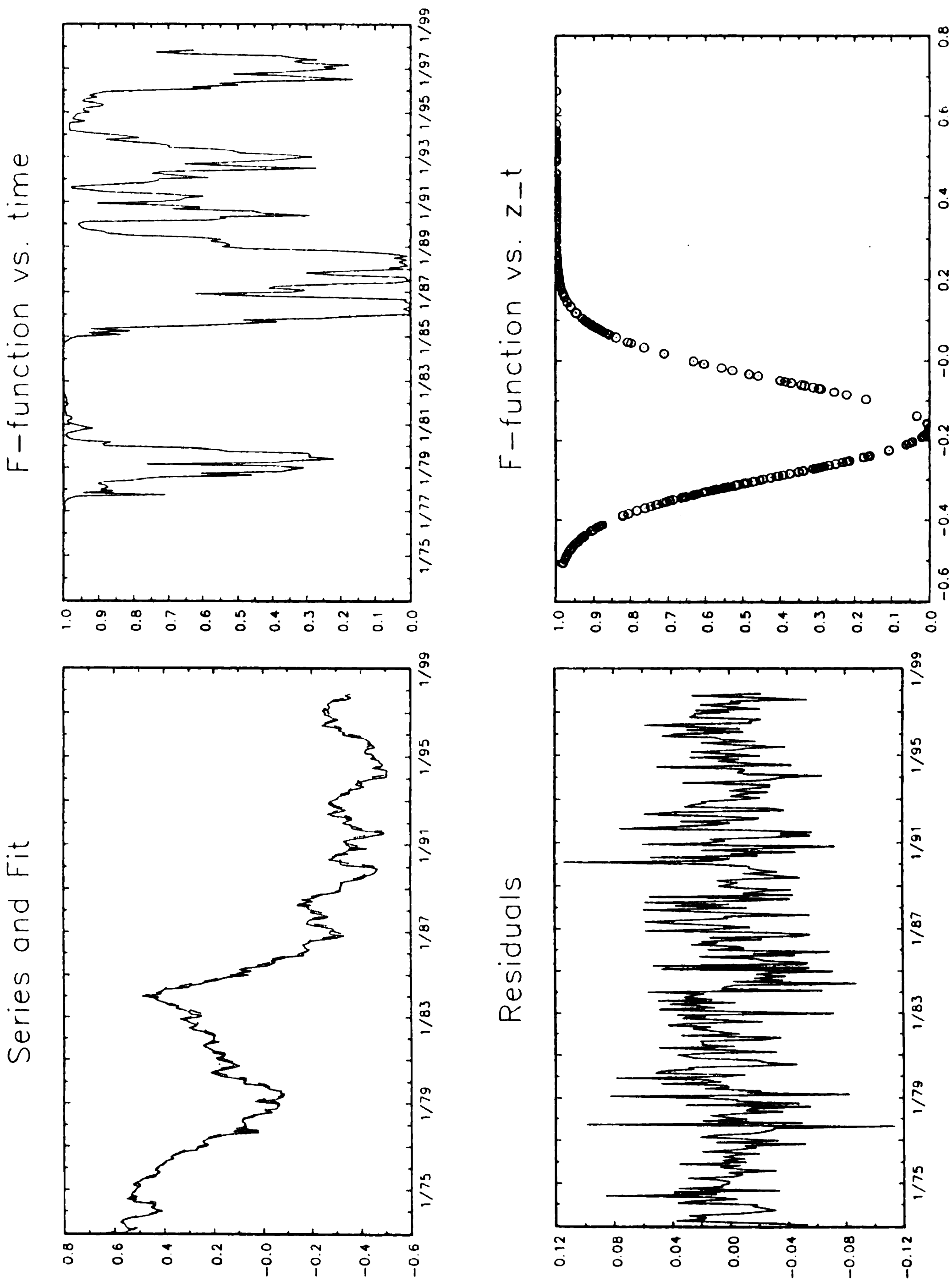


Figure 6.2 (cont'd).

(d) IL

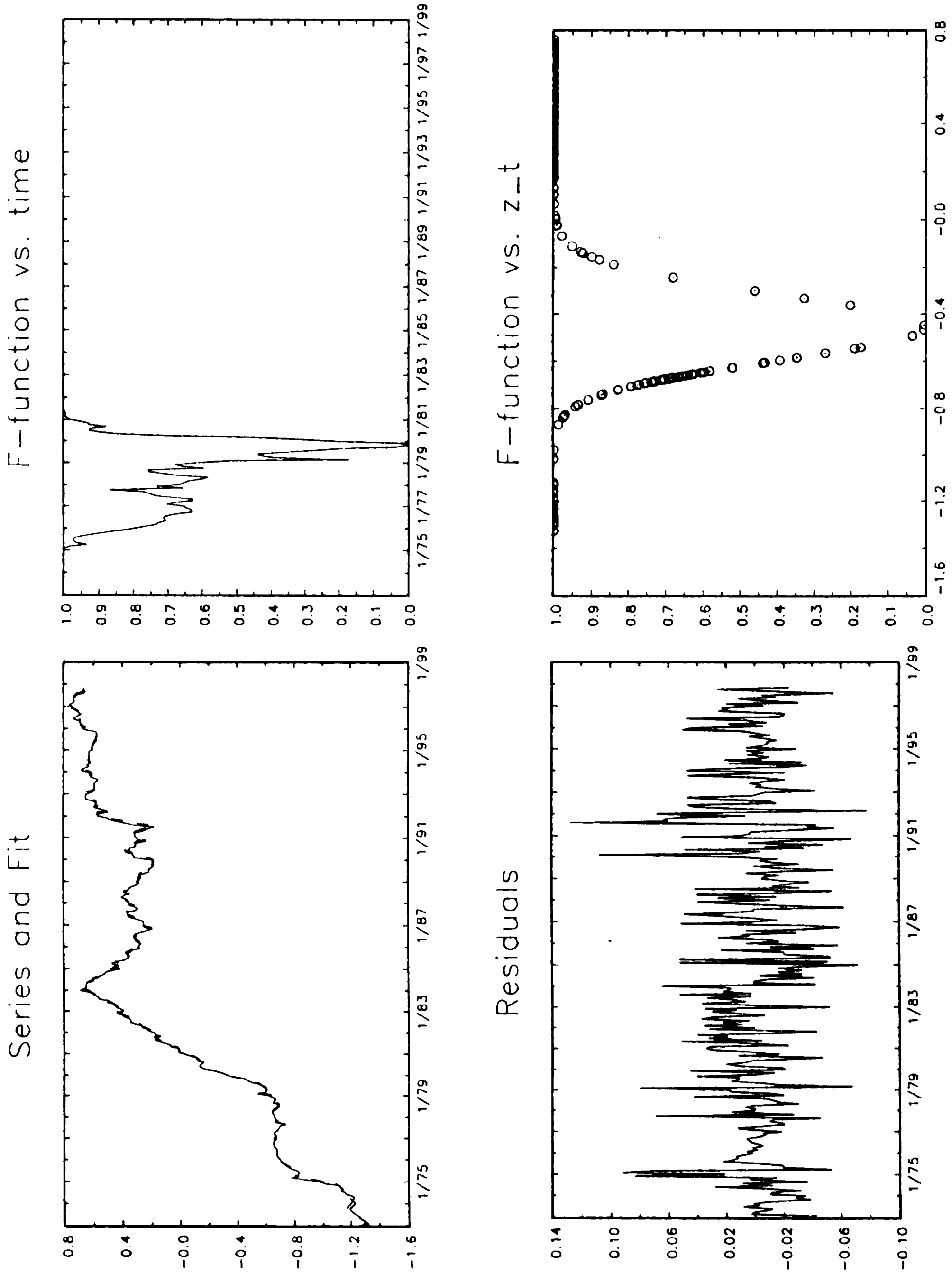


Figure 6.2 (cont'd).

(f) JY

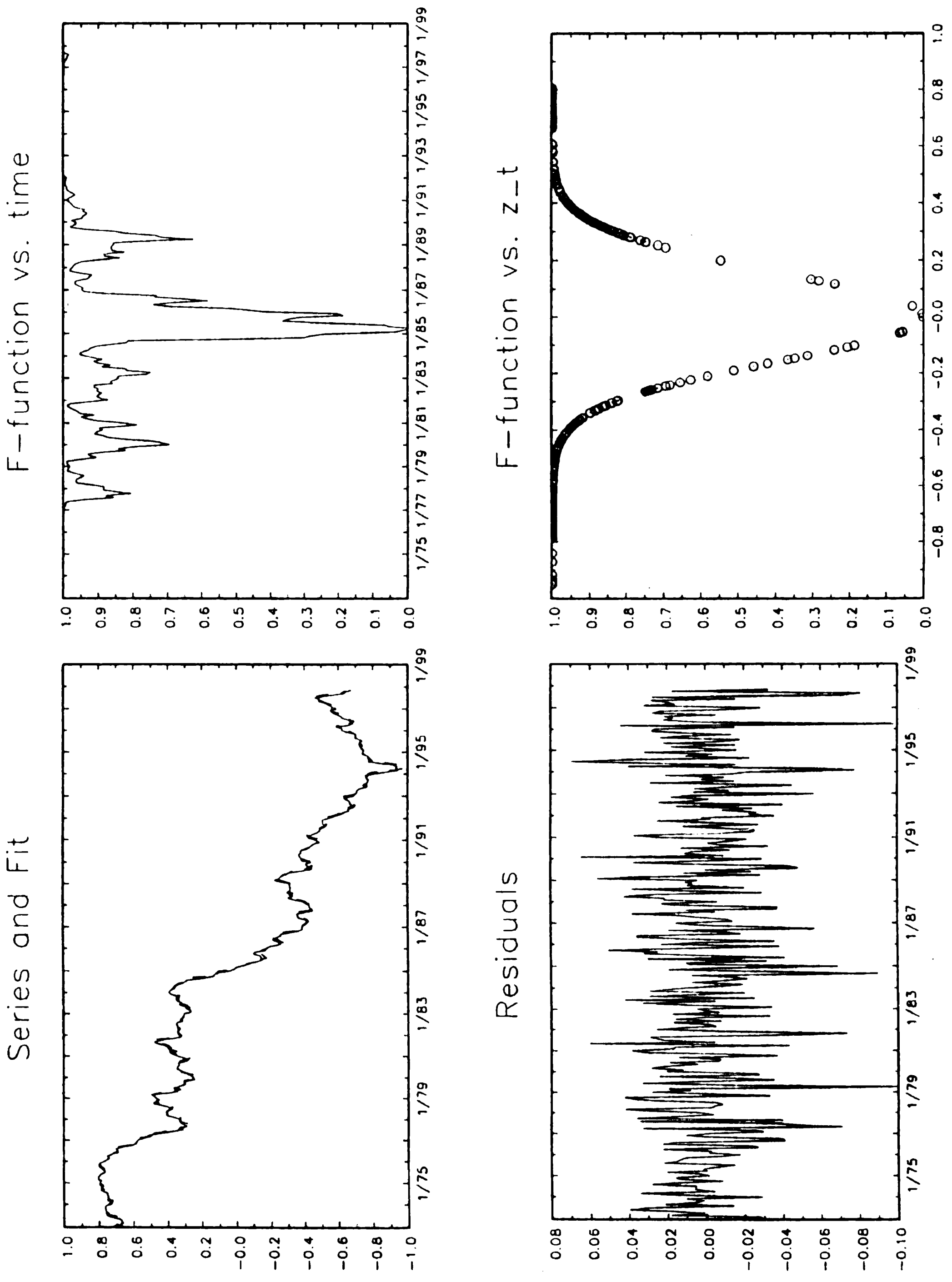


Figure 6.2 (cont'd).

(g) <u>SF</u>

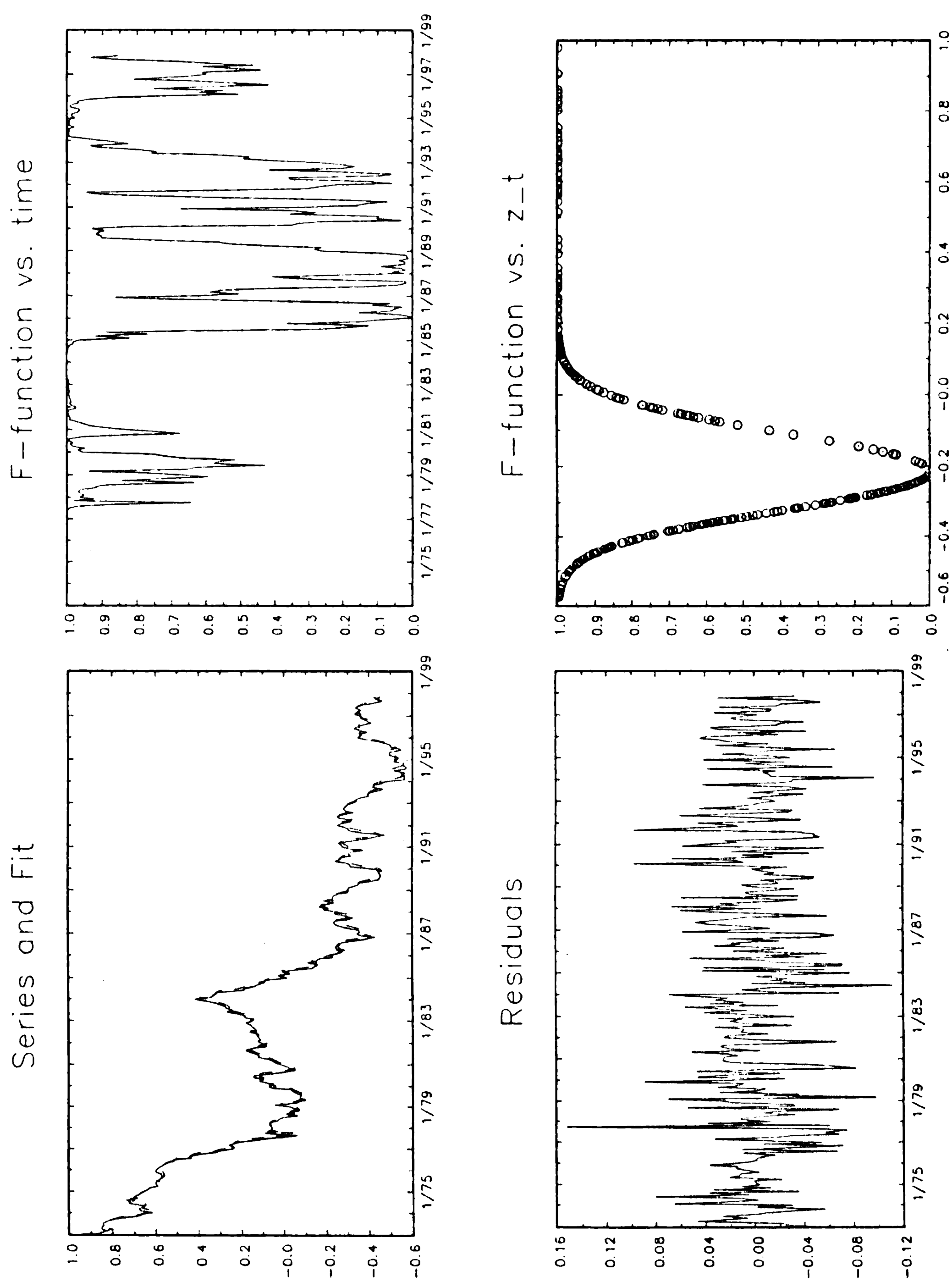
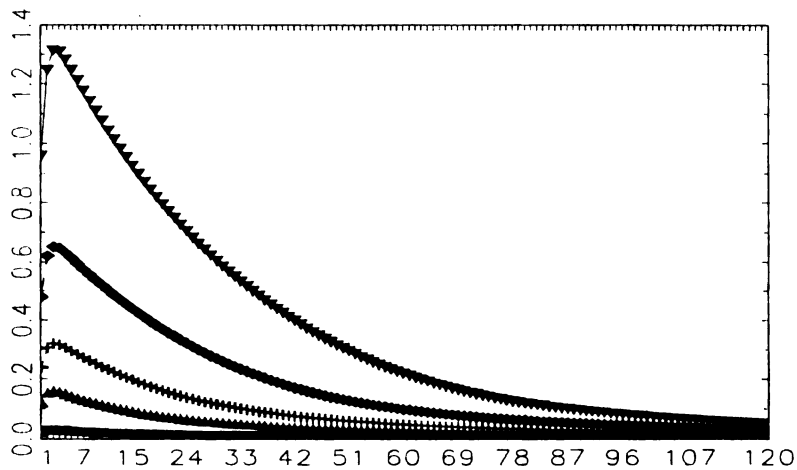


Figure 6.3: Generalized Impulse Response Functions from Estimated ESTAR Models

(a)BP



(b)CD

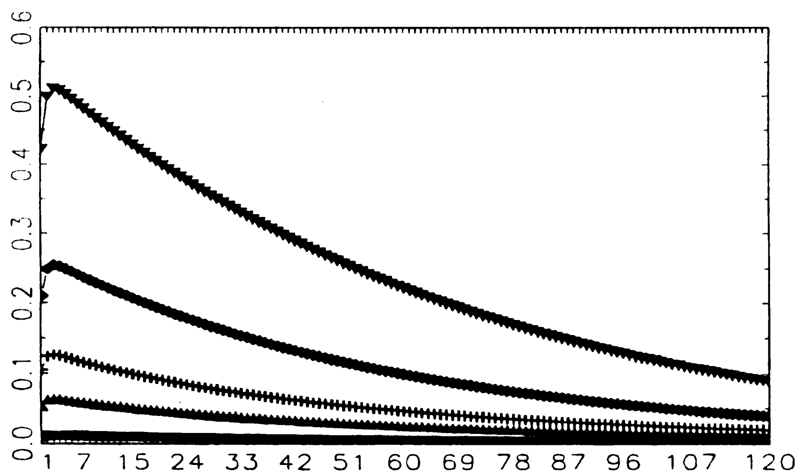
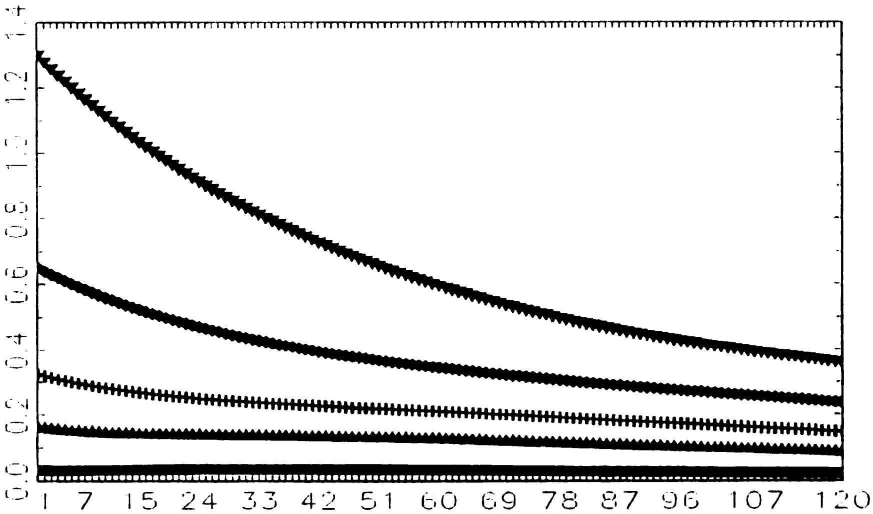


Figure 6.3 (cont'd).

(c)<u>DG</u>



(d)<u>GM</u>

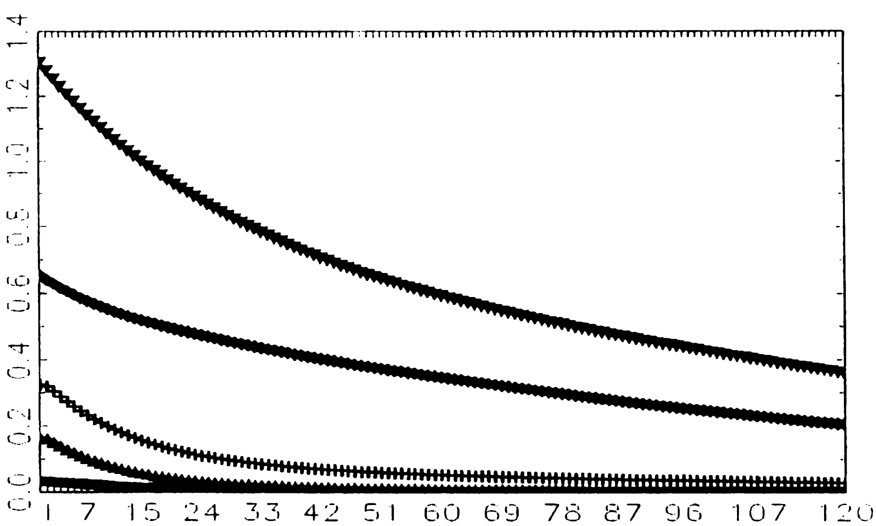
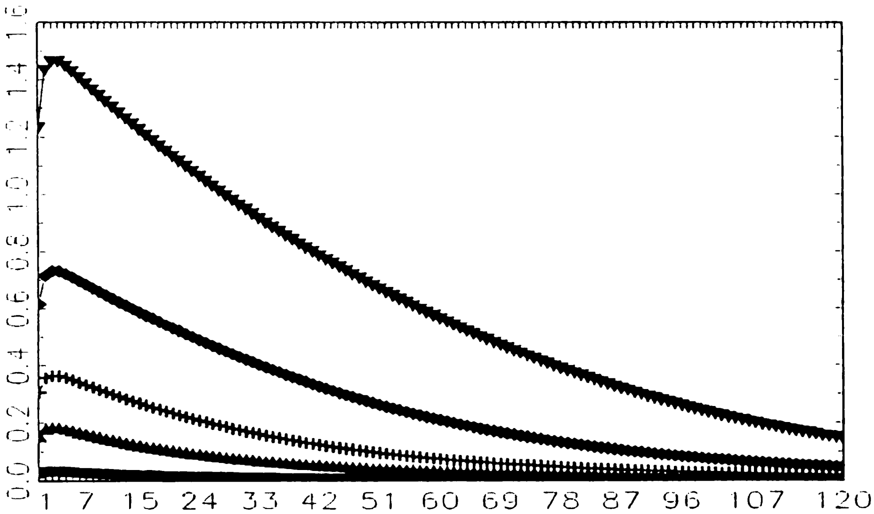


Figure 6.3 (cont'd).

(e) IL



(f)JY

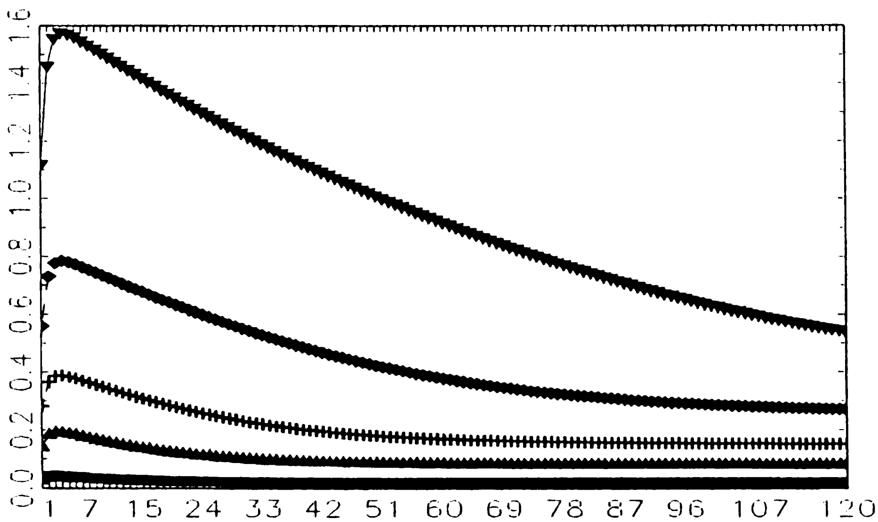


Figure 6.3 (cont'd).

(g) SF

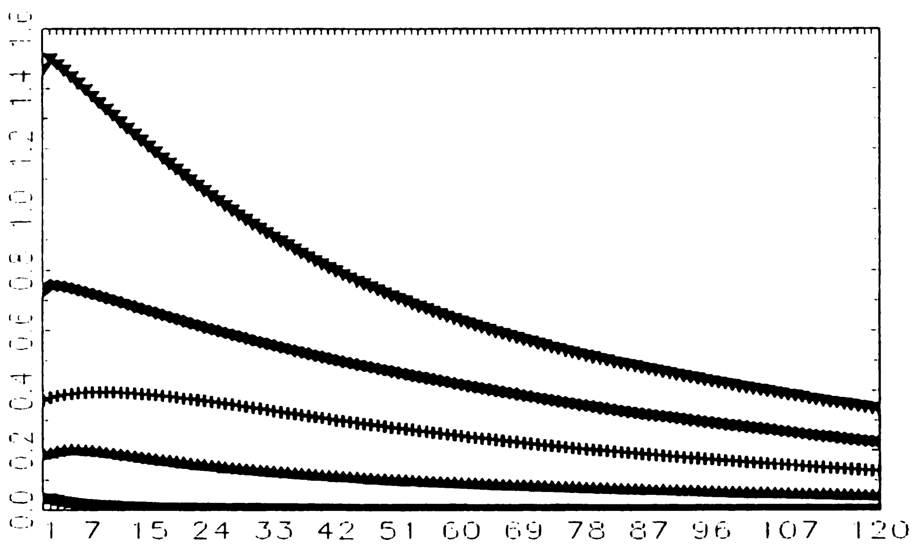
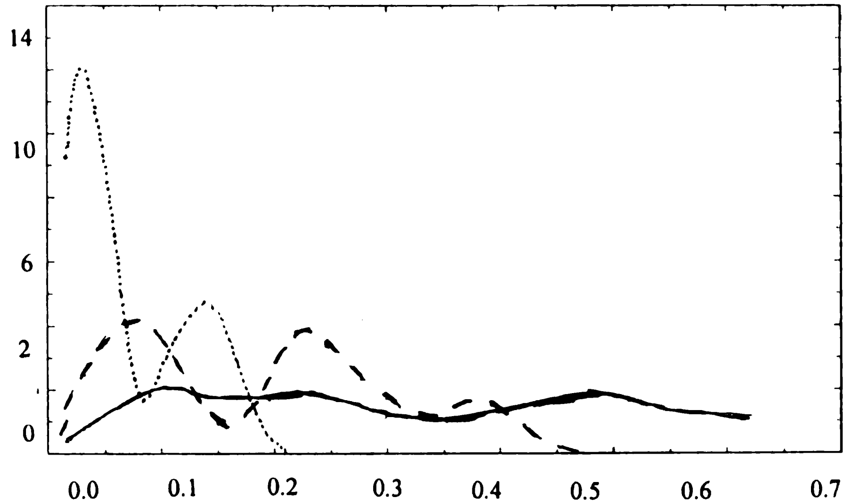


Figure 6.4: Distribution of Generalized Impulse Responses

(a)BP



(b)CD

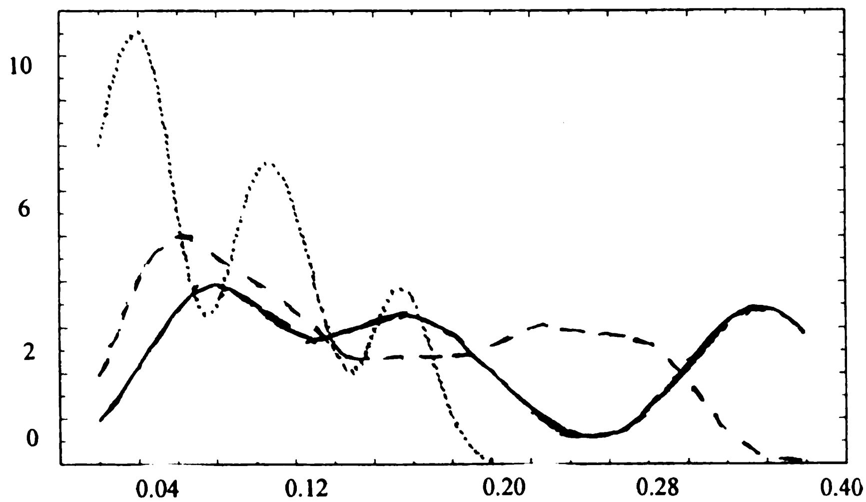
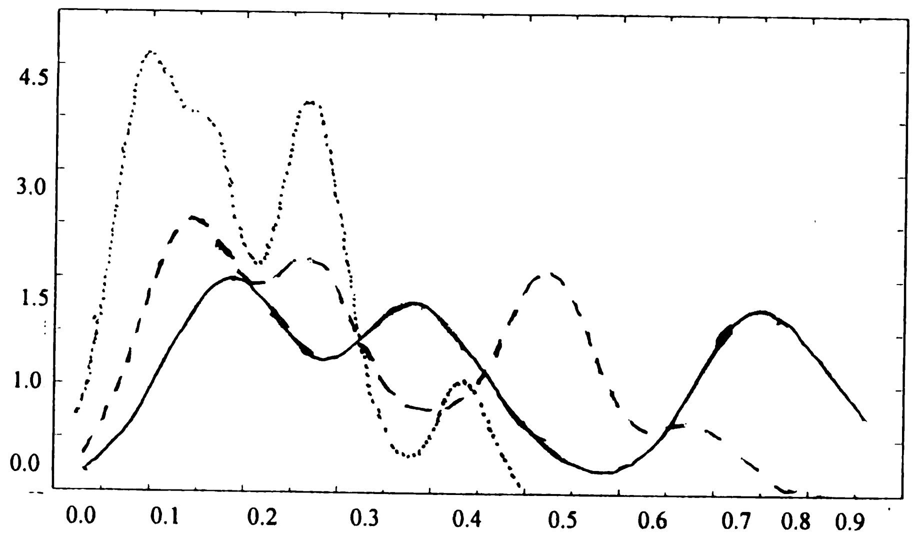


Figure 6.4 (cont'd).

c)DG



(d)GM

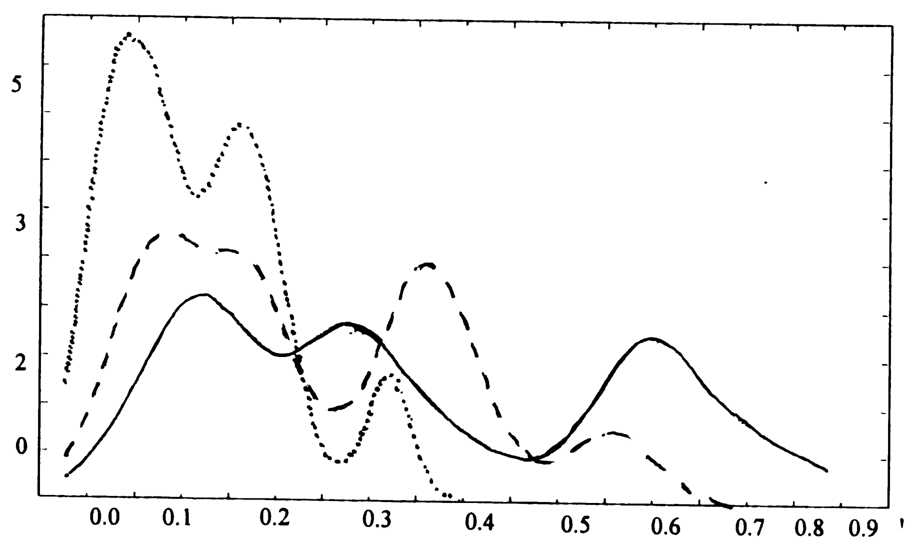
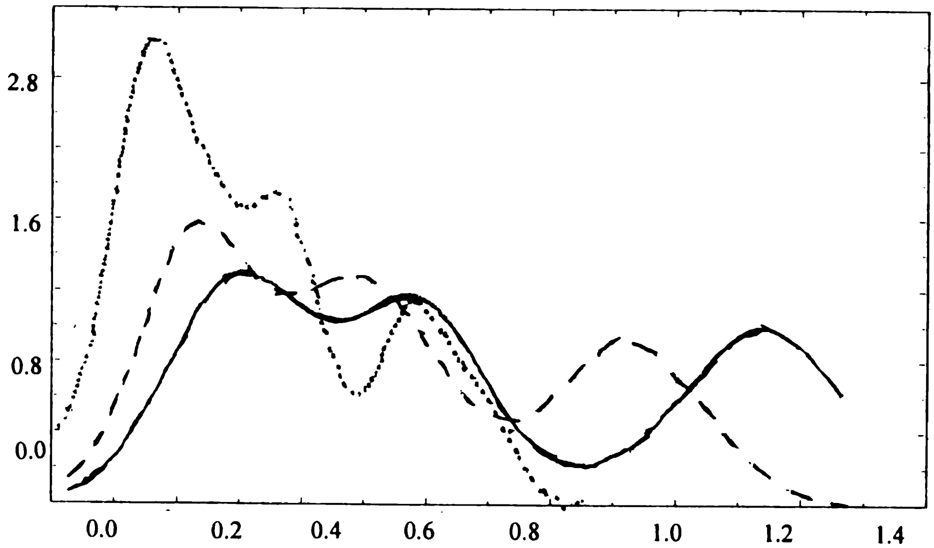


Figure 6.4 (cont'd).

(e) IL



(f)JY

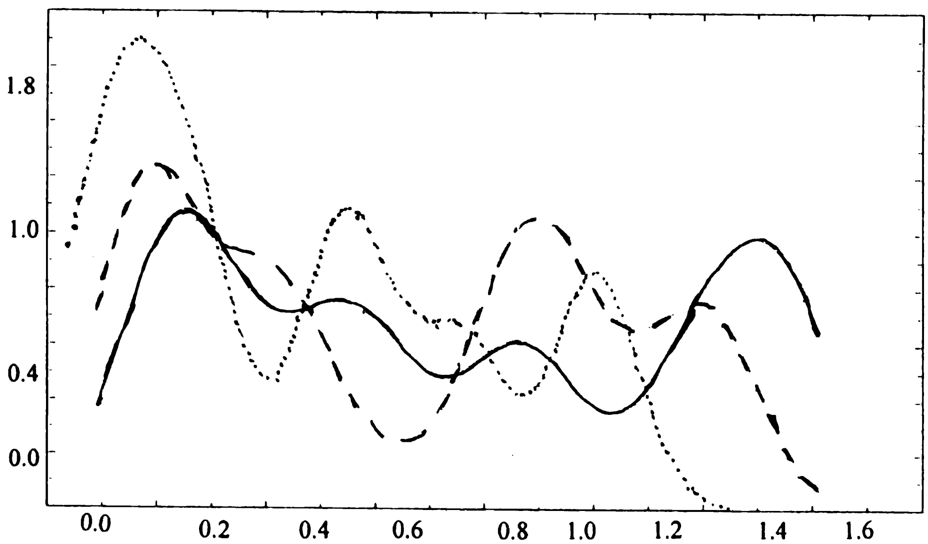


Figure 6.4 (cont'd).

(g) SF

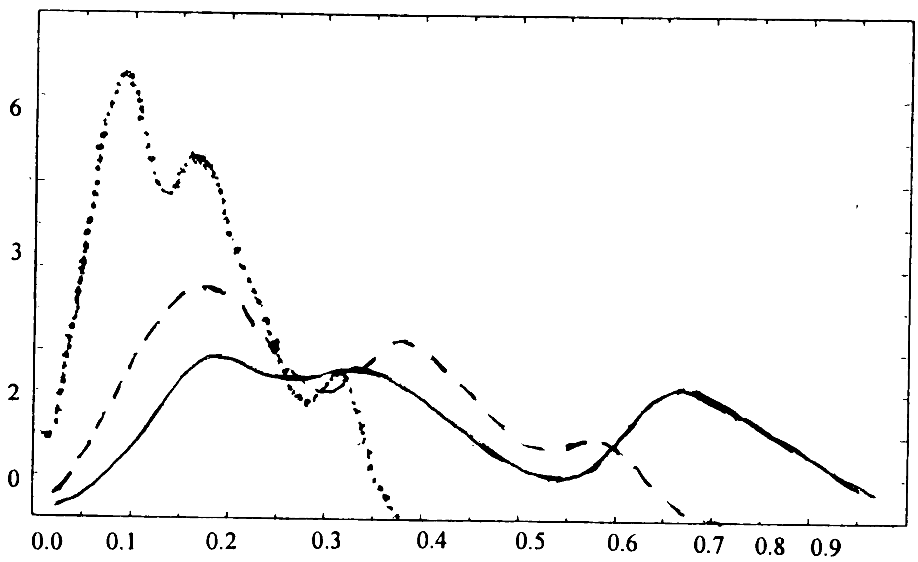


Table 6.1: Empirical critical values of the unit root tests

	1%	5%	10%	15%	20%	80%	85%	90%	95%	99%
$supW$	0.912	1.972	2.756	3.370	3.950	13.140	15.084	17.860	23.849	44.077
$supWh$	0.960	2.094	2.941	3.590	4.230	14.548	16.732	20.268	26.817	45.915
$supW_{\mu}$	0.244	0.945	1.621	2.207	2.733	11.430	13.247	15.886	21.631	41.532
$supWh_{\mu}$	0.286	1.040	1.741	2.391	2.984	12.514	14.427	17.537	23.868	42.473

Notes: $supW$ **and** $supWh$ **stand for the standard and heteroscedasticity robust version of sup Wald test for testing random walk without drift against a stationary ESTAR alternative while** $supW_{\mu}$ **and** $supWh_{\mu}$ **stand for the standard and heteroscedasticity robust versions of the sup Wald tests of random walk with drift against stationary ESTAR alternative. Critical values are computed from 20,000 replications and** $\mu = 0.05$ **and errors are drawn from** $iid\,\aleph(0\,1)$.

Table 6.2: Empirical size of the unit root tests

Theoretical Size	ADF	PP	$supW$	$supWh$	$supW_{\mu}$	$supWh_{\mu}$
1%	0.013	0.012	0.011	0.010	0.011	0.010
5%	0.050	0.051	0.054	0.052	0.044	0.041
10%	0.102	0.102	0.106	0.100	0.078	0.077

Notes:The columns corresponding to $supW$ **and** $supWh$ **give the rejection frequencies of true null hypotheses of random walk without drift, while the columns corresponding to** $supW_{\mu}$ **and** $supWh_{\mu}$ **give the rejection frequencies of true null of random walk with drift. The data is generated under the nulls of** H_0^a **and** H_0^b **with** $\mu = 0$ **and** $\mu = 0.05$**.The rejection frequencies for ADF and PP corresponds to** $\mu = 0$.

Table 6.3: Empirical power of the unit root tests

a. $\gamma = 2.5$, $c = 0.05$, $\mu = \mu* = 0$

Test	$\rho = 1.0\, \rho* = -0.5$			$\rho = 1.0\, \rho* = 0.5$			$\rho = 1.0\, \rho* = 0.95$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
ADF	0.970	0.975	0.978	0.835	0.850	0.865	0.410	0.445	0.450
PP	0.968	0.977	0.980	0.805	0.844	0.866	0.400	0.425	0.448
$supW$	1.000	1.000	1.000	0.995	0.998	0.999	0.479	0.685	0.803
$supWh$	1.000	1.000	1.000	0.995	0.996	0.997	0.446	0.633	0.750
$supW_{\mu}$	1.000	1.000	1.000	0.996	0.998	0.999	0.507	0.712	0.813
$supWh_{\mu}$	1.000	1.000	1.000	0.993	0.995	0.997	0.481	0.666	0.787

b.$\gamma = 15$, $c = 0.05$, $\mu = \mu* = 0$

Test	$\rho = 1.0\, \rho* = -0.5$			$\rho = 1.0\, \rho* = 0.5$			$\rho = 1.0\, \rho* = 0.95$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
ADF	0.961	0.970	0.972	0.828	0.839	0.855	0.3850	0.411	0.420
PP	0.962	0.975	0.977	0.788	0.812	0.846	0.378	0.405	0.417
$supW$	1.000	1.000	1.000	0.998	0.998	1.000	0.499	0.715	0.817
$supWh$	1.000	1.000	1.000	0.996	0.997	0.998	0.476	0.673	0.785
$supW_{\mu}$	1.000	1.000	1.000	0.995	0.998	0.999	0.538	0.740	0.833
$supWh_{\mu}$	1.000	1.000	1.000	0.994	0.996	0.998	0.494	0.688	0.801

c. $\gamma = 2.5$, $c = 0.05$, $\mu = 0.05 \mu* = -0.05$

Test	$\rho = 1.0\, \rho* = -0.5$			$\rho = 1.0\, \rho* = 0.5$			$\rho = 1.0\, \rho* = 0.95$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
ADF	0.935	0.950	0.958	0.810	0.822	0.835	0.377	0.400	0.414
PP	0.932	0.950	0.960	0.776	0.811	0.836	0.375	0.400	0.413
$supW$	1.000	1.000	1.000	0.997	0.996	0.998	0.500	0.714	0.820
$supWh$	1.000	1.000	1.000	0.995	0.996	0.998	0.488	0.675	0.790
$supW_{\mu}$	1.000	1.000	1.000	0.999	1.000	1.000	0.667	0.814	0.885
$supWh_{\mu}$	1.000	1.000	1.000	0.996	0.999	0.999	0.628	0.773	0.850

c. $\gamma = 15$, $c = 0.05$, $\mu = 0.05 \mu* = -0.05$

Test	$\rho = 1.0\, \rho* = -0.5$			$\rho = 1.0\, \rho* = 0.5$			$\rho = 1.0\, \rho* = 0.95$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
ADF	0.935	0.950	0.958	0.812	0.820	0.837	0.377	0.400	0.414
PP	0.930	0.948	0.956	0.789	0.817	0.837	0.375	0.400	0.413
$supW$	1.000	1.000	1.000	0.998	0.998	0.998	0.524	0.746	0.834
$supWh$	1.000	1.000	1.000	0.995	0.997	0.999	0.488	0.713	0.820
$supW_{\mu}$	1.000	1.000	1.000	1.000	1.000	1.000	0.679	0.829	0.890
$supWh_{\mu}$	1.000	1.000	1.000	0.999	1.000	1.000	0.645	0.794	0.861

Notes:The rows corresponding to $supW$ and $supWh$ give the rejection frequencies of false null hypotheses of random walk without drift, while the rows corresponding to $supW_{\mu}$ and $supWh_{\mu}$ give the rejection frequencies of false null of random walk with drift. The data is generated under the alternative hypothesis of globally stationary ESTAR model.

Table 6.4: Results on unit root and stationarity tests:PP, supWald and KPSS tests

	PP	KPSS	ADF	$supW$	$supWh$	$supW_{\mu}$	$supWh_{\mu}$
BP	-2.571	2.242	-3.009	24.505	29.024	n.a.	n.a.
CD	-1.192	2.357	-1.382	1462.232	1749.536	n.a.	n.a
GM	-2.126	2.675	-1.784	2547.000	2617.812	n.a.	n.a.
IL	-2.697	2.675	-2.785	49.679	55.319	13.058	19.663
JY	-0.376	3.041	-0.136	58.269	65.303	34.965	33.632
DG	-1.536	2.570	-1.311	3030.276	3191.674	n.a.	n.a.
SF	-2.440	2.665	-2.112	249.205	269.036	n.a.	n.a.

Key: The reported values for the PP test are based on the regression of the time series on a constant and its lagged value. The lag truncation for the Bartlett kernel is obtained from the formula $floor(4(\frac{T}{100})^{2/9})$. The 1%, 5% and 10% critical values are -3.454, -2.871, and -2.570 respectively for the PP tests. The reported values for the KPSS test are based on a regression of the series on a constant only. The 1%, 5%, and 10% critical values for the KPSS tests are 0.739, 0.463 and 0.347 respectively. The size of the Bartlett window for KPSS is obtained by using $floor(8(\frac{T}{100})^{1/4})$. ADF test is based on the regression of first differenced real exchange rate on a constant, lagged real exchange rate and $p - 1$ lags of the first differenced real exchange rate. The lag length is chosen according to the Ljung-Box statistic and for all real exchange rates found to be 1. The 1%, 5%, and 10% critical values for ADF test are -3.454, -2.871, and -2.570.

Table 6.5: Estimation Results from ESTAR models: Sample size: 312

	BP	CD	DG	GM	IL	JY	SF
ϕ_1	0.004	0.002	.	.	0.001	0.003	.
	(0.001)	(0.001)	.	.	(0.000)	(0.001)	.
ϕ_2	-0.002	.	.	.	.	.	.
	(0.001)	.	.	.	.	.	.
μ	.	.	.	.	0.024	-0.017	.
	.	.	.	.	(0.007)	(0.009)	.
ρ	1 .054	1.002	1.035	1.042	0.946	1.065	1.037
	(0.053)	(0.007)	(0.034)	(0.036)	(0.028)	(0.093)	(0.022)
$\mu*$	.	.	.	.	0.004	-0.004	.
	.	.	.	.	(0.002)	(0.002)	.
$\rho*$	0.983	0.996	0.984	0.981	0.993	0.996	0.978
	(0.007)	(0.020)	(0.008)	(0.007)	(0.003)	(0.003)	(0.006)
γ	9.049	14.011	10.466	11.736	5.120	10.480	16.436
	(0.730)	(1.157)	(1.792)	(1.673)	(0.420)	(0.835)	(1.582)
	[0.032]	[0.007]	[0.025]	[0.021]	[0.028]	[0.018]	[0.013]
c	.	-0.140	-0.017	-0.169	-0.456	.	-0.215
	.	(0.038)	(0.150)	(0.143)	(0.040)	.	(0.120)
Skew	0.344	0.078	0.030	0.050	0.542	-0.694	-0.015
Kurt	3.737	0.210	4.053	3.663	4.229	3.905	3.706
$pLM(1-6)$	0.139	0.136	0.444	0.234	0.236	0.242	0.453
$pLM(1-12)$	0.390	0.064	0.593	0.396	0.277	0.291	0.534
$pNLES_{max}$	0.185	0.873	0.767	0.753	0.205	0.163	0.470
$pNLLS_{max}$	0.114	0.149	0.027	0.389	0.243	0.306	0.072
SSR	0.173	0.034	0.315	0.321	0.277	0.230	0.406
$pLMc$	0.326	0.797	0.659	0.692	0.091	0.153	0.574

Heteroscedasticity robust standard errors are given underneath the parameter estimates. The values in squared parentheses are the computed marginal significance levels. The rows corresponding to $pLM(1-6)$ and $pLM(1-12)$ are the p-values from Lagrange Multiplier test statistics for up to 6th and 12th order serial correlations in residuals respectively, constructed as in Eitrheim and Teräsvirta (1996). $pNLES_{max}$ is the p-value for maximal Lagrange multiplier test statistic for no remaining ESTAR nonlinearity with delay in the range from 2 to 12 (Eitrheim and Teräsvirta, 1996). $pNLLS_{max}$ is the p-value corresponding to no remaining LSTAR nonlinearity with delay in the range 1 to 12 (Eitrheim and Teräsvirta, 1996). SSR is the sum squared residuals of regression. $pLMc$ is p-value for Lagrange multiplier test statistic for parameter constancy in the estimated ESTAR model (Eitrheim and Teräsvirta, 1996).