Nonlinear long memory models with applications in finance

Paolo Zaffaroni

1997

A Thesis submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy in Economics at the London School of Economics and Political Science (University of London)

to Marianna
and my family

Contents

List of Figures

List of Tables

Abstract

The last decade has witnessed a great deal of research in modelling volatility of financial asset returns, expressed by time-varying variances and covariances. The importance of modelling volatility lies in the dependence of any financial investment decision on the expected risk and return as formalized in classical asset pricing theory. Precise evaluation of volatilities is a compulsory step in order to perform correct options pricing according to recent theories of the term structure of interest rates and for the construction of dynamic hedge portfolios. Models of time varying volatility represent an important ground for the development of new estimation and forecasting techniques for situations not reconcilable with the Gaussian or, more generally, a linear time series framework. This is particularly true for the statistical analysis of time series with long range dependence in a nonlinear framework. The aim of this thesis is to introduce parametric nonlinear time series models with long memory, with particular emphasis on volatility models, and to provide a methodology which yields asymptotically exact inference on the parameters of the models. The importance of these results stems from: (i) rigorous asymptotics was lacking from the stochastic volatility literature; (ii) the statistical literature does not cover the analysis of the asymptotic behaviour of quadratic forms in nonlinear non-Gaussian variates that characterizes our problem.

NOTATION:

K denotes any arbitrary constant.

$A(x) \sim B(x)$ as $x \to x_0$ denotes that the ratio $A(x)/B(x)$ converges to one as $x \to x_0$ (asymptotic equivalence).

For any function $h(\lambda;\phi)$, $-\pi \leq \lambda < \pi$, ϕ a $q \times 1$ vector , $\phi \in \Phi \subset R^q$, integrable over $[-\pi,\pi]$, $\gamma_h(u;\phi)$, $u = 0, \pm 1, \ldots$ denote the sequence of its Fourier coefficients.

For any (second-order) real stationary processes $\{Z_t\}$, $\{W_t\}\, t = 0, \pm 1, \ldots$:

$\gamma_Z(l)$, $l = 0, \pm 1, \pm 2$ denotes the autocovariance function,

$\rho_Z(l)$, $l = 0, \pm 1, \ldots$ denotes the autocorrelation function,

$f_Z(\lambda)$, $-\pi \leq \lambda < \pi$ denotes the power spectrum (if it exists).

$\gamma_{ZW}(l)$, $l = 0, \pm 1, \ldots$ denotes the crosscorrelation function,

$f_{ZW}(\lambda)$, $-\pi \leq \lambda < \pi$ denotes the cross spectrum (if it exists) where $f_{WZ}(\omega) = f_{ZW}(-\omega)$.

$Q_{ZZZ}(\lambda_1, \lambda_2, \lambda_3)$ denotes the trispectrum for Z_t, $-\pi \leq \lambda_i < \pi$, $i = 1, 2, 3$.

For any sequence a_t, $t = 1, \ldots, T$, $\bar{a} = \frac{1}{T}\sum_{t=1}^{T} a_t$.

$tr(A)$ expresses the trace and $\mid A \mid$ the determinant for any matrix A.

$\delta(u, v)$ denotes the Kronecker delta.

$I(A)$ denotes the indicator function which takes value one if the event A is true and zero otherwise.

$\delta_{ab}(l;\theta) = \sum_{i=1}^{\infty} a_i(\theta) b_{i+l}(\theta)$, $l > 0$ integer for arbitrary $a_i(\theta), b_i(\theta), i = 1, 2, \ldots$ both functions of some vector θ, including the constant function case $a_i(\theta) = a_i$, $b_i(\theta) = b_i$.

I_T denotes the identity matrix of dimension $T \times T$.

1_T denotes a $T \times 1$ vector of ones.

$D_E(\theta) = \frac{\partial}{\partial \theta} vecE(\theta)$ for some matrix $E(\theta)$ function of a vector θ, denotes the gradient with respect to θ.

$\| . \|$ denotes the Euclidean norm.

$\mathcal{N}_p(\mu, \Sigma)$ denotes a random vector $p \times 1$ normally distributed with mean μ and variance covariance matrix Σ.

χ_p denotes a random variable distributed like a central chi-square with p degrees of freedom.

$\rightarrow_p$ denotes convergence in probability.

$\rightarrow_d$ denotes convergence in distribution.

Acknowledgments

I thank my supervisor, Professor Peter M. Robinson for his continuous interest in my progress, his support and guidance, during the preparation of this thesis. In particular I thank him for proposing the class of models analyzed in the thesis and for many other suggestions. I have also benefited from discussions with Dr. Javier Hidalgo and Professor Jan Magnus and comments by Professor Danny Quah and Professor Enrique Sentana at seminars at the LSE.

I thank Professor Luigi Spaventa and Professor Marco Lippi for encouraging me to continue my studies abroad. I thank Adriana Barone, Luis Gil-Alana, Claudio Michelacci, Giovanni Marseguerra and Carlo Velasco for constant support.

Research support from Banca d'Italia, EC HCM grant n. ERBCHBITC941742 and ESRC grant n. R000235892 is gratefully acknowledged.

I thank my family all for their continuous support during my studies. Finally I thank my wife Marianna for her patience, encouragement and for providing all the serenity that I needed for this work. To her and to my family this effort is dedicated.

Chapter 1

Introduction

In this chapter we consider some aspects of the empirical distribution of financial asset returns thus motivating the importance of volatility models. In particular we review the basic volatility models both in the short and long memory case and we assess the limits of linear modelling in order to capture the dynamic features of financial data.

1.1 Models of time varying volatility

1.1.1 Some stylized facts.

Let us consider seven time series of asset returns. In all cases we calculate the return as $x_t = ln(P_t/P_{t-1})$ for $t = 1, 3, \ldots, 3089$ where P_t denotes the speculative price of the asset. In particular we will consider the exchange rate *Yen/Pound spot* and *forward* , the exchange rate *Dollar/Pound spot* and *forward* and the return indices *FTSE 100* , *FTSE All* and the *S&P 500 All*.

The data are daily and run from 1 Jan 1985 through 1 Nov 1996 and are plotted in Figure 1.1 (forex) and Figure 1.2 (stock returns).

In Table 1.1 we report the Ljung and Box (1978) statistic based on the first 24 and 300

Figure 1.1: Plots of the time series of foreign exchange rate returns. Daily data from 2 Jan 1985 to 1 Nov 1996.

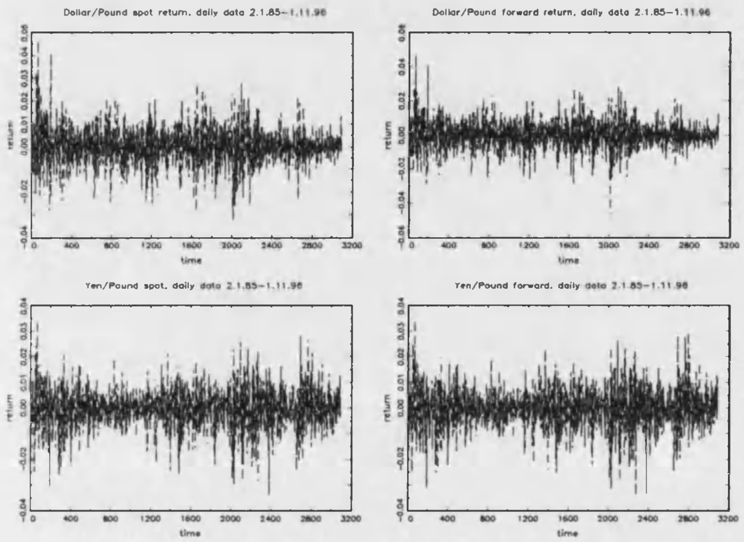


Figure 1.2: Plots of the time series of stock return indexes. Daily data from 2 Jan 1985 to 1 Nov 1996.

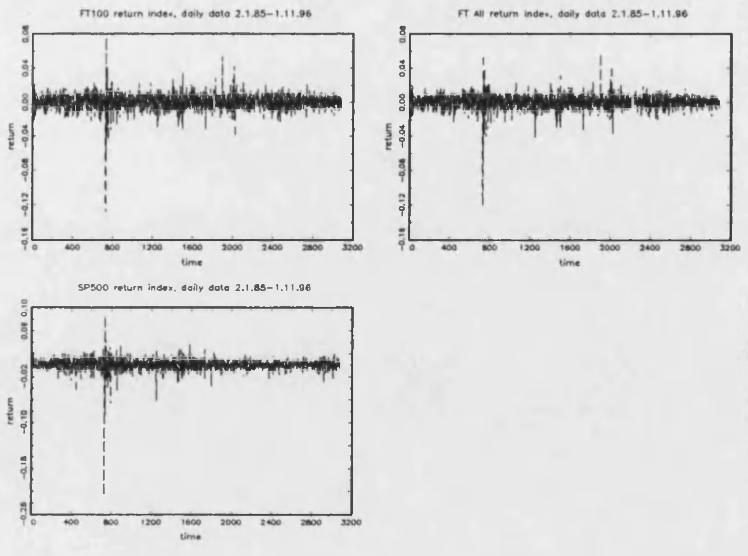


Figure 1.3: Foreign exchange rate returns: correlograms lag $0 - 300$ for the time series of returns in the 'levels' (first row) and in the 'squares' (second row).

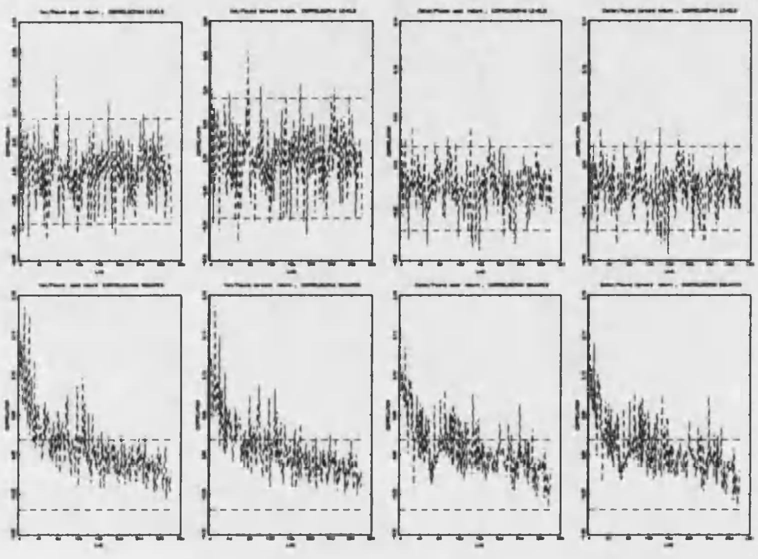


Figure 1.4: Stock return indexes: correlograms lag $0 - 300$ for the time series of returns in the 'levels' (first row) and in the 'squares' (second row).

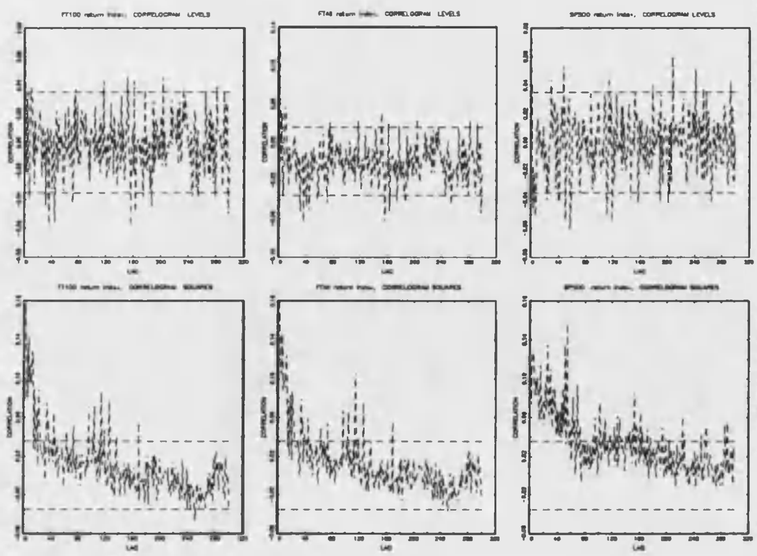


Table 1.1: **Summary Statistics:**

Data	$Q(24)\,(p-value)$	$Q^2(24)$	$Q(300)\,(p-value)$	$Q^2(300)$	kurtosis	skewness
sYP	53.76 (exp-3.34)	738.15	338.68 (exp-1.21)	1665.86	6.23	0.001
fYP	50.70 (exp-2.94)	635.36	338.58 (exp-1.20)	1437.41	6.80	0.001
sUP	73.13 (exp-6.14)	607.01	371.09 (exp-2.49)	1412.63	5.48	0.079
fUP	69.37 (exp-5.56)	580.32	362.45 (exp-2.11)	1363.59	5.68	0.075
F100	48.66 (exp-2.68)	697.64	386.97 (exp-3.29)	1144.66	5.35	0.002
FAll	88.59 (exp-8.61)	706.78	447.15 (exp-7.14)	1131.53	6.84	0.008
S500	38.45 (exp-1.51)	605.49	426.77 (exp-5.71)	1872.24	8.89	0.35

"sYP" = spot Yen/Pound, "fYP" = forward Yen/Pound,
"sUP" = spot Dollar/Pound, "fUP" = forward Dollar/Pound,
"F100" = FTSE 100, "FAll" = FTSE All, "S500" = S&P500 .

The data refers to the period 1^{st} Jan 1986 to 1^{st} Nov 1993 (3088 observations).
Columns $2-5$ report the Ljung and Box (1984) test statistic with 24 and 300 lags respectively for the data in the level (columns 2 and 4) and in the squares (columns 3 and 5), .
with p-value in parentheses (not reported in columns 3 and 5 because negligible).
For the stock return indexes we skipped the week starting on the Monday 17 October 1987.

sample autocorrelations for the raw data x_t and for the squares $y_t = x_t^2$ in the first two columns ($Q(24)$, $Q^2(24)$) and third and fourth ($Q(300)$, $Q^2(300)$) respectively. In parentheses we report the p-value based on the usual χ^2 approximation for the raw data figures only, the others being highly significant. Finally in the last two columns we report the sample coefficients of kurtosis and skewness for the raw returns.

The test statistic is always highly significant for the squares y_t and, to a much smaller degree, about half of the times for the x_t. But where the two cases really differ is in the magnitude of the test statistic itself. In other words the level of significance is around 1:11 and 1:4 for the x_t relative to the corresponding y_t case when one considers 24 and 300 lags respectively.

The results clearly indicate some degree of serial correlation in the levels but with substantial and significant serial correlation in the squares. In particular, for the forex Dollar/Pound and the stock indices the degree of dependence in the squares appears particularly strong, given the high significance of the portmanteau statistic up to the 300th lag. For all the series the kurtosis is much greater than that for Gaussian data, ranging between 5 and 8 in particular for the S & P 500 return index series which also displays the greatest skewness.

In Figures 1.3 and 1.4 we present the correlograms up to lag 300 for the forex data and the return indexes data respectively both in the levels and the squares. From the plots we immediately recognize the striking difference in the dependence structure for the raw and the squared data, the latter being characterized by an approximate hyperbolic decay in the autocovariance function as the lag increases.

1.1.2 Economic interpretations of volatility models.

At the present time it seems difficult to justify an equilibrium asset price with a time varying second moment, solution of a fully specified general equilibrium model. Indeed, the SV models were originally obtained simply as discrete approximations of the diffusion processes used in classic asset pricing models. In terms of ARCH-type models there are even less justifications given its discrete time structure. Bekaert (1996) proposed an asset pricing model in a dynamic programming framework where the equilibrium return has an ARCH behaviour in the squares, but the model, as it stands, appears of limited usefulness for the understanding of the source of time varying volatility. In fact, the result is obtained specifying the state variable as having an ARCH behaviour as well, thus making the model one of partial equilibrium.

Probably the most successful way of motivating changes in volatility is still to be found in the work of Clark (1973), subsequently refined by Tauchen and Pitts (1983), and Andersen (1996) among others. Very simply, this family of models states that price changes occur as a result of a random number of intra-daily price movements in turn caused by information arrivals. Hence in its simplest formulation the model puts

$$x_t = \sum_{i=1}^{n_t} x_{i,t}$$

with the i-th event on period t, $i = 1, \dots, n_t$ changing by $x_{i,t}$ the return on the same day. Usually the model sets the $x_{i,t}$ and the n_t to be independent and identically distributed and independent from each other. The specification of the $\{n_t\}$ process will drive the asymptotic distribution of the x_t.

Standard results apply (Clark 1973, Theorem 2) when the process $\{n_t\}$ is made of a sequence of positive integral valued random variables such that

$$n_t/t \to_p 1, \quad \text{as } t \to \infty,$$

where $\rightarrow_p$ denotes convergence in probability. In other words, if n_t has small variation around t for large t, then x_t approaches the normal distribution as $t \rightarrow \infty$.

On the other hand, assuming that n_t has appreciable variance around t even for large t, for instance when

$$n_t/t \rightarrow_d Z, \quad \text{as } t \rightarrow \infty,$$

$\rightarrow_d$ denoting convergence in distribution, where Z is a non negative random variable with $E(Z) = 1$, independent of the $x_{i,t}$, then (Clark 1973, Theorem 3) $x_t/t^{1/2}$ converges in distribution to a random variable u with density

$$pdf(u) = \frac{1}{(2\pi Z)^{1/2}} exp(-u^2/(2Z)),$$

which is a function of Z and thus not a Gaussian density function. The $\{x_t\}$ is an example of a subordinate process (Clark 1973).

More insightful information can be gained from models which consider the joint dynamic of asset return and trading volume (Tauchen and Pitts 1983) (Gallant, Hsieh, and Tauchen 1991), with both variables being mixture of i.i.d. variates driven by the same mixing variable n_t.

This is a promising field of research yet underdeveloped in order to obtain a full general equilibrium model, perhaps on the line of Campbell, Grossman, and Wang (1993) who explained the joint behaviour of return autocorrelation (in the levels) and trading volume dynamics.

1.1.3 Classifying models of changing volatility

Let us assume that the raw series x_t has conditional mean equal to zero and conditional variance equal to σ_t^2, conditioning on the information set I_t so that

$$x_t|I_t \sim N(0, \sigma_t^2). \tag{1.1}$$

Volatility models differ by the way in which the information set I_t, driving the conditional variance, is specified (Shephard 1996) (Ghysels, Harvey, and Renault 1995).

Two important classes can be drawn with this respect. Volatility models of the first class define I_t as belonging to the sigma-field generated by the past realizations of x_t, viz. $X_{t-1} = \{x_s | s \leq t-1\}$. Most of these kind of models are known as ARCH-type models since the seminal work of Engle (1982) where the conditional variance is a linear function of a finite number p of past squared observations $y_t = x_t^2$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p}. \tag{1.2}$$

Therefore we can reinterpret (1.1) as giving explicitly the one-step-ahead forecast density

$$x_t | X_{t-1} \sim N(0, \sigma_t^2). \tag{1.3}$$

The important thing to notice is that today's conditional variance is completely determined once we know past observations or, in other words, it is said to be observed.

In the second important class of models, I_t is set to be a function of some latent process, say l_t. These kinds of models are known as Stochastic Volatility (SV) models. The dynamic for the logarithm of the variance is defined as a function of the unobservable process l_t. A simple example (Taylor 1986) is :

$$x_t | l_t \sim N(0, exp(l_t)), \;\; l_t = \beta_0 + \beta_1 l_{t-1} + \eta_t, \;\; \eta_t \sim NID(0, \sigma_\eta^2). \tag{1.4}$$

Differently from the ARCH-type models, one-step-ahead forecast densities do not in general have a closed analytic form due to the latent structure of the (logged) variance.

As will be clear in the following sections, this difference between the two classes will entail a great deal of difference in deriving the statistical properties of the process, and, more important, in terms of estimation and testing. In general, the lack of a closed-form expression for the one-step-ahead density in the SV-type models implies that simple evaluation of the likelihood is not straightforward and thus is likelihood-based estimation. On the other hand, the independence between the shock in the latent process and the shock in the observable process yields many simplifications in the derivation and manipulation of the statistical properties of the process.

1.1.4 ARCH-type models

Consider the simplest ARCH(1) model given by:

$$x_t = \epsilon_t \sigma_t, \; \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}, \; \epsilon_t \sim NID(0,1), \tag{1.5}$$

It follows that x_t are martingale difference but that the y_t are autocorrelated and have AR(1) autocorrelation in the sense that

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + v_t, \tag{1.6}$$

with $v_t = \sigma_t^2(\epsilon_t^2 - 1)$ and so a martingale difference. The coefficients α_0, α_1 need to be non-negative to ensure non-negativity for the conditional variance. There is no upper bound on α_0 whereas α_1 cannot be too large in order to ensure covariance stationarity of y_t namely $3\alpha_1^2 < 1$. From simple calculations, given the linear autoregressive scheme (1.6), it follows that the autocorrelation function decays exponentially as $\rho_y(s) = \alpha_1^s$, $s = 0, 1, \ldots$, $\rho_w(u)$, $u = 0, \pm 1, \pm 2, \ldots$ denoting the autocorrelation function for any stationary process $\{w_t\}$. Moreover, the unconditional distribution of x_t is not normal and displays a kurtosis greater than 3.

The basic ARCH model has been modified in different ways, specifying different functional forms for σ_t^2, the more remarkable ones given by the GARCH model of Bollerslev (1986) and the exponential GARCH of Nelson (1991). Further developments assume different distributional assumption for ϵ_t (Bollerslev 1987) (Nelson 1991) so obtaining a more leptokurtic unconditional distribution or including asymmetric effects. Surveys on ARCH-type of models are given in Bollerlsev, Chou, and Kroner (1992) and in Bollerslev, Engle, and Nelson (1995) among others.

In terms of statistical inference, efficient estimation is allowed for ARCH-types of models given the ease by which the true likelihood can be built from the one-step-ahead densities via a prediction decomposition. From (1.3) ignoring constant terms

$$\begin{aligned} log\, pdf(x_1, \ldots, x_T | x_0, \theta) &= \sum_{t=1}^{T} log\, pdf(x_t | X_{t-1}, \theta) \\ &= -1/2 \sum_{t=1}^{T} log\, \sigma_t^2 - 1/2 \sum_{t=1}^{T} y_t / \sigma_t^2, \end{aligned}$$

where θ denotes the parameter vector which indexes the model (for the ARCH(1) model $\theta = (\alpha_0, \alpha_1)'$). Obviously one can compute the likelihood given other distributional assumptions besides the normal. Scores are also very easy to compute. Notice how the evaluation of the likelihood depends on some prior observations.

Although ARCH-type models have been routinely used and estimated via a Maximum Likelihood Estimator (MLE) approach, the asymptotic properties of the MLE have been established just for the ARCH model (under normality) by Weiss (1986)[1] and for the GARCH(1,1) by Lumsdaine (1996) and Lee and Hansen (1994) who considered the more general case of a Pseudo MLE with a Gaussian likelihood. For all the other ARCH-type models the asymp-

[1]Weiss (1986) also established the asymptotic properties of the least squares estimator for the ARCH(p) model, obviously less efficient than the ML.

totic properties of the MLE have been at most conjectured whereas the small-sample ones are routinely estimated through Montecarlo experiments.

This lack of rigorous asymptotic theory mirrors the extreme difficulties encountered in establishing statistical properties of the MLE due to the intrinsic nonlinearity of this model, a property shared with all volatility models.

1.1.5 SV models

The simplest example is given by the log-normal SV model:

$$x_t = \epsilon_t exp(l_t/2),\ l_t = \beta_0 + \beta_1 l_{t-1} + \eta_t, \tag{1.7}$$

where $\epsilon_t \sim NID(0,1)$, $\eta_t \sim NID(0,\sigma_\eta)$ and are independent of each other.

As with the ARCH model, the raw series x_t is martingale difference and is white noise if $\mid \beta_1 \mid < 1$. There is no need of non-negativity constraints on the parameters because this is a model for the logged conditional variance. In fact we can write

$$log\, y_t = l_t + log\, \epsilon_t^2, \tag{1.8}$$

with l_t as in (1.7). This implies that $log\ y_t$ has an ARMA(1,1) structure with $\rho_{log\, y}(s) = \beta_1^s/(1+4.93/\sigma_l^2)$ where $\sigma_l^2 = var(l_t)$.

If l_t is stationary all the moments of x_t exist, with odd ones trivially equal to zero. It can also be shown (Taylor 1986) that $\rho_y(s) \sim \beta_1^s$, implying, differently from the ARCH case, that the autocorrelation for the squared variate y_t can be negative, again an ARMA(1,1) feature. Notice also that by construction there are no constraints of non-negativity on the parameters in the variance equation. Finally this SV model has fatter tails than the normal distribution.

In terms of estimation, the latent structure of the SV models brings a great deal of difficulties due to the impossibility of computing analytically the likelihood function. Simple ways of avoiding this problem have been either to consider a GMM approach (cf. Chesney and Scott (1989), Duffie and Singleton (1993)), or to use the state-space/Kalman filter device (Harvey, Ruiz, and Shephard 1994) based on a Gaussian quasi likelihood. Even if the QMLE seems more efficient than the GMM estimator (Ruiz 1994) it seems that the direct way based on the maximizing the joint density of $(x_1, \ldots, x_T)$ is preferable, T being the sample size. In this case one needs to integrate out the latent process l_t as in

$$pdf(x_1, \ldots, x_T) = \int pdf(x_1, \ldots, x_T|l)g(l)dl,$$

with $l = (l_1, \ldots, l_T)'$ and $g(.)$ denoting the joint density function for l.

This integral does not have a closed form and is also a multiple integral of dimension $T \times dim(l_t)$. Several methods can be used in order to perform this integration as by Montecarlo integration or more recently by Markov chain Montecarlo integration (cf. Shephard (1996) for a survey). It is important to stress that one should establish whether the conditions for Central Limit Theorems to hold (Tierney 1994) are valid for the Markov chain sampler used to get the joint density of $(x_1, \ldots, x_T)$, thus guaranteeing the standard asymptotic properties for the MLE. This aspect appears to be neglected in the SV literature.

1.2 Memory considerations on nonlinear modelling

Linear time series models, in the sense of a linear filter in a white noise sequence with square summable coefficients[2], are appropriate in the case of Gaussian series. They also model the second-moment structure of non-Gaussian series, in the sense that Gaussian inference procedures can be asymptotically justified in many non-Gaussian environments (see e.g. Hannan (1973)). It is also possible to estimate efficiently linear models with non-Gaussian innovations when the latter have a parametric or even a nonparametric distribution. However, linear models can afford only a description of second moment properties, or ones of the higher moment properties that is heavily dependent on the linear dynamics.

Despite the prevalence of linear modelling, methods of time series analysis motivated by possible nonlinearity have a long history. Two rival approaches to nonparametric modelling, using smoothed estimation, focus on higher-order spectra (e.g. Brillinger and Rosenblatt (1966)) on one hand, and probability densities and conditional expectations on the other (Robinson 1983). The properties and application of a number of specific members of the limitless range

[2]More formally, if we replace the white noise sequence assumption with an i.i.d. sequence assumption our definition coincides with the Hannan (1970, p.209) definition of a 'generalized linear process', which allows for long memory.

of nonlinear stochastic processes have been explored, such as nonlinear functions of Gaussian processes, nonlinear autoregressive (Tong 1991) or moving average processes (Robinson 1977), and bilinear processes (Subba Rao and Gabr 1984), especially in a parametric context.

Before employing a specific nonlinear approach as with linear analysis, memory properties have to be borne in mind.

It is well known that most of the Law of Large Numbers and Central Limit Theorems used in modern time series analysis once developed for the i.i.d. case do hold under weaker conditions, in particular allowing a certain degree of serial dependence formalized as a more or less strong form of mixing condition (Rosenblatt 1956) (Ibragimov 1959). We can define the family of processes satisfying such a scheme of dependence as short memory or weak dependent.

For stationary Gaussian processes the autocorrelation function, equivalently the spectral density (if it exists), unambiguously describes the memory properties. Autoregressive moving average processes can be said to have short memory in the sense that autocorrelations decay exponentially and the spectral density is analytic. It is mathematically convenient to adopt a much broader definition of short memory, such as one where the autocovariances are absolutely summable, or where the spectral density is bound. Long memory processes which violate these requirements have increasingly been of interest. These typically imply that autocorrelations decay hyperbolically with exponent exceeding -1, or that the spectral density increases as a power law in the neighbourhood of a singularity, usually at zero frequency. For a recent survey, see Robinson (1994c).

For non-Gaussian processes there are no such global concepts of short and long memory. For linear transformations of non-Gaussian white noise it seems natural to define short memory as absolute summability of the Wold representation weights (implying summability of the autocovariances). For nonlinear processes we may be unable to determine analytically the rate of decay of autocorrelations, in which case we cannot even say whether there is short or long memory in a second order sense. On the other hand many nonlinear processes that have been studied appear to have short memory in the sense that they seem likely to satisfy some more or less strong form of mixing condition. For example many (finite order) nonlinear autoregressive scheme are of this type, in view of the Markov property. On the other hand, there has been little effort to combine nonlinearity and long memory in modelling with such exceptions as Rosenblatt (1987), Taqqu (1987) and Robinson (1991c).

In the following chapters we will introduce a class of nonlinear non-Gaussian processes for which the memory properties are rigorously defined in terms of second order quantities. Before considering this nonlinear approach we will show the way in which the coefficients of linear models determine the memory properties of the squared process, so that long memory in the levels is a precondition for long memory in the squares but does not guarantee it. This of course implies the need for considering an explicitly nonlinear model.

Following Robinson and Zaffaroni (1996a), consider a fourth-order stationary process:

$$v_t = \sum_{j=-\infty}^{\infty} \tau_j \eta_{t-j} \,, \tag{1.9}$$

where:

$$\sum_{j=-\infty}^{\infty} \tau_j^2 < \infty \,, \tag{1.10}$$

and the η_t satisfy:

$$E(\eta_t) = 0 \,, t = 0, \pm 1, \ldots \tag{1.11}$$

$$E(\eta_t \eta_s) = \begin{cases} \sigma_\eta^2 , & s = t \\ 0 , & s \neq t \end{cases} \tag{1.12}$$

$$E(\eta_s \eta_t \eta_u) = 0 \,, \; \forall s, t, u \tag{1.13}$$

$$E(\eta_s \eta_t \eta_u \eta_v) = \begin{cases} 3\sigma_\eta^4 + \kappa & s = t = v = u \\ \sigma_\eta^4 & \begin{cases} s = t \neq v = u \\ s = u \neq t = v \\ s = v \neq t = u \end{cases} \\ 0 & \textit{otherwise.} \end{cases} \tag{1.14}$$

The η_t behave like an independent identically distributed sequence up to fourth moments, κ denoting the fourth cumulant which is zero in the Gaussian case. On the other hand (1.9)-(1.12) are implied if v_t is a covariance stationary process with absolutely continuous spectral distribution function, and (1.9)-(1.12) with $\tau_j = 0$, $j < 0$, are implied if v_t is purely non deterministic.

Henceforth for any stationary process $\{z_t\}$ $\gamma_z(u)$, $u = 0, \pm 1, \ldots$ and $f_z(\lambda)$, $-\pi \leq \lambda < \pi$ defines respectively its autocovariance function and (when it exists) its power spectrum.

The lag-j autocovariance of v_t is given by

$$\gamma_v(j) = \sigma_\eta^2 \sum_{i=-\infty}^{\infty} \tau_i \tau_{i+j} \ , \ j = 0, \pm 1, \dots, \tag{1.15}$$

and the spectral density $f_v(.)$ of v_t (if it exists) satisfies

$$\gamma_v(j) = \int_{-\pi}^{\pi} f_v(\lambda) cos(j\lambda) d\lambda \ , \ j = 0, \pm 1, \dots . \tag{1.16}$$

The v_t process is uncorrelated if and only if $\tau_j = 0, \ \forall j \neq 0$. On the other hand the mild ergodicity condition

$$\gamma_v(j) \to 0 \ , \text{as } j \to \infty \tag{1.17}$$

is implied by integrability of $f_v(\lambda)$, in view of the Riemann-Lebesgue theorem. In particular, it is possible to contemplate $\gamma_v(j)$ which decay to zero at a variety of rates, and to choose the τ_j accordingly.

One choice of particular interest is

$$\gamma_v(j) \sim C j^{2d-1}, \ \text{as } j \to \infty, \tag{1.18}$$

for

$$| C | < \infty, \ \ 0 < d < 1/2, \tag{1.19}$$

where $\sim$ expresses asymptotic equivalence. v_t is then a long memory process in the sense that while (1.17) holds,

$$\sum_{j=0}^{\infty} |\gamma_v(j)| = \infty. \tag{1.20}$$

In the frequency domain we can consider

$$f_v(\lambda) \sim c\lambda^{-2d}, \ \text{as } \lambda \to 0^+ \tag{1.21}$$

for

$$0 < c < \infty, \ \ 0 < d < 1/2, \tag{1.22}$$

so that v_t has long memory in the sense that $f_v(\lambda)$ is unbounded.

The properties (1.18) and (1.21) do not necessarily co-exist, but do so under suitable additional conditions (Yong 1974), and in particular when v_t is a stationary and invertible fractionally integrated autoregressive moving average sequence (ARFIMA), given by

$$(1 - L)^d a(L) v_t = b(L) \eta_t \ , \ t = 0, \pm 1, \dots, \tag{1.23}$$

where L is the lag operator and $a(.)$ and $b(.)$ are finite order polynomials, all of whose zeros are outside the unit circle in the complex plane. In case of (1.23) we have

$$\tau_j \sim K\, j^{d-1} \tag{1.24}$$

for some $K > 0$, and indeed more generally (1.24) implies (1.18).

From

$$\gamma_{v^2}(j) = cov(v_t^2, v_{t+j}^2), \tag{1.25}$$

under (1.9)-(1.14) it is easily shown that

$$\gamma_{v^2}(j) = 2\sigma_\eta^4 \Big(\sum_{i=-\infty}^{\infty} \tau_i \tau_{i+j}\Big)^2 + \kappa \sum_{i=-\infty}^{\infty} \tau_i^2 \tau_{i+j}^2 , \tag{1.26}$$

so that the autocovariance properties of v_t^2 reflect κ as well as the τ_j. When $\kappa = 0$, as in the Gaussian case, we have

$$\gamma_{v^2}(j) = 2(\gamma_v(j))^2 \tag{1.27}$$

so that under (1.18)

$$\gamma_{v^2}(j) \sim 2C^2 j^{4d-2}. \tag{1.28}$$

Then, within the range $0 < d < 1/2$, we have long memory in v_t^2 if and only if there is sufficiently strong long memory in v_t , that is if $1/4 < d < 1/2$. For $0 < d < 1/4$, v_t^2 has short memory and the case $d = 1/4$ is indeterminate.

Similar considerations hold more generally. If $\kappa > 0$ the second term in (1.26) is non-negative so that under (1.24)

$$\gamma_{v^2}(j) \geq 2C^2 j^{4d-2} \tag{1.29}$$

for all sufficiently large j, providing a lower bound to the memory of v_t^2. On the other hand, in general,

$$\sum_{i=-\infty}^{\infty} \tau_i^2 \tau_{i+j}^2 = O(\max_{|i| \geq j/2} \tau_i^2) , \tag{1.30}$$

applying (1.10). Under (1.24)

$$\sum_{i=-\infty}^{\infty} \tau_i^2 \tau_{i+j}^2 = O(j^{2d-2}) \tag{1.31}$$

so that the second term in (1.26) is dominated by the first, and the Gaussian long-run behaviour is manifested for all κ.

The above result can be compared with one for nonlinear functions of Gaussian process of Taqqu (1975). Consider a function $g(x_t)$ where x_t is stationary Gaussian and $Eg(x_t)^2 < \infty$. Then we can consider the Hermite expansion

$$g(x) = \beta_0 + \sum_{j=1}^{\infty} \frac{\beta_j}{j!} H_j(x) \tag{1.32}$$

where $H_j(x)$ is the jth Hermite polynomial. Taqqu (1975) defined the Hermite rank m of $g(.)$ by

$$m = max\{i : \beta_j = 0, j < i\}. \tag{1.33}$$

Further he showed that under (1.18), (1.33), $g(x_t)$ has lag-j autocorrelation decaying like $j^{m(2d-1)}$. Thus $g(x_t)$ has long memory if and only if $d > 1/2 - 1/2m$. Clearly $m = 2$ is the case $g(x_t) = x_t^2$ so we have generalized the result for the square function in the Gaussian case of (1.9).

Taqqu (1975) showed that the limiting distribution of suitably normalized partial sums $\sum_{t=1}^{n} g(x_t)$ is governed by the Hermite rank m. In particular, if $m = 1$ the limiting distribution is Gaussian, whereas if $m > 1$ it is a nonstandard functional of fractional Brownian motion, varying with m. These results of Taqqu have been extended to an important nonparametric technique for analyzing possibly non Gaussian time series, smoothed probability density estimation.

Let $p(y)$ be the probability density function of $y_t = g(x_t)$ and estimate it by

$$\hat{p}(z) = \frac{1}{nh} \sum_{t=1}^{n} k(\frac{y_t - z}{h}) \tag{1.34}$$

where $h = h_n$ is a bandwidth sequence, tending to zero more slowly than $n \to \infty$, and $k(.)$ integrates to 1. In the case of short memory y_t , such as strongly mixing ones with a suitably decaying mixing number and under other regularity conditions, for fixed distinct points z_i , $i = 1, \ldots, r$ the $\hat{p}(z_i)$ are known to be asymptotically normally and independently distributed (see Robinson (1983)). The limiting distributional behaviour of the $\hat{p}(z_i)$ in the long memory case was first considered by Robinson (1991b), observing that $k(\frac{y_t - z}{h})$ is an instantaneous nonlinear function of x_t if y_t is, and employing Taqqu (1975) result for partial sums.

Robinson (1991b) result was extended by Cheng and Robinson (1994). If $m = 1$, then the $\hat{p}(z_i)$ can be asymptotically normal but with a singular limiting covariance matrix (of rank 1 for all r) in contrast to the short memory case, while for $m > 1$ the limiting distribution is nonstandard.

While probability density estimates contain information and can be extended to describe aspects of the joint distribution of finitely many y_t, they cannot satisfactorily describe the intrinsically infinite-dimensional long-memory nonlinear dynamics, and so we turn to the estimation of explicitly nonlinear models for long memory.

1.3 Further developments of ARCH and SV models: the long memory cases.

Both the ARCH-type and the SV models can be shown to satisfy some form of more or less strong form of mixing condition for the squared process $y_t = x_t^2$, expressed by the exponential decay of $\gamma_y(l)$, the raw process x_t behaving usually like a martingale difference in standard formulations. But as shown in section 1.1.1 the sample autocorrelation for squared returns dies away much more slowly than a strong-mixing process would allow for. Long memory processes seem capable of mimicking the autocorrelation behaviour of squared returns when the theoretical autocorrelations decay so slowly that they are not absolutely summable, or when the spectral density is unbounded at the origin.

Robinson (1991a) first proposed a generalized ARCH assuming

$$E(x_t|X_{t-1}) = 0, \; a.s. \tag{1.35}$$

where $\sigma_t^2 = var(x_t|X_{t-1})$ is given by

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^{\infty} \psi_j (y_{t-j} - \sigma^2), \text{ a.s.} \tag{1.36}$$

for some $\sigma^2 > 0$ with the weights ψ_i, $i = 1, 2, \ldots$ chosen in such a way to impart long memory to y_t.

Writing $\psi(L) = 1 - \sum_{i=1}^{\infty} \psi_i L^i$ we can express (1.36) as

$$\psi(L) y_t = \phi + \nu_t, \tag{1.37}$$

with $\phi = \psi(1)\sigma^2$ and the $\nu_t = y_t - \sigma_t^2$ are a sequence of martingale difference $E(\nu_t|X_{t-1}) = 0$.

If follows that y_t has spectral density

$$f_y(\lambda) \propto |\psi(e^{i\lambda})|^{-2}, \; -\pi \leq \lambda < \pi. \tag{1.38}$$

Thus if

$$|\psi(e^{i\lambda})| \sim c\lambda^d, \text{ as } \lambda \to 0^+, \tag{1.39}$$

for $c \geq 0$ and $0 < d < 1/2$, it follows that y_t has long memory in the sense indicated in (1.21).

Hence setting

$$x_t = \epsilon_t \sigma_t, \tag{1.40}$$

where

$$\epsilon_t | X_{t-1} \, i.i.d.(0,1), \tag{1.41}$$

we obtain a data generating process with the desired properties.

A special case is obtained by setting

$$\psi(L) = (1-L)^d \frac{a(L)}{b(L)}, \tag{1.42}$$

with $a(.)$ and $b(.)$ being finite order polynomials whose zeros are outside the unit circle in the complex plane. This case was indeed considered by Baillie, Bollerslev, and Mikkelsen (1996) but with ϕ a free parameter being otherwise equal to zero from (1.42). Notice that in this case for any $d > 0$, y_t does not have a finite unconditional mean which means that the x_t have unbounded variance. Given the nature of the model, Baillie, Bollerslev, and Mikkelsen (1996) proposed PML estimation. Of course, one has to evaluate the likelihood conditional on some prior observations. Lumsdaine (1996) established that for the GARCH(1,1) this influence is asymptotically negligible. In this long memory framework we conjecture that the influence might still be asymptotically negligible but it would probably require a much greater time in order to become so depending on the degree of memory of the process. Moreover, the asymptotic properties of the PMLE have not yet been established due to the great complexity that the combination of nonlinearity and long memory imposes. Baillie, Bollerslev, and Mikkelsen (1996) report some Montecarlo experiments.

Harvey (1993) introduced a SV model with long memory. He considered (1.7) and replaced the difference equation in l_t with

$$l_t = (1-L)^{-d}\eta_t, \; 0 \leq d \leq 1. \tag{1.43}$$

A Gaussian PMLE was proposed, but again the asymptotic properties of the estimator are unknown. Breidt, Crato, and deLima P. (1993) make an attempt to obtain consistency of the PMLE.

The issue of making an (asymptotically) correct inference is particularly important for processes displaying long memory because, differently from the short memory case, a variety of peculiar results are obtained. For instance, statistics based on samples of consecutive observations may not be normal depending on the degree of memory of the process.

Chapter 2

Long memory moving average models: the 'one-shock' case

We introduce a new class of volatility models defined as nonlinear moving averages models in terms of an unobservable sequence. We establish the second order properties of the model and its capability to capture the most important aspects of the distribution of financial asset returns considered in Chapter 1.

2.1 Definitions and assumptions

We introduce a new class of volatility models within the family of observation driven models. A very wide class of X_{t-1}-measurable functions σ_t^2 in (1.40) can give rise under (1.41) to x_t that are uncorrelated but not independent, and a number of these can imply long memory in $y_t = x_t^2$. The model (1.35) can be thought of as an (infinite-order) nonlinear AR,

$$x_t/\sigma_t = \epsilon_t. \tag{2.1}$$

Alternatively we can think in terms of a nonlinear MA

$$x_t = \epsilon_t h_{t-1}, \tag{2.2}$$

where $h_{t-1} = h(\epsilon_{t-1}, \epsilon_{t-2}, \ldots)$. We can think of (2.2) as inverse to (2.1) if there is a well defined solution of (2.1) for x_t of the form $x_t = g(\epsilon_t, \epsilon_{t-1}, \ldots)$, so that

$$\sigma_t = \sigma(x_{t-1}, x_{t-2}, \ldots) = \sigma(g(\epsilon_{t-1}, \epsilon_{t-2}, \ldots), g(\epsilon_{t-2}, \epsilon_{t-3}, \ldots), \ldots) = h_{t-1}.$$

Commencing from (2.2) rather than (2.1) means that we have to choose a functional form for the h_{t-1}.

Nonlinear MA have a long history but they have not really been popular in practical applications, probably due to the difficulty in establishing the invertibility condition (Granger and Andersen 1978). They can be viewed as made of the first terms in the Volterra expansion (Volterra 1931) of a general nonlinear process (Wiener 1958). Robinson (1977) considered an explicit case of nonlinear MA by extending the linear MA(1) to

$$x_t = \epsilon_t + b\epsilon_{t-1} + c\epsilon_t\epsilon_{t-1}. \tag{2.3}$$

Robinson (1977) proposed estimates of b and c and established consistency and asymptotic normality, giving also limit theorems applicable to more general nonlinear MA models than (2.3). When $b = 0$, the x_t in (2.3) are uncorrelated but, when $c \neq 0$, they are not independent (in particular y_t is autocorrelated) and we have a special case of (2.2) with

$$h_{t-1} = 1 + c\epsilon_{t-1}. \tag{2.4}$$

However no instantaneous functions of such x_t have long memory, x_t being m-dependent.

We can impart long memory to such instantaneous transformations as y_t, by allowing for infinitely many lagged ϵ_t and choosing their coefficients so to allow a slow enough rate of decay. The algebraically simplest way of doing so is to extend the linear form of (2.4) to

$$h_{t-1} = \rho + \sum_{j=1}^{\infty} \alpha_j \epsilon_{t-j}, \tag{2.5}$$

Hence the model that we propose is obtained extending further (2.2) allowing for a non zero mean

$$x_t = \mu + \epsilon_t h_{t-1}, \tag{2.6}$$

with h_t given by (2.5).

In terms of the coefficients α_i $i = 1, 2, \ldots$ the following regularity conditions are assumed. Henceforth K denotes any arbitrary constant (not necessarily the same one).

Assumptions A

$A_1 \qquad \alpha_j \sim Kj^{d-1},\, 0 < d < 1/2,\, 0 <\mid K \mid < \infty,\ \ as\ j \to \infty$

$A_2 \qquad \mid \alpha_j - \alpha_{j+1} \mid \leq K \dfrac{\mid \alpha_j \mid}{j},\ \forall j > J,\ some\ J < \infty, 0 < K < \infty.$

We also introduce the following

Assumption B

The unobservable process $\{\epsilon_t\}$ *is i.i.d. with*

$$\begin{aligned} E(\epsilon_t) &= 0. \\ E(\epsilon_t^2) &= \sigma^2, 0 < \sigma^2 < \infty, \\ E(\epsilon_t^3) &= \mu_3, \mid \mu_3 \mid < \infty, \\ cum_4(\epsilon_t, \epsilon_t, \epsilon_t, \epsilon_t) &= \kappa_4 < \infty. \end{aligned}$$

Obviously the fourth moment of ϵ_t, $\mu_4 = E(\epsilon_t^4)$ is obtained by standard arguments as $\mu_4 = \kappa_4 + 3\sigma^4$.

Let $\mathcal{F}_t$ be the $\sigma-$ field generated by the realizations $\{\epsilon_s; s \leq t\}$. The two sigma-fields $\mathcal{F}_t$ and X_t are formally equivalent when the conditions for the invertibility of the model hold but we will not address this issue. Henceforth we will always consider the probabilistic space defined by $\mathcal{F}_t$.

Henceforth $\delta_{ab}(l;\theta) = \sum_{i=1}^{\infty} a_i(\theta) b_{i+l}(\theta)$, $l > 0$ integer for arbitrary $a_i(\theta), b_i(\theta), i = 1, 2, \ldots$ both functions of some vector θ, including the constant function case $a_i(\theta) = a_i$, $b_i(\theta) = b_i$.

It follows easely that x_t is weakly stationary under B and A_1 given that under these assumptions $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$. Indeed under the same assumptions the process is also strict stationary as it is formally shown below.

Let us now consider the statistical properties of the x_t and y_t processes.

2.2 Statistical properties

2.2.1 The autocovariance functions for x_t and y_t

Theorem 1 *Under Assumption B the x_t are a martingale difference sequence and thus for any $l \neq 0$*

$$\gamma_x(l) = 0. \tag{2.7}$$

Proof: Let us take without loss of generality $l > 0$. In fact if $l < 0$ we will set $t + l = t'$ and $t = t' + \mid l \mid$ and thus we will consider $cov(x_{t'}, x_{t'+|l|})$.

Then

$$\begin{aligned} E(x_t - \mu | \mathcal{F}_{t-1}) &= E(\epsilon_t (\rho + \sum_{j=1}^{\infty} \alpha_j \epsilon_{t-j} | \mathcal{F}_{t-1}) & (2.8) \\ &= (\rho \sum_{j=1}^{\infty} \alpha_j \epsilon_{t-j}) E(\epsilon_t | \mathcal{F}_{t-1}) = 0, & (2.9) \end{aligned}$$

so that in particular one obtains

$$\gamma_x(l) = E\left((x_t - \mu)(x_{t+l} - \mu)\right) = E\left((x_t - \mu) E(x_{t+l} - \mu | \mathcal{F}_t)\right) = 0.$$

□

The following definition and result will be used.

Definition 1 *(Leonov and Shiryaev 1959)*

Consider a (not necessarily rectangular) two-way table

$$\begin{matrix} (1,1) & \ldots & (1,J_1) \\ \vdots & & \vdots \\ (I,1) & \ldots & (I,J_I) \end{matrix} \tag{2.10}$$

and a partition $P_1 \cup P_2 \cup \ldots \cup P_M$ of its entries. We shall say that sets $P_{m'}, P_{m''}$, of the partition, **hook** *if there exist $(i_1, j_1) \in P_{m'}$ and $(i_2, j_2) \in P_{m''}$ such that $i_1 = i_2$. We shall say that the sets*

$P_{m'}, P_{m''}$ **communicate** *if there exists a sequence of sets* $P_{m_1} = P_{m'}, P_{m_2}, \ldots, P_{m_N} = P_{m''}$ *such that* P_{m_n} *and* $P_{m_{n+1}}$ *hook for* $n = 1, 2, \ldots, N-1$. *A partition is said to be* **indecomposable** *if all sets communicate (cf. Brillinger (1975) for further results on indecomposable partitions).*

A very important result is given by the following proposition.

Proposition 1 *(Leonov and Shiryaev 1959)*

Consider a two-way array of random variables $X_{ij}; j = 1, \ldots, J_i; i = 1, \ldots, I$. *Consider the* I *random variables*

$$Y_i = \prod_{j=1}^{J_i} X_{ij}, \quad i = 1, \ldots, I.$$

The joint cumulant $cum_I(Y_1, \ldots, Y_I)$ *is then given by*

$$\sum_{\nu} cum(X_{ij};\, ij \in \nu_1) \ldots cum(X_{ij};\, ij \in \nu_p),$$

where the summation is over all indecomposable partitions $\nu = \nu_1 \cup \ldots \cup \nu_p$ *of the table (2.10).*

Theorem 2 *Under Assumption B, and (without loss of generality) for any integer* $l \neq 0$

$$\begin{aligned}\gamma_y(l) \quad &= (\kappa_4\sigma^4 + 2\sigma^8)\alpha_{|l|}^2\delta_{\alpha\alpha}(0) + \sigma^4\kappa_4\delta_{\alpha^2\alpha^2}(l) + 2\sigma^8\delta_{\alpha\alpha}^2(l) + 2\mu_3^2\alpha_{|l|}\delta_{\alpha^2\alpha}(l) + 4\rho^2\sigma^6\delta_{\alpha\alpha}(l) \\ &+2\sigma^2\mu\alpha_{|l|}\left[\rho\alpha_{|l|}\mu_3 + 2\rho^2\sigma^2 + 2\sigma^4\delta_{\alpha\alpha}(l)\right] + 2\rho^2\sigma^6\alpha_{|l|}^2 + 2\rho\sigma^4\mu_3\left[\delta_{\alpha\alpha^2}(l) + 2\alpha_{|l|}\delta_{\alpha\alpha}(l)\right] \\ &+2\rho^3\mu_3\sigma^2\alpha_{|l|} + 2\rho\sigma^4\mu_3\left[\alpha_{|l|}\delta_{\alpha\alpha}(0) + \delta_{\alpha^2\alpha}(l)\right].\end{aligned}$$

Proof: Let us consider the case $l > 0$.

Setting for any integer a, b, c such that $a > b \geq c$

$$h_{a,(b,c)} = \rho + w_{a,(b,c)},$$

$$w_{a,(b,c)} = \sum_{k=a-b}^{a-c} \alpha_k \epsilon_{t+a-k},$$

one obtains

$$
\begin{aligned}
&\gamma_y(l) = \\
&cov(\mu^2 + \epsilon_t^2 h_{0,(-1,-\infty)}^2 + 2\mu\epsilon_t h_{0,(-1,-\infty)}, \mu^2 + \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2 + 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}) \\
&= cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2) + cov(2\mu\epsilon_t h_{0,(-1,-\infty)}, 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}) \\
&+ cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}) + cov(2\mu\epsilon_t h_{0,(-1,-\infty)}, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2) \\
&= (i) + (ii) + (iii) + (iv),
\end{aligned}
$$

with

$$
\begin{aligned}
(i) = &\quad cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2), \\
(ii) = &\quad cov(2\mu\epsilon_t h_{0,(-1,-\infty)}, 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}), \\
(iii) = &\quad cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, 2\mu\epsilon_{t+l} h_{l,(l-1,-\infty)}), \\
(iv) = &\quad cov(2\mu\epsilon_t h_{0,(-1,-\infty)}, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2).
\end{aligned}
$$

By the independence assumption and $E(\epsilon_t) = 0$ the terms (ii) and (iii) are equal to zero. Thus

$$(i) = cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2 h_{l,(l-1,-\infty)}^2) = (i.1) + (i.2) + (i.3).$$

with

$$
\begin{aligned}
&(i.1) = \rho^2 cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2), \\
&(i.2) = cov(\epsilon_t^2(\rho^2 + w_{0,(-1,-\infty)}^2 + 2\rho w_{0,(-1,-\infty)}), \epsilon_{t+l}^2 w_{l,(l-1,-\infty)}^2), \\
&(i.3) = 2\rho cov(\epsilon_t^2 h_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2 w_{l,(l-1,-\infty)}).
\end{aligned}
$$

By Proposition 1

$$
\begin{aligned}
(i.1) &\quad = 0. \\
(i.3) &\quad = 2\rho^3\alpha_l\mu_3\sigma^2 + 2\rho\sigma^4\mu_3(\alpha_l\delta_{\alpha\alpha}(0) + \delta_{\alpha^2\alpha}(l)) + 4\rho^2\sigma^6\delta_{\alpha\alpha}(l), \\
(i.2) &\quad = (i.2.1) + (i.2.2) + (i.2.3),
\end{aligned}
$$

with

$$
\begin{aligned}
&(i.2.1) = 2\rho^2\sigma^6\alpha_l^2, \\
&(i.2.2) = 2\rho\sigma^4\mu_3(\delta_{\alpha\alpha^2}(l) + 2\alpha_l\delta_{\alpha\alpha}(l)), \\
&(i.2.3) = cov(\epsilon_t^2 w_{0,(-1,-\infty)}^2, \epsilon_{t+l}^2 w_{l,(l-1,-\infty)}^2).
\end{aligned}
$$

We will now analyze the third term of (i.2), namely (i.2.3), which requires more details. Thus

$$E(\epsilon_t w_{0,(-1,-\infty)})^2 = \sigma^4 \delta_{\alpha\alpha}(0)\,.$$

$$E((\epsilon_t w_{0,(-1,-\infty)})^2(\epsilon_{t+l} w_{l,(l-1,-\infty)})^2) = E((\epsilon_t \sum_{j=1}^{\infty} \alpha_j \epsilon_{t-j})^2\,(\epsilon_{t+l} \sum_{j=1}^{\infty} \alpha_j \epsilon_{t+l-j})^2)$$
$$= \sigma^2 E(\epsilon_t^2\, w_{0,(-1,-\infty)}^2 [w_{l,(l-1,1)} + w_{l,(0,0)} + w_{l,(-1,-\infty)}]^2)\,.$$

Then developing the square the last expression becomes

$$\sigma^2 E\left[\epsilon_t^2\, w_{0,(-1,-\infty)}^2\, \left(w_{l,(l-1,1)}^2 + w_{l,(0,0)}^2 + w_{l,(-1,-\infty)}^2 + 2w_{l,(l-1,1)} w_{l,(0,0)}\right.\right.$$
$$\left.\left. +\ 2w_{l,(l-1,1)} w_{l,(-1,-\infty)} + 2w_{l,(0,0)} w_{l,(-1,-\infty)}\right)\right]\,.$$

Then given that

$$E\left[\epsilon_t^2 w_{0,(-1,-\infty)}^2\, w_{l,(l-1,1)}(w_{l,(0,0)} + w_{l,(-1,-\infty)})\right] = 0,$$
$$E\left[\epsilon_t^2 w_{0,(-1,-\infty)}^2\, w_{l,(l-1,1)}^2\right] = \sigma^6 \delta_{\alpha\alpha}(0) \sum_{j=1}^{l-1} \alpha_j^2,$$
$$E\left[\epsilon_t^2\, w_{0,(-1,-\infty)}^2\, w_{l,(0,0)}^2\right] = \mu_4 \alpha_l^2 \sigma^2 \delta_{\alpha\alpha}(0),$$
$$E\left[\epsilon_t^2\, w_{0,(-1,-\infty)}^2\, w_{l,(0,0)}\, w_{l,(-1,-\infty)}\right] = \mu_3^2 \alpha_l \delta_{\alpha^2\alpha}(l),$$
$$E\left[w_{0,(-1,-\infty)}^2\, w_{l,(-1,-\infty)}^2\right]$$
$$= \mu_4\, \delta_{\alpha^2\alpha^2}(0) + \sigma^4 \sum_{j\neq i} \alpha_i^2 \alpha_{j+l}^2 + 2\sigma^4 \sum_{i\neq j} \alpha_j \alpha_{j+l} \alpha_i \alpha_{i+l},$$

one obtains that

$$(i.2.3) = \sigma^8 \delta_{\alpha\alpha}(0)\,(\sum_{j=1}^{l-1} \alpha_j^2) + \mu_4 \alpha_l^2 \sigma^4 \delta_{\alpha\alpha}(0) + 2\mu_3^2 \alpha_l \delta_{\alpha^2\alpha}(l)\ +\ \mu_4 \sigma^4\, \delta_{\alpha^2\alpha^2}(l)$$
$$+\sigma^8 \sum_{j\neq i} \alpha_i^2 \alpha_{j+l}^2 + 2\sigma^8 \sum_{i\neq j} \alpha_j \alpha_{j+l} \alpha_i \alpha_{i+l} - \sigma^8 \delta_{\alpha\alpha}^2(0)\,.$$

Now, considering that $\kappa_4 = \mu_4 - 3\sigma^4$ and substituting it in the former expressions we obtain

$$(i.2.3) = \sigma^8 \delta_{\alpha\alpha}(0) \left(\sum_{i=1}^{\infty} \alpha_i^2 - \alpha_l^2 - \sum_{i=1}^{\infty} \alpha_{i+l}^2\right) + (\kappa_4 + 3\sigma^4)\sigma^4 \left(\alpha_l^2 \delta_{\alpha\alpha}(0) + \delta_{\alpha^2\alpha^2}(0)\right)$$
$$+\sigma^8 \left(\delta_{\alpha\alpha}(0) \sum_{i=1}^{\infty} \alpha_{i+l}^2 - \delta_{\alpha^2\alpha^2}(l)\right) + 2\sigma^8 \left(\delta_{\alpha\alpha}^2(l) - \delta_{\alpha^2\alpha^2}(l)\right) + 2\alpha_l \mu_3 \delta_{\alpha^2\alpha}(l) - \sigma^8 \delta_{\alpha\alpha}^2(0).$$

Simple algebraic steps yield:

$$(i.2.3) = \kappa_4\sigma^4\left(\alpha_l^2\delta_{\alpha\alpha}(0) + \delta_{\alpha^2\alpha^2}(l)\right) + 2\alpha_l\mu_3^2\delta_{\alpha^2\alpha}(l) + 2\sigma^8\left(\alpha_l^2\delta_{\alpha\alpha}(0) + \delta_{\alpha\alpha}^2(l)\right).$$

Then considering (iv) one gets

$$\begin{aligned}
&(iv) = cov(2\mu\epsilon_t(\rho + w_{0,(-1,-\infty)}), \epsilon_{t+l}^2(\rho + w_{l,(l-1,-\infty)})^2) \\
&= (iv.1) + (iv.2) + (iv.3) + (iv.4) + (iv.5) + (iv.6), \\
&\text{with} \\
&(iv.1) = 2\mu cov(\epsilon_t\rho, \epsilon_{t+l}^2\rho^2), \\
&(iv.2) = 2\mu cov(\epsilon_t\rho, \epsilon_{t+l}^2 w_{l,(l-1,-\infty)}^2), \\
&(iv.3) = cov(\epsilon_t\rho, 2\rho\epsilon_{t+l}^2 w_{l,(l-1,-\infty)}), \\
&(iv.4) = cov(\epsilon_t w_{0,(-1,-\infty)}, \epsilon_{t+l}^2\rho^2), \\
&(iv.5) = cov(\epsilon_t w_{0,(-1,-\infty)}, \epsilon_{t+l}^2 w_{l,(l-1,-\infty)}^2), \\
&(iv.6) = cov(\epsilon_t w_{0,(-1,-\infty)}, 2\rho\epsilon_{t+l}^2 w_{l,(l-1,-\infty)}).
\end{aligned}$$

Direct evaluation gives $(iv.1) = (iv.4) = (iv.6) = 0$ so that evaluating the remaining terms one obtains

$$(iv) = (iv.2) + (iv.3) + (iv.5) = 2\mu\left[\rho\sigma^2\alpha_l^2\mu_3 + 2\rho^2\sigma^4\alpha_l + 2\sigma^4\alpha_l\delta_{\alpha\alpha}(l)\right].$$

□

The process x_t can be seen as an example of a process uncorrelated yet not independent.

Finally the variance of the y_t process is obtained as follows.

Theorem 3 *Under Assumption B,*

$$\begin{aligned}
var(y_t) &= \rho^4(\kappa_4 + 2\sigma^4) + 6\rho^2\sigma^2\delta_{\alpha\alpha}(0)(\kappa_4 + 2\sigma^4) + 4\rho\mu_3(3\sigma^4 + \kappa_4)\delta_{\alpha^2\alpha}(0) \\
&+\sigma^4\left(\kappa_4\delta_{\alpha^2\alpha^2}(0) + 2\sigma^4\delta_{\alpha\alpha}^2(0)\right) + (2\sigma^4 + \kappa_4)\left(3\sigma^4\delta_{\alpha\alpha}^2(0) + \kappa_4\delta_{\alpha^2\alpha^2}(0)\right) \\
&+4\mu^2\left(\rho^2\sigma^2 + \sigma^4\delta_{\alpha\alpha}(0)\right) + 4\mu\left(\rho^3\mu_3 + 3\rho\mu_3\sigma^2\delta_{\alpha\alpha}(0) + \mu_3^2\delta_{\alpha^2\alpha}(0)\right).
\end{aligned}$$

Proof:

$$
\begin{aligned}
& var(y_t) = \\
& = var((\epsilon_t h_{0,(-1,-\infty)})^2) + 4\mu^2 var(\epsilon_t h_{0,(-1,-\infty)}) + 4\mu cov(\epsilon_t h_{0,(-1,-\infty)}, (\epsilon_t h_{0,(-1,-\infty)})^2) \\
& = (i) + (ii) + (iii), \\
& \text{ where} \\
(i) \quad & = \rho^4(\kappa_4 + 2\sigma^4) + 6\rho^2\sigma^2\delta_{\alpha\alpha}(0)(\kappa_4 + 2\sigma^4) + 4\rho\mu_3(3\sigma^4 + \kappa_4)\delta_{\alpha\alpha^2}(0) \\
& + \sigma^4(\kappa_4\delta_{\alpha^2\alpha^2}(0) + 2\sigma^4\delta^2_{\alpha\alpha}(0)) + (2\sigma^4 + \kappa_4)(3\sigma^4\delta^2_{\alpha\alpha}(0) + \kappa_4\delta_{\alpha^2\alpha^2}(0)), \\
(ii) \quad & = 4\mu^2(\rho^2\sigma^2 + \sigma^4\delta_{\alpha\alpha}(0)), \\
(iii) \quad & = 4\mu(\rho^3\mu_3 + 3\rho\mu_3\sigma^2\delta_{\alpha\alpha}(0) + \mu_3^2\delta_{\alpha\alpha^2}(0)).
\end{aligned}
$$

□

Now, given Assumption A_1, it can be shown that the squared process displays long-memory. We will show this at first in terms of the long-lags behaviour of the autocovariances.

In order to do this, we will need to evaluate the asymptotic behaviour of the convolutions of the α_j given by Lemma 27 in Appendix A.

In the next section we will obtain the expression for the power spectrum for the y_t and then we will derive its behaviour near zero frequency.

Henceforth $\delta(u, v)$ denotes the Kronecker delta.

Theorem 4 *Under Assumptions A_1, B, as $l \to \infty$*

$$
\gamma_y(l) \sim \begin{cases} K\, l^{4d-2} & , \rho = 0, \quad 1/4 < d < 1/2 \\ K\, l^{2d-1} & , \rho \neq 0, \quad 0 < d < 1/2. \end{cases} \tag{2.11}
$$

Proof: By direct use of Lemma 27 and by A_1 in the expression for $\gamma_y(u)$ in Theorem 2.

□

Notice that when $\rho = 0$ the squared process displays long-memory in the sense indicated by (1.21) for $1/4 < d < 1/2$ only. Hence considering that the instantaneous transformation given by the squares implies that a great deal of the degree of dependence in the h_{t-1} component of the raw series is lost due to the transformation [1].

In fact, let us consider the linear process:

$$\xi_t = \epsilon_t + \sum_{i=1}^{\infty} \alpha_i \, \epsilon_{t-i} \, , \tag{2.12}$$

so that $h_{t-1} = \xi_t - \epsilon_t$. Its autocovariance function is equal to

$$\gamma_\xi(l) = \sigma^2(\alpha_l + \sum_{i=1}^{\infty} \alpha_i \, \alpha_{i+l}) = \sigma^2(\alpha_l + \delta_{\alpha\alpha}(l)) \tag{2.13}$$

Then, by using Lemma 27 as $l \to \infty$

$$\gamma_\xi(l) \sim K \, l^{2d-1} . \tag{2.14}$$

So, the linear process[2] displays long-memory, keeping stationarity, for

$$0 < d < 1/2 .$$

Thus we can see how the instantaneous filter expressed by the square operator change the pattern of dependence of the process, restricting the 'relevant' parameter space.

2.2.2 The power spectrum for x_t and y_t.

Let us replace Assumption B with

Assumption B'

$$\begin{aligned} \epsilon_t & \quad i.i.d. , & (2.15) \\ E(\epsilon_t) &= 0 , & (2.16) \end{aligned}$$

[1] This 'not-invariance' property of nonlinear operators with respect to the memory properties if the series is well known within the unit root literature (Granger 1991) (Corradi 1995).

[2] Which behaves, in terms of autocovariance function, as the Wold representation of an $ARFIMA(p, d, q)$.

$$
\begin{aligned}
E(\epsilon_t^2) &= \sigma^2, 0 < \sigma^2 < \infty\,, &(2.17)\\
E(\epsilon_t^3) &= \mu_3 = 0\,, &(2.18)\\
cum_4(\epsilon_t, \epsilon_t, \epsilon_t, \epsilon_t) &= \kappa_4 = 0\,. &(2.19)
\end{aligned}
$$

Then we obtain the following

Corollary 1 *Under Assumption B', for any $l \neq 0$*

$$
\begin{aligned}
\gamma_y(l) \quad &= 2\sigma^8\alpha_{|l|}^2\delta_{\alpha\alpha}(0) + 2\sigma^8\delta_{\alpha\alpha}^2(l) + 4\rho^2\sigma^6\delta_{\alpha\alpha}(l) \\
&+4\sigma^6\mu\alpha_{|l|}\left[\rho^2 + \sigma^2\delta_{\alpha\alpha}(l)\right] + 2\rho^2\sigma^6\alpha_{|l|}^2\,.
\end{aligned}
$$

With respect to the linear process ξ_t as defined above, let us define its power spectrum as $f_\xi(\omega)$, assuming that it exists, so that:

$$\gamma_\xi(l) = \int_{-\pi}^{\pi} f_\xi(\omega)e^{i\omega l}d\omega\,, \tag{2.20}$$

and by letting

$$\alpha_l = \frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha(\omega)e^{i\omega l}d\omega\,,\ \ l > 0\,, \tag{2.21}$$

with $\int_{-\pi}^{\pi} \alpha(\omega)d\omega = 2\pi$ we get

$$f_\xi(\omega) = \sigma^2\frac{|\alpha(\omega)|^2}{2\pi}\,. \tag{2.22}$$

Also let us define

$$\beta(\omega) = 2(\mathcal{R}e\ (\alpha(\omega))\ - 1)\,, \tag{2.23}$$

with $\mathcal{R}e\,(z)$ denoting the real part of any given complex number z.

Now we can derive the power spectrum respectively for the process x_t and for y_t. Obviously, for the former, given its nature as a martingale property, the power spectrum will be a constant function with respect to the frequencies in $[-\pi, \pi)$. In fact we obtain

Theorem 5 *Under Assumption B' the power spectrum for the x_t is*

$$f_x(\lambda) = \frac{\sigma^2}{2\pi}(\rho^2 + \sigma^2\delta_{\alpha\alpha}(0))\,,\ -\pi \leq \lambda < \pi. \tag{2.24}$$

Proof: From the uncorrelatedness of x_t, it follows that the only term appearing in the series expression for the power spectrum is the variance

$$f_x(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \gamma_x(u)\, e^{iu\lambda} = \frac{\sigma^4}{2\pi}\delta_{\alpha\alpha}(0) \sum_{u=-\infty}^{\infty} \delta(u,0)e^{iu\lambda}\,. \tag{2.25}$$

□

Concerning the power spectrum for the y_t process we obtain

Theorem 6 *Under Assumption B' for $-\pi \leq \omega < \pi$*

$$\begin{aligned} f_y(\omega) &= \frac{a_1}{4\pi^2}\int_{-\pi}^{\pi} \beta(\lambda)\beta(\omega-\lambda)d\lambda + \frac{a_2}{4\pi^4}\int_{-\pi}^{\pi} \mid \alpha(\lambda) \mid^2 \mid \alpha(\omega-\lambda) \mid^2 d\lambda \\ &+ \frac{a_3}{4\pi^2}\int_{-\pi}^{\pi} \beta(\lambda) \mid \alpha(\omega-\lambda) \mid^2 d\lambda + \frac{a_4}{2\pi} \mid \alpha(\omega) \mid^2 + \frac{a_5}{2\pi}\beta(\omega) + \frac{\nu_y}{2\pi}, \end{aligned}$$

with $a_1 = 2\sigma^6(\gamma_\xi(0) + \rho^2 - 2\mu)$, $a_2 = 2\sigma^8$, $a_3 = 4\sigma^6(\mu - \sigma^2)$, $a_4 = 4\sigma^6\rho^2$, $a_5 = 4\sigma^4\rho^2(\mu - \sigma^2)$, $\nu_y = 2\rho^4\sigma^4 + 8\rho^2\sigma^6\delta_{\alpha\alpha}(0) + 6\sigma^8\delta^2_{\alpha\alpha}(0) + 4\mu^2\sigma^2(\rho^2 + \sigma^2\delta_{\alpha\alpha}(0))$.

Proof: Expressing the autocovariance of y_t in terms of the autocovariance of the process ξ_t we get

$$\begin{aligned} \gamma_y(l) &= 2\sigma^6\alpha^2_{|l|}(\gamma_\xi(0) - \sigma^2) + 2\sigma^4(\gamma_\xi(l) - \sigma^2\alpha_{|l|})^2 \\ &+4\mu\sigma^4\left[\rho^2\alpha_{|l|} + \alpha_{|l|}(\gamma_\xi(l) - \sigma^2\alpha_{|l|})\right] \\ &+2\rho^2\sigma^6\alpha^2_{|l|} + 4\rho^2\sigma^4(\gamma_\xi(l) - \sigma^2\alpha_{|l|}) \\ &= 2\sigma^6\alpha^2_{|l|}\gamma_\xi(0) - 2\sigma^8\alpha^2_{|l|} + 2\sigma^4\gamma^2_\xi(l) + 2\sigma^8\alpha^2_{|l|} - 4\sigma^6\gamma_\xi(l)\alpha_{|l|} \\ &+4\mu\rho^2\sigma^4\alpha_{|l|} - 4\rho^2\sigma^6\alpha_{|l|} + 2\rho^2\sigma^6\alpha^2_{|l|} - 4\mu\sigma^6\alpha^2_{|l|} + 4\mu\sigma^4\alpha_{|l|}\gamma_\xi(l) + 4\rho^2\sigma^4\gamma_\xi(l)\,. \end{aligned}$$

Thus summing up we obtain

$$\begin{aligned} \gamma_y(l) &= (2\sigma^6\gamma_\xi(0) + 2\rho^2\sigma^6 - 4\mu\sigma^6)\alpha^2_{|l|} \\ &+4(\mu-\sigma^2)\sigma^4\gamma_\xi(l)\alpha_{|l|} + 4(\mu-\sigma^2)\rho^2\sigma^4\alpha_{|l|} + 2\sigma^4\gamma^2_\xi(l) + 4\rho^2\sigma^4\gamma_\xi(l)\,. \end{aligned}$$

Then, by using (2.20) (2.21) and (2.22), we can write the latter expression in terms of the spectrum and the transfer function for the process ξ_t as:

$$\gamma_y(l) = 2\sigma^6(\gamma_\xi(0) + \rho^2 - 2\mu)\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \beta(\omega)\beta(\lambda)e^{i\omega l}e^{i\lambda l}d\omega d\lambda$$

$$+\frac{2\sigma^8}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | \alpha(\omega) |^2 | \alpha(\lambda) |^2 e^{i\omega l} e^{i\lambda l} d\omega d\lambda$$
$$+4\sigma^6(\mu-\sigma^2)\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | \alpha(\omega) |^2 \beta(\lambda) e^{i\omega l} e^{i\lambda l} d\omega d\lambda$$
$$+\frac{4\rho^2\sigma^4}{2\pi}(\mu-\sigma^2)\int_{-\pi}^{\pi}\beta(\nu)e^{i\nu l}d\nu+\frac{4\rho^2\sigma^6}{2\pi}\int_{-\pi}^{\pi} | \alpha(\nu) |^2 e^{i\nu l}d\nu .$$

and so by setting $\omega + \lambda = \nu$ we obtain

$$\gamma_y(l) =$$
$$= 2\sigma^6(\gamma_\xi(0)+\rho^2-2\mu)\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\left[\int_{-\pi}^{\pi}\beta(\lambda)\beta(\nu-\lambda)d\lambda\right]e^{i\nu l}d\nu$$
$$+\frac{2\sigma^8}{4\pi^4}\int_{-\pi}^{\pi}\left[\int_{-\pi}^{\pi} | \alpha(\lambda) |^2 | \alpha(\nu-\lambda) |^2 d\lambda\right]e^{i\nu l}d\nu$$
$$+4\sigma^6(\mu-\sigma^2)\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\left[\int_{-\pi}^{\pi}\beta(\lambda) | \alpha(\nu-\lambda) |^2 d\lambda\right]e^{i\nu l}d\nu$$
$$+\frac{4\rho^2\sigma^6}{2\pi}\int_{-\pi}^{\pi} | \alpha(\nu) |^2 e^{i\nu l}d\nu+\frac{4(\mu-\sigma^2)\rho^2\sigma^4}{2\pi}\int_{-\pi}^{\pi}\beta(\nu)e^{i\nu l}d\nu .$$

Then by equating the last expression to

$$\int_{-\pi}^{\pi} f_y(\nu)e^{i\nu l}d\nu . \tag{2.26}$$

and summing ν_y,the part of the variance which is not included in $\gamma_y(0)$ as from Theorem 3, the result follows.

□

In order to identify and thus to estimate the model a different parameterization will be needed. Let us set

$$\bar{\rho}^2 = \frac{\rho^2}{\sigma^2},\ \bar{\mu} = \frac{\mu}{\sigma^2}.$$

Then the power spectrum for y_t can be rewritten as

$$f_y(\nu) = \sigma^8 K_y(\nu), \tag{2.27}$$

where

$$K_y(\nu) = 2(\delta_{\alpha\alpha}(0)+1+\bar{\rho}-2\bar{\mu})\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\beta(\lambda)\beta(\nu-\lambda)d\lambda$$
$$+\frac{2}{4\pi^4}\int_{-\pi}^{\pi} | \alpha(\lambda) |^2 | \alpha(\nu-\lambda) |^2 d\lambda$$
$$+\frac{4(\bar{\mu}-1)}{4\pi^2}\int_{-\pi}^{\pi}\beta(\lambda) | \alpha(\nu-\lambda) |^2 d\lambda$$

$$+\frac{4\bar{\rho}^2}{2\pi} \mid \alpha(\nu) \mid^2 + \frac{4(\bar{\mu}-1)\bar{\rho}^2}{2\pi}\beta(\nu) + \frac{\nu'_y}{2\pi},$$

with $\nu'_y = 2\bar{\rho}^4 + 8\bar{\rho}^2\delta_{\alpha\alpha}(0) + 6\delta^2_{\alpha\alpha}(0) + 4\bar{\mu}^2(\bar{\rho}^2 + \delta_{\alpha\alpha}(0))$.

Alternative expressions for $f_y(\lambda)$ are given as follows.

Corollary 2 *Under Assumption B' two alternative expressions for $f_y(\lambda)$ are given by*

(I)

$$\begin{aligned} f_y(\nu) &= \frac{1}{2\pi}\sum_{l=-\infty}^{\infty} 2\sigma^8\delta_{\alpha\alpha}^2(l)e^{i\nu l} + \frac{1}{2\pi}\sum_{l=-\infty}^{\infty} 4\rho^2\sigma^6\delta_{\alpha\alpha}(l)e^{i\nu l} + \frac{\nu_y}{2\pi} \\ &+ \frac{1}{2\pi}\sum_{l=-\infty}^{\infty} 2\sigma^4\alpha_{|l|}\left(\sigma^2\alpha_{|l|}(\sigma^2\delta_{\alpha\alpha}(0)+\rho^2)+2\mu(\rho^2+\sigma^2\delta_{\alpha\alpha}(l))\right)e^{i\nu l}, \end{aligned}$$

(II)

$$\begin{aligned} f_y(\nu) &= \frac{\sigma^8}{\pi}\Bigg[(\delta_{\alpha\alpha}(0)+2\bar{\mu}-\bar{\rho}^2+1)\sum_{l=-\infty}^{\infty}\alpha_{|l|}^2 e^{il\nu} + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\sum_{l=-\infty}^{\infty}\alpha_m\alpha_{m+l}\alpha_n\alpha_{n+l}e^{il\nu} \\ &+ 2\bar{\mu}\sum_{m=1}^{\infty}\sum_{l=-\infty}^{\infty}\alpha_m\alpha_{m+l}\alpha_{|l|}e^{i\nu l} + 2\bar{\rho}^2\sum_{m=1}^{\infty}\sum_{l=-\infty}^{\infty}\alpha_m\alpha_{m+l}e^{i\nu l} + \frac{\nu_y'}{2}\Bigg], \nu_y' = \frac{\nu_y}{\sigma^8}. \end{aligned}$$

The power spectrum for the process obtained as a result of this particular nonlinear filter, is a linear combination of convolutions, not surprisingly, given the particular nonlinearity involved viz. squaring the observable process .

2.2.3 Kurtosis and 'leverage' effect for x_t

Defining the coefficient of Kurtosis as

$$kurt(x_t) = \frac{E(x_t-\mu)^4}{[E(x_t-\mu)^2]^2}, \tag{2.28}$$

we obtain the expression for the coefficient of kurtosis of the process x_t as follows.

Theorem 7 *Under Assumption B ,*

$$\begin{aligned} kurt(x_t) &= 3+\left(\frac{2\delta_{\alpha\alpha}(0)(4\bar{\rho}^2+\delta_{\alpha\alpha}(0))}{(\bar{\rho}^2+\delta_{\alpha\alpha}(0))^2}\right)+\frac{\kappa_4}{\sigma^4}\left(1+\frac{3\bar{\rho}^2\delta_{\alpha\alpha}(0)+2\delta_{\alpha\alpha}^2(0)+\delta_{\alpha^2\alpha^2}(0)}{(\bar{\rho}^2+\delta_{\alpha\alpha}(0))^2}\right) \\ &+\frac{\kappa_4^2\delta_{\alpha^2\alpha^2}(0)+4\rho(2\sigma^4+\kappa_4)\mu_3\delta_{\alpha^2\alpha}(0)}{(\bar{\rho}^2+\delta_{\alpha\alpha}(0))^2}. \end{aligned}$$

Proof: Writing

$$kurt(x_t) = \frac{var((x_t - \mu)^2)}{[E(x_t - \mu)^2]^2} + 1,$$

we can make use of Theorem 4 which will give the denominator. Thus from

$$var(x_t) = \sigma^4(\bar{\rho}^2 + \delta_{\alpha\alpha}(0)),$$

the result follows.

□

Corollary 3 *Under Assumption B' we obtain*

$$kurt(x_t) = 3 + \phi(a) \text{ setting } a = \bar{\rho}^2/\delta_{\alpha\alpha}(0),$$

where the non-negative function $\phi(.)$ is defined as

$$\phi(a) = \frac{2(4a+3)}{(a+1)^2}.$$

Given the (not increasing) monotonicity of the function $\phi(.)$ we obtain that

$$3 < kurt(x_t) \leq 8 \text{ for } 0 \leq a < \infty,$$

with $\lim_{a\to\infty} kurt(x_t) = 3$.

Furthermore, since Black (1976), another characteristic of returns distribution is the so called 'leverage' effect, the negative correlation between current returns and future volatility. This reflects the increase of the debt-to-equity ratio, in turn due to a reduction in the equity value and thus an increase in the firm riskiness.

Henceforth for any bivariate stationary process $\{w_t, k_t\}$ we define the cross-covariance function $\gamma_{wk}(u)$, $u = 0, \pm 1, \pm 2, \ldots$ and, when it exists , the cross spectral density function $f_{wk}(\lambda)$, $-\pi \leq \lambda < \pi$.

Theorem 8 *Under Assumption B,*

$$\gamma_{xy}(l) = \begin{cases} 2\sigma^4\alpha_l(\sigma^2\delta_{\alpha\alpha}(l) + \rho^2) + \mu_3\rho\sigma^2\alpha_l^2, & l > 0, \\ \mu_3\left(\rho^3 + 2\rho\sigma^2\delta_{\alpha\alpha}(0) + \mu_3\delta_{\alpha\alpha^2}(0) + 2\rho\sigma^2\delta_{\alpha\alpha}(0)\right) + & \\ 2\mu\sigma^4(\sigma^2\delta_{\alpha\alpha}(0) + \rho^2), & l = 0, \\ 0, & l < 0. \end{cases}$$

Proof: The result follows by repeating the arguments in Theorem 2.

□

Hence the model does take into account the leverage effect either for $\mu_3 < 0$, reflecting a strong asymmetry in the ϵ_t which would be consistent with the nature of this aspect of the returns distribution. More in general leverage effect, at lag L, arises whenever $\alpha_L < 0$. Furthermore the model is apt to display negative 'contemporaneous' correlation between asset return and asset volatility (Campbell and Hentschel 1990). Even with respect to this asymmetric effect, this family of nonlinear MA appear to capture a great deal of the 'stylized facts' characterizing the empirical distribution of asset return[3]

We derive the cross spectrum for the x_t and the y_t which will be considered when discussing the estimation of the model (cf. Chapter 3, section 3.3).

Theorem 9 *Under Assumption B', for $-\pi \leq \lambda < \pi$*

$$f_{xy}(\lambda) = \frac{2\sigma^6}{4\pi^2}\int_{-\pi}^{\pi} \mid \alpha(\lambda - \omega) \mid^2 \alpha(\omega)d\omega + \frac{2\sigma^4(\rho^2 - \sigma^2)}{2\pi}\alpha(\lambda) + \frac{\nu_{xy}}{2\pi},$$

with $\nu_{xy} = 2\mu\sigma^4(\sigma^2\delta_{\alpha\alpha}(0) + \rho^2)$.

Proof: The result follows by repeating the arguments in Theorem 6.

□

[3]This dynamic asymmetry is taken into account by further refinements of the ARCH-type models (Nelson 1991, the EGARCH) (Sentana 1995, the QARCH) .

2.2.4 Behaviour of the power spectrum near zero frequency

At first we shall consider the following definition which plays a crucial role in all the successive analysis.

Definition 3 . *(Yong 1974, Def I-1)*
A sequence $\{a_n\}$ *of positive numbers is said to be quasi- monotonically convergent to zero (QMC) if* $a_n \to 0$ *and for some constant* $\beta \geq 0$,

$$a_{n+1} \leq a_n(1+\frac{\beta}{n}) \text{ for all } n \geq n_0(\beta).$$

Then the following proposition will be used.

Proposition 2 *(Yong 1974, Theorem III - 12) Let* $\{u_k, k \geq 1\}$ *be of bounded variation and quasi-monotonically convergent to zero. Let* $0 < \alpha < 1$. *Let* $L(x)$ *being a slowly varying function at infinity (Yong 1974) and let* $C(\alpha) = \frac{\pi}{2\Gamma(\alpha)cos(\alpha\pi/2)}$. *Then:*

$$u_k \sim k^{-\alpha}L(k) \text{ as } k \to \infty$$

if and only if

$$\sum_{k=1}^{\infty} u_k \, cos(k\omega) \sim C(\alpha)\omega^{\alpha-1}L(\frac{1}{\omega}) \text{ as } \omega \to 0^+.$$

To make use of Proposition 2 we need to prove the QMC of $\gamma_y(u)$.

We obtain the following

Theorem 10 *Under Assumptions* B', A_1, A_2 , *as* $\lambda \to 0^+$

$$f_y(\lambda) \sim \begin{cases} K \mid \lambda \mid^{1-4d}, & \rho = 0, \quad 1/4 < d < 1/2, \\ K \mid \lambda \mid^{-2d}, & \rho \neq 0, \quad 0 < d < 1/2. \end{cases} \tag{2.29}$$

Proof: From expression (I) of Corollary 6, we can write the power spectrum as

$$f_y(\nu) = f_1(\nu) + f_2(\nu;\rho) + f_3 + f_4(\nu) + f_5(\nu;\mu,\rho) + f_6(\nu;\mu),$$

with

$$
\begin{aligned}
f_1(\nu) &= \frac{2\sigma^8}{2\pi}\sum_{l=-\infty}^{\infty}\delta^2_{\alpha\alpha}(l)e^{i\nu l}\,,\\
f_2(\nu;\rho) &= \frac{4\rho^2\sigma^6}{2\pi}\sum_{l=-\infty}^{\infty}\delta_{\alpha\alpha}(l)e^{i\nu l}\,,\\
f_3 &=\\
&= \frac{\sigma^6\delta_{\alpha\alpha}(0)}{2\pi}(6\sigma^2+8\rho^2\delta_{\alpha\alpha}(0))+\frac{1}{2\pi}\left(2\rho^4\sigma^4+4\mu^2\sigma^2(\rho^2+\sigma^2\delta_{\alpha\alpha}(0))\right).\\
f_4(\nu) &= \frac{2\sigma^6(\sigma^2\delta_{\alpha\alpha}(0)+\rho^2)}{2\pi}\sum_{l=-\infty}^{\infty}\alpha^2_{|l|}e^{i\nu l}.\\
f_5(\nu;\mu,\rho) &= \frac{4\mu\sigma^6\rho^2}{2\pi}\sum_{l=-\infty}^{\infty}\alpha_{|l|}e^{i\nu l}.\\
f_6(\nu;\mu) &= \frac{4\mu\sigma^8}{2\pi}\sum_{l=-\infty}^{\infty}\alpha_{|l|}\delta_{\alpha\alpha}(l)e^{i\nu l}.
\end{aligned}
$$

Thus f_3 is part of the variance of the process and so constant in the frequency ν. The notation used reflects for each of the addenda whether they depend on the parameters ρ and μ in such a way that for $a \in \{\mu,\rho\}$ and for all $\nu \in [-\pi,\pi)$

$$f_i(\nu;a)\mid_{a=0}=0,\ i=2,5,6.$$

By use of Lemma 27 and assumption A_1 we obtain that the dominating term is $f_2(.)$ when $\rho\neq 0$ and $f_1(.)$ when $\rho=0$. The addenda that involve μ are always dominated in both cases.

In fact we obtain, as $l \to \infty$

$$
\begin{aligned}
&\text{for } f_1(.),\ \delta^2_{\alpha\alpha}(l)\sim K\,l^{4d-2},\\
&\text{for } f_2(.),\ \delta_{\alpha\alpha}(l)\sim K\,l^{2d-1},\\
&\text{for } f_4(.),\ \alpha^2_{|l|}\sim K\,l^{2d-2},\\
&\text{for } f_5(.),\ \alpha_{|l|}\sim K\,l^{d-1},\\
&\text{for } f_6(.),\ \alpha_{|l|}\delta_{\alpha\alpha}(l)\sim K\,l^{3d-2}.
\end{aligned}
$$

By Assumptions A_1, A_2 and by Lemma 29, by applying Proposition 2 to each of the $f_i(.)$, $i=1,2,4,5,6$, we get, as $\nu \to 0^+$

$$f_1(\nu)\sim K\mid\nu\mid^{1-4d},\ \text{for } 1/4<d<1/2,$$

$$f_2(\nu;\rho) \sim K \mid \nu \mid^{-2d}, \text{ for } 0 < d < 1/2\,,$$
$$f_5(\nu;\mu,\rho) \sim K \mid \nu \mid^{-d}, \text{ for } 0 < d < 1\,,$$
$$f_6(\nu;\mu) \sim K \mid \nu \mid^{1-3d}, \text{ for } 1/3 < d < 2/3\,,$$

where we stressed the values of d which enable to apply Proposition 2. But by A_1 to achieve stationarity $d < 1/2$.

□

Remarks: (1) With respect of notation of Proposition 2 the slowly varying function is just a constant function of the frequency.

(2) When $\rho = 0$ we can see from Theorem 6 that the 'interesting' part of the spectrum for the squared process is a convolution of the spectrum for a linear process. By using the result in Zygmund (1977) about the p-integrability of a function, convolution of two q- integrable functions, viz. $p = 2/q - 1$, it follows that $f_y(.)$ inherits directly the integrability property (i.e. stationarity) and that it is continuous (p=0), and so not long-memory, when the spectrum of the linear component $f_\xi(.)$ is square-integrable (q=2), that is for $d < 1/4$, confirming the results obtained by analyzing the behaviour of $\gamma_y(l)$ at long lags (Theorem 4) or the behaviour of $f_y(.)$ near the zero frequency (Theorem 10).

2.2.5 The trispectrum for y_t

The process $\{y_t\}$ is both nonlinear and non-Gaussian. Non-Gaussianity implies that other aspects of the distribution of y_t are contained in high-order moments other than the mean and the variance. Given that an estimation approach based on a Gaussian pseudo maximum likelihood estimator is proposed in the following chapter, nonlinearity of the model is likely to generate a 'cumulant' term in the asymptotic variance-covariance matrix.

Let us evaluate the fourth cumulant for the squared process and from it the trispectrum. For simplicity let us set $\mu = 0$, otherwise the number of terms involved becomes intractable.

Let us define for any integer j

$$cum_j(\underbrace{.,.,\ldots,.}_{j\,terms}),$$

the cumulant operator in j arguments.

By stationarity

$$cum_{yyy}(a,b,c) = cum_4(y_t, y_{t+a}, y_{t+b}, y_{t+c}), \tag{2.30}$$

$cum_{yyy}(.,.,.)$ denoting the fourth-order cumulant for the y_t.

Let us set for simplicity

$$k_1 = \bar{\rho}^2 + \delta_{\alpha\alpha}(0).$$

Theorem 11 *Under Assumption B' and setting for simplicity $\delta_{\alpha\alpha}(l) = \delta_{(l)}$, $l = 0, \pm 1, \ldots$ we have*

$$\begin{aligned}
&cum_{yyy}(a,b,c) = 8\sigma^{16}\alpha_a^2\alpha_{b-a}^2\alpha_{c-a}^2 k_1 \\
&+\sigma^4\Big[2\sigma^4\alpha_{c-a}^2\Big\{2\sigma^4\alpha_a^2\left(2\sigma^4\delta_{(b)}^2 + 4\rho^2\sigma^2\delta_{(b)}\right) + 2\sigma^4\alpha_b^2\left(2\sigma^4\delta_{(a)}^2 + 4\rho^2\sigma^2\delta_{(a)}\right) \\
&+\sigma^2 k_1 8\alpha_a\alpha_b\left(\sigma^6\delta_{(b-a)} + \rho^2\sigma^4\right)\Big\} \\
&+2\sigma^4\alpha_{b-a}^2\Big\{2\sigma^4\alpha_a^2\left(2\sigma^4\delta_{(c)}^2 + 4\rho^2\sigma^2\delta_{(c)}\right) + 2\sigma^4\alpha_c^2\left(2\sigma^4\delta_{(a)}^2 + 4\rho^2\sigma^2\delta_{(a)}\right) \\
&+\sigma^2 k_1 8\alpha_a\alpha_c\left(\sigma^6\delta_{(c-a)} + \rho^2\sigma^4\right)\Big\} \\
&+2\sigma^8\alpha_{c-b}^2\Big\{2\sigma^4\alpha_b^2\left(2\sigma^4\delta_{(a)}^2 + 4\rho^2\sigma^2\delta_{(a)}\right) + 2\sigma^4\alpha_a^2\left(2\sigma^4\delta_{(b)}^2 + 4\rho^2\sigma^4\delta_{(b)}\right) \\
&+\sigma^2 8k_1\alpha_a\alpha_b\left(\sigma^6\delta_{(b-a)} + \rho^2\sigma^4\right)\Big\}\Big] + \Big\{2k_1\sigma^{10}\alpha_a^2 8\alpha_{b-a}\alpha_{c-a}\left(\sigma^6\delta_{(c-b)} + \rho^2\sigma^4\right)\Big\}
\end{aligned}$$

$$+\sigma^6 \Big\{2\alpha_c^2\sigma^4 \Big[8\sigma^6\delta_{(b)}\delta_{(a)}\delta_{(b-a)} + 8\rho^2\sigma^4 \Big(\delta_{(a)}\delta_{(b)} + \delta_{(b)}\delta_{(b-a)} + \delta_{(a)}\delta_{(b-a)}\Big)\Big]$$

$$+2\alpha_b^2\sigma^4 \Big[8\sigma^6(\delta_{(a)}\delta_{(c)}\delta_{(c-a)} + 8\rho^2\sigma^4 \Big(\delta_{(a)}\delta_{(c)} + \delta_{(a)}\delta_{(c-a)} + \delta_{(c)}\delta_{(c-a)}\Big)\Big]$$

$$+2\alpha_a^2\sigma^4 \Big[8\sigma^6\delta_{(b)}\delta_{(c)}\delta_{(c-b)} + 8\rho^2\sigma^4 \Big(\delta_{(b)}\delta_{(c)} + \delta_{(b)}\delta_{(c-b)} + \delta_{(c)}\delta_{(c-b)}\Big)\Big]$$

$$+\sigma^2k_1 \Big[16\sigma^8 \Big(\alpha_b\alpha_c\delta_{(b-a)}\delta_{(c-a)} + \alpha_a\alpha_b\delta_{(c-a)}\delta_{(c-b)} + \alpha_a\alpha_c\delta_{(b-a)}\delta_{(c-b)}\Big)$$

$$+16\rho^2\sigma^4 \Big(\alpha_b\alpha_c(\delta_{(b-a)} + \delta_{(c-a)}) + \alpha_a\alpha_b(\delta_{(c-b)} + \delta_{(c-a)}) + \alpha_a\alpha_c(\delta_{(c-b)} + \delta_{(b-a)})\Big)\Big]$$

$$+16\rho^2\sigma^4 \Big(\alpha_b\alpha_c(\delta_{(b-a)} + \delta_{(c-a)}) + \alpha_a\alpha_b(\delta_{(c-b)} + \delta_{(c-a)}) + \alpha_a\alpha_c(\delta_{(c-b)} + \delta_{(b-a)})\Big)\Big]$$

$$+8\alpha_a\alpha_c \Big(2\sigma^4\delta_{(b)}^2 + 4\rho^2\sigma^2\delta_{(b)}\Big) \Big(\sigma^6\delta_{(c-a)} + \rho^2\sigma^4\Big)$$

$$+8\alpha_b\alpha_c \Big(2\sigma^4\delta_{(a)}^2 + 4\rho^2\sigma^2\delta_{(a)}\Big) \Big(\sigma^6\delta_{(c-b)} + \rho^2\sigma^4\Big)\Big\}$$

$$+8\sigma^{14}\alpha_{c-b}^2\alpha_{b-a}^2 \Big(\sigma^2\delta_{(a)}^2 + 2\rho^2\delta_{(a)}\Big)$$

$$+\sigma^6 \Big\{16\sigma^4\alpha_{c-a}^2 \Big[\sigma^6\delta_{(b)}\delta_{(a)}\delta_{(b-a)} + \rho^2\sigma^4 \Big(\delta_{(b)}\delta_{(c)} + \delta_{(b)}\delta_{(c-b)} + \delta_{(c)}\delta_{(c-b)}\Big)\Big]$$

$$+16\sigma^4\alpha_{b-a}^2 \Big[\sigma^6\delta_{(a)}\delta_{(c)}\delta_{(c-a)} + \rho^2\sigma^4 \Big(\delta_{(a)}\delta_{(c)} + \delta_{(a)}\delta_{(c-a)} + \delta_{(c)}\delta_{(c-a)}\Big)\Big]$$

$$+16\sigma^4\alpha_{c-b}^2 \Big[\sigma^6\delta_{(a)}\delta_{(b)}\delta_{(b-a)} + \rho^2\sigma^4 \Big(\delta_{(a)}\delta_{(b)} + \delta_{(a)}\delta_{(b-a)} + \delta_{(b)}\delta_{(b-a)}\Big)\Big]\Big\}$$

$$+8\sigma^6\alpha_{b-a}\alpha_{c-a}\delta_{(a)} \Big(\sigma^6\delta_{(c-b)} + \rho^2\sigma^4\Big) \Big(2\sigma^4\delta_{(a)} + 4\rho^2\sigma^2\Big)$$

$$+16\sigma^8 \Big\{\sigma^8 \Big[\delta_{(a)}\delta_{(b)}\delta_{(c-b)}\delta_{(c-a)} + \delta_{(a)}\delta_{(c)}\delta_{(b-a)}\delta_{(c-b)} + \delta_{(b)}\delta_{(c)}\delta_{(b-a)}\delta_{(c-a)}\Big]$$

$$+\rho^2\sigma^6 \Big[\delta_{(a)}\delta_{(b)}(\delta_{(c-b)} + \delta_{(b-a)}) + \delta_{(a)}\delta_{(c)}(\delta_{(c-b)} + \delta_{(c-a)}) + \delta_{(b)}\delta_{(c)}(\delta_{(b-a)} + \delta_{(c-a)})\Big]\Big\}$$

$$+16\rho^2\sigma^{14}\Big[\delta_{(b)}\delta_{(c-a)} \Big(\delta_{(c-b)}+\delta_{(b-a)}\Big)+\delta_{(c)}\delta_{(b-a)} \Big(\delta_{(c-a)}+\delta_{(c-b)}\Big)+\delta_{(a)}\delta_{(c-b)} \Big(\delta_{(b-a)}+\delta_{(c-a)}\Big)\Big].$$

Proof: Let us set $0 \le a \le b \le c$. Setting

$$\bar{\epsilon}_t^2 = \epsilon_t^2 - \sigma^2 , \tag{2.31}$$

then we can express the fourth cumulant as

$$cum_{y,y,y}(a,b,c) = (1) + (2) + (3) + (4) , \tag{2.32}$$

where

$$(1) = \quad cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, \bar{\epsilon}_{t+c}^2 h_{t-1+c}^2) + \tag{2.33}$$

$$cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) + \tag{2.34}$$

$$cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, h_{t-1+c}^2 \bar{\epsilon}_{t+c}^2) + \tag{2.35}$$

$$cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) . \tag{2.36}$$

$$
\begin{aligned}
(2) = \quad & cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, \bar{\epsilon}_{t+c}^2 h_{t-1+c}^2) + && (2.37)\\
& cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) + && (2.38)\\
& cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, h_{t-1+c}^2 \bar{\epsilon}_{t+c}^2) + && (2.39)\\
& cum_4(\bar{\epsilon}_t^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) . && (2.40)\\
(3) = \quad & cum_4(\sigma^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, h_{t-1+c}^2 \bar{\epsilon}_{t+c}^2) + && (2.41)\\
& cum_4(\sigma^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) + && (2.42)\\
& cum_4(\sigma^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, h_{t-1+c}^2 \bar{\epsilon}_{t+c}^2) + && (2.43)\\
& cum_4(\sigma^2 h_{t-1}^2, \bar{\epsilon}_{t+a}^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) . && (2.44)\\
(4) = \quad & cum_4(\sigma^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, h_{t-1+c}^2 \bar{\epsilon}_{t+c}^2) + && (2.45)\\
& cum_4(\sigma^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \bar{\epsilon}_{t+b}^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) + && (2.46)\\
& cum_4(\sigma^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, h_{t-1+c}^2 \bar{\epsilon}_{t+c}^2) + && (2.47)\\
& cum_4(\sigma^2 h_{t-1}^2, \sigma^2 h_{t-1+a}^2, \sigma^2 h_{t-1+b}^2, \sigma^2 h_{t-1+c}^2) . && (2.48)
\end{aligned}
$$

Now by direct evaluation of the cumulant, given the independence of the ϵ_t and the fact that

$$E(\bar{\epsilon}_t^2) = 0 ,$$

we have that for each of the term as

$$cum_4(.,.,., \bar{\epsilon}_{t+c}^2 h_{t-1+c}^2) = 0 , \tag{2.49}$$

so that eight out of sixteen terms in (2.32) are equal to zero.

Finally, setting as in Theorem 2

$$w_{a,(b,c)} = \sum_{k=a-b}^{a-c} \alpha_k \epsilon_{t+a-k} \text{ where } a > b \geq c$$

we have that

$$h_{t+a-1} = \rho + w_{a,(a-1,-\infty)} ,$$

$$h_{t+b-1} = \rho + w_{b,(b-1,-\infty)} ,$$

with

$$w_{b,(b-1,-\infty)} = w_{b,(b-1,a+1)} + w_{b,(a,a)} + w_{b,(a-1,-\infty)} ,$$

$$h_{t+c-1} = \rho + w_{c,(c-1,-\infty)} ,$$

with

$$w_{c,(c-1,-\infty)} = w_{c,(c-1,b+1)} + w_{c,(b,b)} + w_{c,(b-1,a+1)} + w_{c,(a,a)} + w_{b,(a-1,-\infty)} .$$

Thus

$$h^2_{t+a-1} = \rho^2 + w^2_{a,(a-1,-\infty)} + 2\rho w_{a,(a-1,-\infty)}\,, \tag{2.50}$$

$$h^2_{t+b-1} = \rho^2 + w^2_{b,(b-1,a+1)} + w^2_{b,(a,a)} + w^2_{b,(a-1,-\infty)}$$
$$+2w_{b,(b-1,a+1)}\left(w_{c,(a,a)} + w_{c,(a-1,-\infty)}\right) + 2w_{b,(a,a)}w_{b,(a-1,-\infty)} + 2\rho w_{b,(b-1,-\infty)}\,.$$

$$h^2_{t+c-1}$$
$$= \rho^2 + w^2_{c,(c-1,b+1)} + w^2_{c,(b,b)} + w^2_{c,(b-1,a+1)} + w^2_{c,(a,a)} + w^2_{c,(a-1,\infty)}$$
$$+2w_{c,(c-1,b+1)}\left(w_{c,(b,b)} + w_{c,(b-1,a+1)} + w_{c,(a,a)} + w_{c,(a-1,-\infty)}\right)$$
$$+2w_{c,(b,b)}\left(w_{c,(b-1,a+1)} + w_{c,(a,a)} + w_{c,(a-1,-\infty)}\right)$$
$$+2w_{c,(b-1,a+1)}\left(w_{c,(a,a)} + w_{c,(a-1,-\infty)}\right)$$
$$+2w_{c,(a,a)}w_{c,(a-1,-\infty)} + 2\rho w_{c,(c-1,-\infty)}\,.$$

Let us now evaluate each term from (2.34) to (2.48). We report in Appendix C the graphs representing the 'links' on which the evaluation of the terms (2.34)-(2.48) is based.

With respect to element (2.34), from

$$E(\bar{\epsilon}^2_t \epsilon_t) = 0\,, \tag{2.51}$$

we have that the only relevant element in the expression of h^2_{t+c-1} is given by

$$w_{c,(b,b)} = \alpha_{c-b}\epsilon_{t+b}\,,$$

so that (2.34) becomes

$$\sigma^2\alpha^2_{c-b}cum_4(\bar{\epsilon}^2_t h^2_{t-1}, \bar{\epsilon}^2_{t+a}h^2_{t-1+a}, \bar{\epsilon}^2_{t+b}h^2_{t-1+b}, \epsilon^2_{t+b})\,. \tag{2.52}$$

Then, by the cumulants' reduction theorem (Proposition 1) the cumulant expression of (2.52) can be written as

$$2\sigma^4 cum_2(\bar{\epsilon}^2_t, h^2_{t-1+a})cum_1(h^2_{t-1})cum_2(\,\bar{\epsilon}^2_{t+a}, h^2_{t-1+b})\,, \tag{2.53}$$

so that

$$(2.34) = \alpha^2_a\alpha^2_{b-a}\alpha^2_{c-a}8(\sigma^2)^8 k_1\,. \tag{2.54}$$

For the other terms we have in turn

$$(2.36) = \quad \sigma^4 \Big[cum_2(\bar{\epsilon}^2_{t+a}, h^2_{t-1+c}) \Big\{cum_2(\bar{\epsilon}^2_t, h^2_{t-1+a}) cum_2(h^2_{t-1}, h^2_{t-1+b})$$
$$+cum_2(\bar{\epsilon}^2_t, h^2_{t-1+b}) cum_2(h^2_{t-1}, h^2_{t-1+a})$$
$$+ cum_3(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+b}) cum_1(h^2_{t-1})\Big\}$$
$$+cum_2(\bar{\epsilon}^2_{t+a}, h^2_{t-1+b}) \Big\{cum_2(\bar{\epsilon}^2_t, h^2_{t-1+a}) cum_2(h^2_{t-1}, h^2_{t-1+c})$$
$$+cum_2(\bar{\epsilon}^2_t, h^2_{t-1+c}) cum_2(h^2_{t-1}, h^2_{t-1+a})$$
$$+cum_3(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+c}) cum_1(h^2_{t-1})\Big\} +$$
$$+cum_3(\bar{\epsilon}^2_{t+a}, h^2_{t-1+b}, h^2_{t-1+c}) \Big\{cum_1(h^2_{t-1}) cum_2(\bar{\epsilon}^2_t, h^2_{t-1+a})\Big\}\Big] .$$

$$(2.38) = \quad \sigma^4 cum_2(\bar{\epsilon}^2_{t+b}, h^2_{t-1+c}) \Big\{cum_2(\bar{\epsilon}^2_t, h^2_{t-1+b}) cum_2(h^2_{t-1}, h^2_{t-1+a})$$
$$+cum_2(\bar{\epsilon}^2_t, h^2_{t-1+a}) cum_2(h^2_{t-1}, h^2_{t-1+b}) + cum_3(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+b}) cum_1(h^2_{t-1})\Big\} .$$

$$(2.40) = \quad \sigma^6 \Big\{cum_2(\bar{\epsilon}^2_t, h^2_{t-1+c}) cum_3(h^2_{t-1}, h^2_{t-1+a}, h^2_{t-1+b})$$
$$+cum_2(\bar{\epsilon}^2_t, h^2_{t-1+b}) cum_3(h^2_{t-1}, h^2_{t-1+a}, h^2_{t-1+c})$$
$$+cum_2(\bar{\epsilon}^2_t, h^2_{t-1+a}) cum_3(h^2_{t-1}, h^2_{t-1+b}, h^2_{t-1+c})$$
$$+cum_4(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+b}, h^2_{t-1+c}) cum_1(h^2_{t-1})$$
$$+cum_3(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+b}) cum_2(h^2_{t-1}, h^2_{t-1+c})$$
$$+cum_3(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+c}) cum_2(h^2_{t-1}, h^2_{t-1+b})$$
$$+ cum_3(\bar{\epsilon}^2_t, h^2_{t-1+b}, h^2_{t-1+c}) cum_2(h^2_{t-1}, h^2_{t-1+a})\Big\} .$$

$$(2.42) = \sigma^4 cum_2(\bar{\epsilon}^2_{t+b}, h^2_{t-1+c}) cum_2(\bar{\epsilon}^2_{t+a}, h^2_{t-1+b}) cum_2(h^2_{t-1}, h^2_{t-1+a}) .$$

$$(2.44) = \quad \sigma^6 \Big\{cum_2(\bar{\epsilon}^2_{t+a}, h^2_{t-1+c}) cum_3(h^2_{t-1}, h^2_{t-1+a}, h^2_{t-1+b})$$
$$+cum_2(\bar{\epsilon}^2_{t+a}, h^2_{t-1+b}) cum_3(h^2_{t-1}, h^2_{t-1+a}, h^2_{t-1+c})$$
$$+ cum_3(\bar{\epsilon}^2_{t+a}, h^2_{t-1+b}, h^2_{t-1+c}) cum_2(h^2_{t-1}, h^2_{t-1+a})\Big\} .$$

$$(2.46) = \sigma^6 cum_2(\bar{\epsilon}^2_{t+b}, h^2_{t-1+c}) cum_3(h^2_{t-1}, h^2_{t-1+a}, h^2_{t-1+b}) .$$

$$(2.48) = \sigma^8 cum_4(h^2_{t-1}, h^2_{t-1+a}, h^2_{t-1+b}, h^2_{t-1+c}) .$$

All the cumulants in (2.36)-(2.48) are reported in Appendix B (section B.1- B.7). Hence by simple substitution the result follows.

□

Corollary 4 *Under Assumption B' and for $\rho = 0$ we have*

$$
\begin{aligned}
& cum_{yyy}(a,b,c) = 8\sigma^{16}\alpha_a^2\alpha_{b-a}^2\alpha_{c-a}^2 k_1 \\
& +8\sigma^{16}\left[\alpha_{c-a}^2\left\{\alpha_a^2\delta_{(b)}^2 + \alpha_b^2\delta_{(a)}^2 + 2k_1\alpha_a\alpha_b\delta_{(b-a)}\right\} + \alpha_{b-a}^2\left\{\alpha_a^2\delta_{(c)}^2 + \alpha_c^2\delta_{(a)}^2 + 2k_1\alpha_a\alpha_c\delta_{(c-a)}\right\}\right. \\
& \left.+\alpha_{c-b}^2\left\{\alpha_b^2\delta_{(a)}^2 + \alpha_a^2\delta_{(b)}^2 + 2k_1\alpha_a\alpha_b\delta_{(b-a)}\right\}\right] + \left[16k_1\sigma^{16}\alpha_a^2\alpha_{b-a}\alpha_{c-a}\delta_{(c-b)}\right] \\
& +16\sigma^{16}\left[\alpha_c^2\delta_{(b)}\delta_{(a)}\delta_{(b-a)} + \alpha_b^2\delta_{(a)}\delta_{(c)}\delta_{(c-a)} + \alpha_a^2\delta_{(b)}\delta_{(c)}\delta_{(c-b)}\right] \\
& +16\sigma^{16}k_1\left[\alpha_b\alpha_c\delta_{(b-a)}\delta_{(c-a)} + \alpha_a\alpha_b\delta_{(c-a)}\delta_{(c-b)} + \alpha_a\alpha_c\delta_{(b-a)}\delta_{(c-b)}\right] \\
& +16\sigma^{16}\left[\alpha_a\alpha_b\delta_{(c)}^2\delta_{(b-a)} + \alpha_a\alpha_c\delta_{(b)}^2\delta_{(c-a)} + \alpha_b\alpha_c\delta_{(a)}^2\delta_{(c-b)}\right] \\
& +8\sigma^{16}\alpha_{c-b}\alpha_{b-a}\delta_{(a)}^2\left[\alpha_{c-b}\alpha_{b-a} + 2\delta_{(c-b)}\right] \\
& +16\sigma^{16}\left[\alpha_{c-a}^2\delta_{(b)}\delta_{(a)}\delta_{(b-a)} + \alpha_{b-a}^2\delta_{(a)}\delta_{(c)}\delta_{(c-a)} + \alpha_{c-b}^2\delta_{(a)}\delta_{(b)}\delta_{(b-a)}\right. \\
& \left.+\delta_{(a)}\delta_{(b)}\delta_{(c-b)}\delta_{(c-a)} + \delta_{(a)}\delta_{(c)}\delta_{(b-a)}\delta_{(c-b)} + \delta_{(b)}\delta_{(c)}\delta_{(b-a)}\delta_{(c-a)}\right].
\end{aligned}
$$

Recalling that

$$\beta(\lambda) = Re\left(\alpha(\lambda) - 1\right),$$

and setting

$$\zeta(\lambda) = \mid \alpha(\lambda) - 1 \mid^2,$$

and for any functions $a(\lambda)$, $b(\lambda)$, $c(\lambda)$, $d(\lambda)$, $-\pi \leq \lambda < \pi$,

$$
\begin{aligned}
& C_{ab}(\lambda) = \frac{1}{(2\pi)}\int_{-\pi}^{\pi} a(\omega)b(\lambda-\omega)d\omega,\ -\pi \leq \lambda < \pi, \\
& D_{abc}(\lambda,\mu) = \frac{1}{(2\pi)}\int_{-\pi}^{\pi} a(\omega)b(\lambda-\omega)c(\mu-\omega)d\omega\,, -\pi \leq \lambda < \pi,\ -\pi \leq \mu < \pi \\
& E_{abcd}(\lambda,\mu,\nu) = \frac{1}{(2\pi)}\int_{-\pi}^{\pi} a(\omega)b(\lambda-\omega)c(\mu-\omega)d(\nu-\omega)d\omega, \\
& -\pi \leq \lambda < \pi,\ -\pi \leq \mu < \pi,\ -\pi \leq \nu < \pi,
\end{aligned}
$$

so that each of the integrals is well defined, we obtain the following:

Theorem 12 *Denoting by $Q_{yyy}(\omega_1,\omega_2,\omega_3)$, $-\pi \leq \omega_i < \pi$, $i = 1,2,3$ the fourth-order spectrum (trispectrum) of the y_t, under Assumption B' for $\rho = 0$ and setting for simplicity $\tilde{\omega} = \omega_1+\omega_2+\omega_3$ we have*

$$Q_{yyy}(\omega_1,\omega_2,\omega_3) = \frac{\sigma^{16}}{(2\pi)^3}\{$$

$$8\delta_{(0)}C_{\beta\beta}(\tilde{\omega})C_{\beta\beta}(\omega_2)C_{\beta\beta}(\omega_3)$$

$$+8\Big[C_{\beta\beta}(\omega_3)\Big(C_{\beta\beta}(\omega_1)C_{\zeta\zeta}(\omega_1+\omega_3)+C_{\beta\beta}(\omega_1+\omega_3)C_{\zeta\zeta}(\omega_1)+2\delta_{(0)}D_{\beta\beta\zeta}(\tilde{\omega},\omega_1+\omega_3)\Big)$$

$$C_{\beta\beta}(\omega_2)\Big(C_{\beta\beta}(\omega_3)C_{\zeta\zeta}(\omega_1+\omega_2)+C_{\beta\beta}(\omega_1+\omega_2)C_{\zeta\zeta}(\omega_3)+2\delta_{(0)}D_{\beta\beta\zeta}(\tilde{\omega},\omega_1+\omega_2)\Big)$$

$$C_{\beta\beta}(\omega_3)\Big(C_{\beta\beta}(\omega_1)C_{\zeta\zeta}(\omega_2+\omega_3)+C_{\beta\beta}(\omega_2+\omega_3)C_{\zeta\zeta}(\omega_1)+2\delta_{(0)}D_{\beta\beta\zeta}(\tilde{\omega},\omega_2+\omega_3)\Big)\Big]$$

$$+16\delta_{(0)}C_{\beta\beta}(\tilde{\omega})D_{\beta\beta\zeta}(\omega_2+\omega_3,\omega_2)$$

$$+16\delta_{(0)}\left[C_{\beta\beta}(\omega_1)D_{\zeta\zeta\zeta}(\omega_2+\omega_3,\omega_2)+C_{\beta\beta}(\omega_2)D_{\zeta\zeta\zeta}(\omega_1+\omega_3,\omega_3)+C_{\beta\beta}(\omega_3)D_{\zeta\zeta\zeta}(\omega_1+\omega_2,\omega_2)\right]$$

$$+16\delta_{(0)}\left[E_{\beta\beta\zeta\zeta}(\tilde{\omega},\omega_2,\omega_1+\omega_2)+E_{\beta\beta\zeta\zeta}(\tilde{\omega},\omega_1,\omega_1+\omega_3)+E_{\beta\beta\zeta\zeta}(\tilde{\omega},\omega_1,\omega_1+\omega_2)\right]+$$

$$16\left[C_{\zeta\zeta}(\omega_3)D_{\beta\beta\zeta}(\omega_1+\omega_2,\omega_1)+C_{\zeta\zeta}(\omega_2)D_{\beta\beta\zeta}(\omega_1+\omega_3,\omega_1)+C_{\zeta\zeta}(\omega_1)D_{\beta\beta\zeta}(\omega_2+\omega_3,\omega_2)\right]$$

$$8C_{\beta\beta}(\omega_3)C_{\beta\beta}(\omega_2+\omega_3)C_{\zeta\zeta}(\tilde{\omega})+16C_{\zeta\zeta}(\tilde{\omega})C_{\beta\zeta}(\omega_3)\beta(\omega_2+\omega_3)$$

$$+16\left[C_{\beta\beta}(\omega_3)D_{\zeta\zeta\zeta}(\tilde{\omega},\omega_2)+C_{\beta\beta}(\omega_2)D_{\zeta\zeta\zeta}(\tilde{\omega},\omega_3)+C_{\beta\beta}(\omega_1)D_{\zeta\zeta\zeta}(\tilde{\omega},\omega_3)\right]$$

$$+16\left[E_{\zeta\zeta\zeta\zeta}(\tilde{\omega},\omega_1+\omega_3,\omega_1)+E_{\zeta\zeta\zeta\zeta}(\tilde{\omega},\omega_1+\omega_2,\omega_1)+E_{\zeta\zeta\zeta\zeta}(\tilde{\omega},\omega_1+\omega_2,\omega_2)\right]\Big\}.$$

Proof: Considering Corollary 4 writing

$$\alpha_{|a|}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\beta(\omega)e^{ia\omega}d\omega,$$
$$\delta_b=\frac{1}{2\pi}\int_{-\pi}^{\pi}\zeta(\omega)e^{ib\omega}d\omega,$$

the result follows by equating $cum_{yyy}(a,b,c)$ to

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}Q_{yyy}(\omega_1,\omega_2,\omega_3)e^{i(a\omega_1+b\omega_2+c\omega_3)}d\omega_1 d\omega_2 d\omega_3$$

on the lines of Theorem 6.

$\square$

Chapter 3

The 'one-shock' model: estimation

In this chapter we propose an estimation procedure for the nonlinear moving average model introduced in Chapter 2. We first discuss standard theory of Gaussian estimation. Then we consider possible estimation procedures for our model. First we consider a MLE under Gaussian unobservables. Finally we propose a Gaussian Pseudo Maximum Likelihood Estimator (PMLE) based on the pretense that y_t is Gaussian. Regularity conditions, which the asymptotic theory in the following chapter is based on, are introduced and discussed. Then we introduce a bivariate PMLE, assuming that the (x_t, y_t) are jointly Gaussian, and we perform an efficiency analysis of the two estimators.

3.1 Review on asymptotic theory for Gaussian estimates

Let $u_t, t = 0, \pm 1, \ldots$ be a stationary Gaussian sequence with mean 0 and spectral density $f_u(\lambda)$, $\lambda \in [-\pi, \pi)$. Considering now a parametric family of functions $f_u(\lambda; \beta)$, $\beta \in B$ for a $p \times 1$ vector of parameter β, we shall call $\beta_0 \in B$ the true parameter value such that $f_u(\lambda) = f_u(\lambda; \beta_0)$. Given a sample of T consecutive observations $u = (u_1, \ldots, u_T)'$ the ML estimate of β will be defined by

$$\hat{\beta} = argmin_{\beta \in B} Q_T(\beta),$$

where

$$Q_T(\beta) = log(2\pi) + \frac{1}{T}log \mid A_T(\beta) \mid + \frac{1}{T}u'A_T^{-1}(\beta)u, \tag{3.1}$$

is $-2/T$ times the log (Gaussian) likelihood and $A_T(\beta)$ is such that $\{A_T(\beta)\}_{a,b} = \int_{-\pi}^{\pi} e^{i(a-b)\omega} f_u(\omega;\beta)d\omega$ and the set B is assumed to be compact.

Under regularity conditions as $T \to \infty$

$$T^{1/2}(\hat{\beta} - \beta_0) \to_d \mathcal{N}_p(0, B^{-1}(\beta_0)), \tag{3.2}$$

with

$$B(\beta) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial log f_u(\omega;\beta)}{\partial\beta}\frac{\partial log f_u(\omega;\beta)}{\partial\beta'}d\omega, \tag{3.3}$$

where the asymptotically efficient property yields

$$\lim_{T\to\infty}\frac{1}{T}E(\frac{\partial Q_T(\beta_0)}{\partial\beta}\frac{\partial Q_T(\beta_0)}{\partial\beta'}) = B(\beta_0). \tag{3.4}$$

Asymptotically equivalent inference is obtained when considering the Gaussian likelihood in the frequency domain, introduced by Whittle (1962)

$$\bar{Q}_T(\beta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} log f_u(\omega;\beta)d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I_u(\omega)}{f_u(\omega;\beta)}d\omega, \tag{3.5}$$

where

$$I_u(\lambda) = \frac{1}{2\pi T} \mid \sum_{t=1}^{T} u_t e^{it\lambda} \mid^2 .$$

This approximation of the Gaussian likelihood, known as the Whittle function, is much easier to handle particularly because there is no need anymore to evaluate $A_T^{-1}(\beta)$ relying on the fact that (cf. Grenander and Szego (1958))

$$\lim_{T\to\infty} det\, A_T(\beta)^{1/T} = 2\pi exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} log f_u(\omega;\beta)d\omega\right).$$

On the other hand, from (3.5), one now needs to calculate two integrals for every evaluation of the objective function but these are routinely approximated by sums over the Fourier frequency. The surprising thing is that (3.2), viz. the same $T^{1/2}$ norming and the same asymptotic distribution and the efficiency property (3.4) hold for the short memory case (Mann and Wald 1943) (Whittle 1962) as well as for the long memory case (Fox and Taqqu 1986) (Dahlhaus 1989). In the former case the regularity conditions state the power spectrum $f_u(\lambda;\beta)$ to be at least continuous in the λ, differently from the latter case where the power spectrum has a discontinuity (of type one) at the origin.

In order to understand why this is so, let us consider the following result on spectral estimates where (3.2) is a particular case of

$$\frac{T^{1/2}}{4\pi}\int_{-\pi}^{\pi} g(\lambda)\,[I_u(\omega)-f_u(\omega)]\,d\omega \rightarrow_d N(0,h(g)), \quad \text{as } T\rightarrow\infty, \tag{3.6}$$

and where

$$h(g)=\frac{1}{4\pi}\int_{-\pi}^{\pi} g^2(\omega)f_u^2(\omega;\beta_0)d\omega,$$

for some function $g(\omega)$, $-\pi \leq \omega < \pi$. Obviously proper regularity conditions on the function $g(.)$ and the power spectrum $f_u(.)$ are needed in order to achieve (3.6), the minimum condition being evenness of $g(.)$.

Many different statistics are embedded within the spectral estimate framework, by a proper choice of the $g(.)$ function. For instance setting $g(\lambda) = cos(k\lambda)$, $k = 0, \pm1, \ldots$ one obtains the result for sample autocovariances (Hannan 1976) and for $g(\lambda) = g_T(\lambda - \mu)$, with $g_T(.)$ being a kernel one gets a spectral density estimate at frequency $\mu \in [-\pi, \pi)$.

Finally when $g(\lambda) = \frac{\partial}{\partial\beta} f_u^{-1}(\lambda;\beta_0)$, we are back in the case of Gaussian estimation.

In the short memory case, assuming that $f_u(.)$ is bounded away from zero, we can take $g(.)$ to be at least continuous but a certain degree of complementarity in smoothness between $g(.)$ and $f_u(.)$ is allowed for. In fact a pole of $f_u(.)$ can match a zero of $g(.)$ of a suitable order so that the asymptotic variance $h(g)$ is finite and (3.6) holds. This is exactly the mechanism which yields (3.2) for long memory processes, where the result goes through with $f_u(\lambda;\beta_0) = O(\lambda^{-\alpha})$ and $g(\lambda) = O(\lambda^{\alpha})$ as $\lambda \rightarrow 0^+$ where $0 < \alpha < 1$.

Result (3.2) was generalized in the case where u_t is a linear process in i.i.d. variates as well as in martingale differences both in the short memory case (Hannan 1973) (Hosoya and Taniguchi 1982) and the long memory case (Giraitis and Surgailis 1990a) (Heyde and Gay 1993), the latter being valid both for short and long memory processes.

In the weak memory case it is well known that when the innovation variance does not depend on the parameter β, namely (cf. Hannan (1970))

$$\int_{-\pi}^{\pi} log f_u(\omega;\beta)d\omega = 0,$$

the PMLE is as efficient as in the Gaussian case. In fact in this case the asymptotic covariance matrix is independent of the fourth cumulant of the innovation in the short memory (Whittle

1962) (Hannan 1973) (Dunsmuir 1979) and in the long memory case (Giraitis and Surgailis 1990a).

The difference concerns the approach used to establish (3.2). Long memory rules out all the CLT results based on some form of mixing condition (McLeish 1975). Indeed the approach pioneered by Fox and Taqqu (1986) (1987) is based on showing that the cumulants of order greater than 3 of the right hand side of (3.6) were converging to zero as the sample size goes to infinity, making use of the general expression for the cumulant of a quadratic form in Gaussian variates (Grenander and Szego 1958)

$$cum_k(\sum_{s,t=1}^{T} b_{(t-s)}u_s u_t) = 2^{k-1}(k-1)!tr(A_T B_T), \tag{3.7}$$

where $A_T = A_T(\beta_0)$ and B_T is a Toeplitz matrix made of the coefficients $b_{(s-t)}$, $s,t = 1,\ldots,T$.

3.2 Exact versus pseudo MLE

3.2.1 The exact MLE under Gaussian unobservables

The long memory parameterization (assumption A_1) implies that it would be implausible, albeit possible, to estimate this nonlinear moving average by some method of moments, differently from the weak memory case of Robinson (1977). In fact it is not be clear which moments should one look at and how the efficiency of the estimator should be affected.

On the other hand, we will now show how an exact MLE although fully efficient does not seem to be the preferable approach to be considered (cf. Robinson and Zaffaroni (1996a)). In this case we need to make some distributional assumptions on the ϵ_t, a much stronger assumption than B'. Assuming Gaussian $\{\epsilon_t\}$

$$pdf(\epsilon_1,\ldots,\epsilon_T) = \frac{1}{\sigma^T(2\pi)^{T/2}} exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{\infty}\epsilon_t^2\right), \tag{3.8}$$

so then we get

$$pdf(x_1,\ldots,x_T) = \frac{1}{\sigma^T(2\pi)^{T/2}} exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{T}\frac{(x_t-\mu)^2}{h_{t-1}^2}\right) \mid \frac{\partial\epsilon}{\partial x'} \mid, \tag{3.9}$$

where $x = (x_1, \ldots, x_T)'$, $\epsilon = (\epsilon_1, \ldots, \epsilon_T)'$.

But the Jacobian matrix $\mid \frac{\partial \epsilon}{\partial x'} \mid$ is triangular because

$$\frac{\partial \epsilon_t}{\partial x_s} = \begin{cases} 1/h_{t-1}, & s = t \\ 0, & s > t \end{cases}$$

so that

$$\mid \frac{\partial \epsilon}{\partial x'} \mid = (\prod_{i=1}^{T} h_{t-i})^{-1}. \tag{3.10}$$

This yields the log likelihood

$$log\, pdf(x_1, \ldots, x_T) = -T log\, \sigma - \frac{T}{2} log\,(2\pi) - \frac{1}{4\sigma^2} \sum_{i=1}^{T} \frac{(x_t - \mu)^2}{h_{t-1}^2} - \sum_{t=1}^{T} log\, h_{t-1}.$$

Obviously, even if the ϵ_t and so the h_t are Gaussian, by the fact that the product of normal random variables is not normal (Aroian 1947), the x_t are not.

As in Robinson (1977), in order to evaluate it we would need to evaluate h_{t-1} in terms of the x_t's, that is, writing the model as a nonlinear autoregression. For instance, setting,

$$\bar{h}_{t-1} = \begin{cases} \rho, & t = 1 \\ \rho + \alpha_1 \bar{\epsilon}_{t-1} + \ldots + \alpha_{t-1} \bar{\epsilon}_1, & t > 1 \end{cases} \tag{3.11}$$

where $\bar{\epsilon}_t = \epsilon_t I(t > 0)$, $I(\,.\,)$ being the indicator function, and from

$$\epsilon_t = (x_t - \mu)/h_{t-1}, \tag{3.12}$$

we obtain

$$\begin{aligned} &\bar{h}_0 = \rho, \\ &\bar{h}_1 = \rho + \alpha_1 \frac{x_1 - \mu}{\bar{h}_0} = \rho + \frac{x_1 - \mu}{\rho}, \\ &\bar{h}_2 = \rho + \alpha_1 \bar{\epsilon}_2 + \alpha_2 \bar{\epsilon}_1 = \rho + \alpha_1 \frac{x_2 - \mu}{\bar{h}_1} + \alpha_2 \frac{x_1 - \mu}{\bar{h}_0} \\ &= \rho + \alpha_1 \frac{\rho(x_2 - \mu)}{(\rho^2 + \alpha_1(x_1 - \mu))} + \alpha_2 \frac{x_1 - \mu}{\rho}, \end{aligned}$$

and so on. Of course one would need to establish that this invertibility condition holds such that, at least, as $t \to \infty$

$$h_t - \bar{h}_t \to_p 0 \tag{3.13}$$

so to be able to perform correct asymptotical inference. As was noticed by Granger and Andersen (1978) the theory of linear processes is not readily extendable to the nonlinear case including our nonlinear moving average case. The same observation applies for all the ARCH-type models where there are still no formal results establishing the invertibility condition, routinely used when estimating and forecasting.

Moreover, even if we can expect that the truncation (used in order to evaluate the likelihood substituting in the barred residual $\bar{\epsilon}_t$) is asymptotically negligible due to the long memory nature of the process, we can expect this 'initial condition' effect to die away very slowly in function of the memory of the process $\{y_t\}$. Lumsdaine (1996) showed that for the GARCH(1,1) the effect of conditioning on the value of the conditional variance of the process at time 0 (before the process was sampled) dies away faster than $T^{-1/2}$ and actually we conjecture it to die exponentially fast. This of course is a consequence of the short memory of the process.

3.2.2 The Gaussian PMLE

We propose a PMLE based on a Gaussian likelihood as if the process $\{y_t\}$ were Gaussian without of course forgetting that it cannot be so, y_t being always non-negative by construction. The Gaussian likelihood is based by construction on all the autocovariance function of y_t and therefore we expect it to be very sensitive to the degree of dependence, in a second order sense, of the process. The success of the Gaussian likelihood principle is also witnessed in a semi-parametric framework (Robinson 1995a).

We will consider the frequency domain approximation (3.5) so that we will lose dependency from initial conditions as well as from the invertibility condition, which the exact MLE was based on.

Let us assume that $\{\alpha_j(\theta) : j = 1, 2, \ldots\}$ represents a sequence of functions of a $p \times 1$ vector parameter $\theta \in \Theta \subset R^p$.

We define the normalized parameters $\bar{\mu} = \mu/\sigma^2$ and $\bar{\rho}^2 = \rho^2/\sigma^2$, and therefore we shall call ψ the $(p+2) \times 1$ vector of parameters made by

$$\psi = (\bar{\mu}, \bar{\rho}^2, \theta')'.$$

Hence ψ will parameterize a family of functions $K_y(\lambda; \psi)$, $\lambda \in [-\pi, \pi)$, $\psi \in \Psi$ obtained by replacing the α_i by the $\alpha_i(\theta)$ in (2.27).

Note that we did not include σ^8, the variance of the unobservables ϵ_t, in ψ. In fact given

$$f_y(\lambda;\psi) = \sigma^8\, K_y(\lambda;\psi),\ -\pi \leq \lambda < \pi,$$

we are able to factorize σ^2, and thus to concentrate it out from the objective function. As shown below, this implies that we do obtain a closed-form expression for its estimator, making unnecessary the distinction between the true value σ^2 and any admissible values.

We shall call $\psi_0 \in \Psi$ the true parameter so that in particular

$$f_y(\lambda) = \sigma^8 K_y(\lambda;\psi_0),\ \ \alpha_j = \alpha_j(\theta_0),\ j = 1, 2, \ldots. \tag{3.14}$$

We will assume that ψ_0 lies in the interior of $\Psi \subset R^{p+2}$, Ψ being assumed compact.

Two possible objective functions, asymptotically equivalent, are considered:

$$\overline{Q}_T(\sigma^2,\psi) \ = \ (1/2\pi)\int_{-\pi}^{\pi}\left[log(\sigma^8 K_y(\lambda;\psi)) + \frac{I(\lambda)}{\sigma^8 K_y(\lambda;\psi)}\right]d\lambda\,, \tag{3.15}$$

$$\widehat{Q}_T(\sigma^2,\psi) \ = \ (1/T)\sum_{j=1}^{T-1}\left[log(\sigma^8 K_y(\lambda_j;\psi)) + \frac{I(\lambda_j)}{\sigma^8 K_y(\lambda_j;\psi)}\right], \tag{3.16}$$

where $I(\omega)$ is the the periodogram,

$$I(\omega) = \frac{\mid \sum_{t=1}^{T}(y_t - \bar{y})e^{it\omega}\mid^2}{2\pi T},\ \ \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t, \tag{3.17}$$

$\lambda_j = \frac{2\pi j}{T}$ denotes the Fourier frequency and the zero frequency is skipped due to mean correction.

With respect to $\widehat{Q}_T(\sigma^2,\psi)$ we concentrate the parameter σ^8 out. With the index j running over $1 \leq t \leq T-1$, from

$$\widehat{Q}_T(\sigma^2,\psi) = (1/T)\sum_j\left[log\sigma^8 + logK_y(\lambda_j;\psi) + \frac{I(\lambda_j)}{\sigma^8 K_y(\lambda_j;\psi)}\right],$$

differentiating with respect to σ^8 and equating the result to zero yields

$$\frac{\partial\,\widehat{Q}_T(\hat{\sigma}^2,\psi)}{\partial\sigma^8} = (1/T)\sum_j\left(\frac{1}{\hat{\sigma}^8} - \frac{I(\lambda_j)}{\hat{\sigma}^{16}K_y(\lambda_j;\psi)}\right) = 0\,,$$

obtaining

$$\hat{\sigma^8} = \hat{\sigma^8}_T(\psi) = (1/T)\sum_j \frac{I(\lambda_j)}{K_y(\lambda_j;\psi)}. \tag{3.18}$$

Plugging it in the first expression, one obtains the concentrated Pseudo likelihood

$$\widehat{Q}_T(\psi) = log\left[(1/T)\sum_j \frac{I(\lambda_j)}{K_y(\lambda_j;\psi)}\right] + \frac{\sum_j log\, K_y(\lambda_j;\psi)}{T} + 1. \tag{3.19}$$

Likewise we can write the concentrated likelihood in terms of the first kind of objective function, i.e. $\bar{Q}_T(\sigma^2,\psi)$, replacing the summations with integrals, and obtaining then

$$\bar{\sigma}_T^8(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I(\omega)}{K_y(\omega;\psi)}d\omega, \tag{3.20}$$

$$\bar{Q}_T(\psi) \stackrel{def}{=} log(\bar{\sigma}_T^8) + \frac{1}{2\pi}\int_{-\pi}^{\pi} logK_y(\omega;\psi)d\omega + 1. \tag{3.21}$$

Let us define the pseudo maximum likelihood estimators (PMLE) $\hat{\psi}_T$ and $\bar{\psi}_T$ by

$$\hat{\psi}_T = argmin_{\psi\in\Psi}\widehat{Q}_T(\psi), \tag{3.22}$$

$$\bar{\psi}_T = argmin_{\psi\in\Psi}\bar{Q}_T(\psi). \tag{3.23}$$

The potentially most cumbersome aspect from a computational point of view is induced by the two convolutions contained in the expression for $f_y(\lambda;\psi)$ which are a result of the nonlinearity. To indicate how to deal with this (cf. Robinson and Zaffaroni (1996b)) let us suppose that we wish to evaluate

$$\tilde{h}(\mu) = \int_{-\pi}^{\pi} \tilde{f}(\lambda)\tilde{g}(\mu-\lambda)d\lambda,$$

for some functions $\tilde{f}(\lambda),\tilde{g}(\lambda)$. By a standard result in harmonic analysis the Fourier transform of $\tilde{h}(\lambda)$ is the product of the Fourier transforms of $\tilde{f}(\lambda),\tilde{g}(\lambda)$. Thus to approximate $\tilde{h}(\mu)$ we can use the fast Fourier transform in order to convolve $\tilde{f}(\lambda_j),\tilde{g}(\lambda_j), j = 1,\ldots,T-1$, taking the product of the results and then deconvolving.

'Time domain' assumptions

To develop the asymptotic theory of the PMLE, we introduce the following assumptions which extend the previous Assumptions A to the sequence $\alpha_i(\theta)$.

Assumptions A'

For any $\theta, \theta^* \in \Theta$ *and* $i = 1, \ldots, p$

A_1' $\quad \alpha_j(\theta) \sim K j^{d(\theta)-1}, \; 0 <| K |< \infty, \; as \; j \to \infty$

where the function $d : \Theta \to (0, 1/2)$ *is continuous.*

A_2' $\quad | \alpha_j(\theta) - \alpha_{j+1}(\theta) | \leq K \frac{| \alpha_j(\theta) |}{j}, \forall j > J, \; some \; J < \infty, 0 < K < \infty,$

A_3' $\quad \frac{\partial \alpha_j(\theta)}{\partial \theta_i} \sim K log\,(j) j^{d(\theta)-1}, 0 <| K |< \infty, \; as \; j \to \infty,$

A_4' $\quad | \frac{\partial \alpha_j(\theta)}{\partial \theta_i} - \frac{\partial \alpha_{j+1}(\theta)}{\partial \theta_i} | \leq \frac{K}{j} | \frac{\partial \alpha_j(\theta)}{\partial \theta_i} |, \forall j > J, \; some \; J < \infty, 0 < K < \infty,$

$A_5(k)'$ $\quad \frac{\partial^k \alpha_j}{\partial \theta_{i_1} \ldots \partial \theta_{i_k}} \sim K log^k(j) j^{d(\theta)-1} \; as \; j \to \infty,$

$A_6(k)'$ $\quad | \frac{\partial^k \alpha_j}{\partial \theta_{i_1} \ldots \partial \theta_{i_k}} - \frac{\partial^k \alpha_{j+1}}{\partial \theta_{i_1} \ldots \partial \theta_{i_k}} | \leq \frac{K}{j} | \frac{\partial^k \alpha_j}{\partial \theta_{i_1} \ldots \partial \theta_{i_k}} | \forall j > J, \; some \; J < \infty, \; 0 < K < \infty,$

where $i_u \in \{1, 2, \ldots, p\}, u = 1, 2, \ldots, k,$

A_7' $\quad \alpha_j(\theta) \neq \alpha_j(\theta^*),$ *for any integer* j *and* $\theta \neq \theta^*$.

Remarks: (1) Assumptions A_1' imply that the $\alpha_i(\theta)$ behave asymptotically as the Wold coefficients of a stationary $ARFIMA(p, d, q)$, thus potentially imparting long memory to the y_t. By no means are we restricting the short run dynamics so that A_1' is perfectly compatible for instance with seasonal and asymmetric effects. Imposing an exact rate rather than an upper bound (cf. Robinson and Zaffaroni (1996a)) plays a crucial role in the asymptotic results, for instance with A_7' to guarantee identifiability of the model. Assumption A_2' implies that the $\alpha_i(\theta)$ behaves as QMC sequence. This will determine the behaviour near the zero frequency as well as imposing an approximate Lipschitz condition (Zygmund 1977) over the interval $(0, \pi]$ for $f_y(\lambda; \psi)$. Assumptions A_3', A_4', imposing the same degree smoothness to $\frac{\partial}{\partial \psi} f_y(\lambda; \psi)$, allow us to make use of limit results for Toeplitz matrices (Fox and Taqqu 1986) in the proof of the CLT of the PMLE. Furthermore, with $A_5'(2)$, $A_6'(2)$, we can apply the Delta method and establish the limit (a.s.) of the hessian matrix.

(2) 'Time domain' regularity assumptions are not common in the long memory parametric (Fox and Taqqu 1986) (Fox and Taqqu 1987) (Giraitis and Surgailis 1990a) and semi-nonparametric literature (Robinson 1995b) (Robinson 1995a). Indeed regularity conditions are usually expressed in terms of a certain degree of smoothness in the power spectrum of the underlying

process[1], henceforth defined as the 'usual' assumptions. But whereas in the latter case frequency domain assumptions represent a natural choice it is not as well motivated in the parametric framework, except in the case of linear models when efficient inference can be drawn from the second order properties of the process. On the other hand, it might be very difficult in general, within a specific nonlinear parameterization, to check that the family of power spectra driven by such parameterization would satisfy the 'usual' assumptions. In our nonlinear parametric framework the statistical properties of the model are completely determined by the behaviour of the unobservables ϵ_t and the coefficients $\alpha_i(\theta)$. In this respect it seems natural to impose regularity conditions as stated by Assumptions A' above. These conditions appear very easy to check and as an example we show in the next section how the $ARFIMA(0,d,0)$ Wold coefficients do satisfy Assumptions A'. Our 'time domain' Assumption A' imply certain but not all of the 'usual' assumptions. In this respect our assumptions are sometimes stronger and sometimes weaker, again a by-product of the difference of our model from the linear framework.

An example: the ARFIMA(0,d,0) coefficients.

Let us consider the following parameterization,

$$d(\theta) = \theta ,$$

and

$$\alpha_i(\theta) = \begin{cases} \prod_{k=1}^{i} \frac{k-1+\theta}{k} & , i \geq 1 , \\ 1 & , i = 0 . \end{cases} \tag{3.24}$$

In terms of transfer function these coefficients correspond to the following

$$\alpha(\lambda) = (1 - e^{i\lambda})^{-\theta} .$$

Thus in order to confirm the capability of these coefficients we need to check that Assumptions A'_1 to A'_7 are satisfied.

Regarding Assumption A'_1, from using Stirling's formula (Brockwell and Davis 1987) it follows that as $i \to \infty$,

$$\alpha_i(\theta) \sim K \, i^{\theta - 1} ,$$

[1]Fox and Taqqu (1986, Remark at p.529) state how in certain cases a Tauberian theorem (as Proposition 2) can be applied in order to get the asymptotic behaviour of the Fourier coefficients of the power spectrum and its derivatives.

where the constant $K = K(\theta)$ does not depend on i.

From (3.24) we get

$$\alpha_i(\theta) - \alpha_{i+1}(\theta) = \alpha_i(\theta)\left(\frac{1-\theta}{i+1}\right),$$

and thus A_2' is trivially satisfied.

Concerning A_3' we get

$$log\, \alpha_i(\theta) = \sum_{k=1}^{i} log(\frac{k-1+\theta}{k}).$$

Differentiating we get

$$\frac{\partial log\, \alpha_i(\theta)}{\partial\theta} = \sum_{k=1}^{i} log(\frac{1}{k-1+\theta}).$$

Then from (Knopp 1961, Theorem 10 pp. 224) as $N \to \infty$ $\sum_{k=1}^{N} \frac{1}{k} \sim log(N)$ so that as $i \to \infty$

$$\frac{\partial\alpha_i(\theta)}{\partial\theta} \sim K\, i^{\theta-1} log(i).$$

Likewise from

$$\frac{\partial log\, \alpha_i(\theta)}{\partial\theta} - \frac{\partial log\, \alpha_{i+1}(\theta)}{\partial\theta} = log(i+\theta),$$

we get, as $i \to \infty$,

$$\frac{\partial\alpha_i(\theta)}{\partial\theta} - \frac{\partial\alpha_{i+1}(\theta)}{\partial\theta} = \frac{1-\theta}{i+\theta}\frac{\partial\alpha_{i+1}(\theta)}{\partial\theta} + \alpha_i(\theta) log(i+\theta) \sim K\alpha_i(\theta)\, log(i).$$

By an identical argument it is straightforward to show that Assumptions $A_5(k)'$ and $A_6(k)'s$ for any integer k are satisfied.

Finally regarding A_7' and imposing the condition

$$\alpha_k(\theta) = \alpha_k(\theta^*), \tag{3.25}$$

for any integer $k \geq 1$ yields

$$(\theta^k - (\theta^*)^k)c_k + (\theta^{k-1} - (\theta^*)^{k-1})c_{k-1} + \ldots + c_0 = 0,$$

where the coefficients c_i, $i = 0, \ldots, k$ are positive functions of the integers $1, 2, \ldots, k$ only. Hence condition (3.25) is true if and only if $\theta = \theta^*$.

3.3 The bivariate Gaussian PMLE

In general there is always a loss in efficiency in estimating the model by the Gaussian PMLE with respect to the exact MLE, when the distribution is known. In this section we consider a possible way to achieve some efficiency gain without having to impose some distributional assumptions.

We propose a Gaussian PMLE as if the bivariate process $z_t = (x_t, y_t)'$ is multivariate normal. In this way we should obtain an efficiency gain with respect to the univariate PMLE based on the y_t only, because we consider the information concerning the x_t and the cross relations between the x_t and the y_t.

Let us consider for simplicity the objective function in the 'integral form'. The bivariate Gaussian likelihood for the z_t with power spectrum matrix $f_z(\lambda)$ is

$$\bar{Q}_{2,T}(\sigma^2, \psi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Big(log \mid f_z(\lambda; \sigma^2, \psi) \mid + tr(f_z^{-1}(\lambda; \sigma^2, \psi) I(\lambda) \Big) \, d\lambda, \tag{3.26}$$

with

$$f_z(\omega; \sigma^2, \psi) = \begin{pmatrix} \sigma^4 K_x(\omega; \psi) & \sigma^6 K_{xy}(\omega; \psi) \\ \sigma^6 K_{yx}(\omega; \psi) & \sigma^8 K_y(\omega; \psi) \end{pmatrix}, f_z(\lambda) = f_z(\lambda; \psi_0),$$

$$I(\omega) = \begin{pmatrix} I_x(\omega) & I_{xy}(\omega) \\ I_{yx}(\omega) & I_y(\omega) \end{pmatrix},$$

where for any sequences a_t, b_t, $t = 1, \ldots, T$

$$I_{ab}(\omega) = \frac{1}{2\pi T} \sum_{s,t=1}^{T} e^{-i\omega(t-s)} (a_t - \bar{a})(b_s - \bar{b}), \;\; I_{ab}(\omega) = I_a(\omega) \text{ when } a = b,$$

$\bar{a}$, $\bar{b}$ denoting the sample means, and $K_x(\omega; \psi)$, $K_{xy}(\omega; \psi)$ are obtained, similarly to $K_y(\omega; \psi)$, from Theorems 5 and 9 replacing the α_i by the $\alpha_i(\theta)$ and factorizing for a suitable power of σ^2.

It follows that

$$f_z^{-1}(\omega; \sigma^2, \psi) = \frac{1}{\sigma^{12} |K_z(\omega; \psi)|} \begin{pmatrix} \sigma^8 K_y(\omega\psi) & -\sigma^6 K_{xy}(\omega) \\ -\sigma^6 K_{yx}(\omega) & \sigma^4 K_x(\omega) \end{pmatrix}, \tag{3.27}$$

where

$$K_z(\omega; \psi) = \begin{pmatrix} K_x(\omega; \psi) & K_{xy}(\omega; \psi) \\ K_{yx}(\omega; \psi) & K_y(\omega; \psi) \end{pmatrix}.$$

Setting for simplicity $\tau = \sigma^2$, we get

$$\begin{aligned}\bar{Q}_{2,T}(\tau,\psi) \quad &= 6\log\tau + \frac{1}{2\pi}\int_{-\pi}^{\pi} log|K_z(\lambda;\psi)|d\lambda \\ &+\frac{1}{\tau^2}b_1(\psi) - \frac{1}{\tau^3}b_2(\psi) + \frac{1}{\tau^4}b_3(\psi),\end{aligned} \tag{3.28}$$

where

$$b_1(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{K_y(\lambda;\psi)I_x(\lambda)}{|K_z(\lambda;\psi)|}d\lambda,$$

$$b_2(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{K_{xy}(\lambda;\psi)I_{yx}(\lambda) + K_{yx}(\lambda;\psi)I_{xy}(\lambda)}{|K_z(\lambda;\psi)|}d\lambda,$$

$$b_3(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{K_x(\lambda;\psi)I_y(\lambda)}{|K_z(\lambda;\psi)|}d\lambda.$$

Then differentiating with respect to τ yields

$$\frac{\partial\bar{Q}_{2,T}(\hat{\tau},\psi)}{\partial\tau} = \frac{6}{\hat{\tau}} - \frac{2}{\hat{\tau}^3}b_1(\psi) + \frac{3}{\hat{\tau}^4}b_2(\psi) - \frac{4}{\hat{\tau}^5}b_3(\psi) = 0, \tag{3.29}$$

which is equivalent to the following fourth order equation in $\hat{\tau} = \hat{\tau}(\psi)$:

$$6\hat{\tau}^4 - 2b_1(\psi)\hat{\tau}^2 + 3b_2(\psi)\hat{\tau} - 4b_3(\psi) = 0.$$

From the sign of the coefficient[2] we can conclude that there is one positive consistent root which is possible to obtain explicitly by a standard result of linear algebra . Plugging $\hat{\tau}$ in (3.28) we get

$$\begin{aligned}\bar{Q}_{2,T}(\psi) \quad &= 6\log\hat{\tau} + \frac{1}{2\pi}\int_{-\pi}^{\pi} log|K_z(\lambda;\psi)|d\lambda \\ &+3 + \frac{1}{2\hat{\tau}^3}b_2(\psi) - \frac{1}{\tau^4}b_3(\psi),\end{aligned} \tag{3.30}$$

where we use (3.29) to concentrate out the term involving $b_1(\psi)$. This is one of the possible ways of using (3.29) which might be preferable from a computational point of view, because $b_1(\psi)$ involves the highest number of integrals.

We define the bivariate PMLEs, $\bar{\psi}_{2,T}$ and $\hat{\psi}_{2,T}$ by

$$\bar{\psi}_{2,T} \quad = argmin_{\psi\in\Psi}\bar{Q}_{2,T}(\psi). \tag{3.31}$$

$$\hat{\psi}_{2,T} \quad = argmin_{\psi\in\Psi}\hat{Q}_{2,T}(\psi), \tag{3.32}$$

where $\hat{Q}_{2,T}$ is obtained by substituting sums with integrals in $\bar{Q}_{2,T}$.

[2]In fact $b_1(\psi)$ and $b_3(\psi)$ are always positive while $b_2(\psi)$ becomes a.s. positive as $T \to \infty$.

3.4 Asymptotic efficiency of the bivariate PMLE vs. (univariate) PMLE

At first let us state the asymptotic distribution of the normal and bivariate PMLE. We will see that it seems a difficult task to obtain some efficiency comparison given the structure of the asymptotic covariance matrices involving complicated trispectra expressions. Thus we will consider 'time domain' Gaussian likelihoods which appear to be much easier to manipulate in order to get an answer to our problem.

3.4.1 Asymptotic distribution of the PMLEs with 'frequency domain' Gaussian likelihood

In Chapter 4 we establish the following result, valid also for $\hat{\psi}_T$.

Henceforth $D_E(\theta) = \frac{\partial}{\partial\phi} vecE(\phi)$ for some matrix $E(\phi)$ function of a vector ϕ, denotes the gradient with respect to ϕ.

Theorem 14

Under Assumptions $A_1', A_2', A_3', A_4', A_5('2), A_6'(2), A_7', B'$ *and Gaussian* ϵ_t *as* $T \to \infty$

$$T^{1/2}(\bar{\psi}_T - \psi_0) \to_d \mathcal{N}_p\left(0, M^{-1}VM^{-1}\right), \tag{3.33}$$

where

$$M(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} N(\lambda;\psi)N(\lambda;\psi)'d\lambda, \tag{3.34}$$

with

$$N(\lambda;\psi) = \frac{\partial log\left(K_y(\lambda;\psi)\right)}{\partial\psi} - \left[\int_{-\pi}^{\pi} \frac{\partial log\left(K_y(\omega;\psi)\right)}{\partial\psi}d\omega\right], \tag{3.35}$$

and

$$\begin{aligned} V(\psi) &= 4\pi\int_{-\pi}^{\pi} \frac{\partial K_y^{-1}(\omega;\psi)}{\partial\psi}\frac{\partial K_y^{-1}(\omega;\psi)}{\partial\psi'}K_y^2(\omega)d\omega \\ &+ 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\partial K_y^{-1}(\omega_1;\psi)}{\partial\psi}\frac{\partial K_y^{-1}(\omega_2;\psi)}{\partial\psi'}\bar{Q}_{yyy}(-\omega_1,\omega_2,-\omega_2)d\omega_1 d\omega_2, \end{aligned}$$

and where $M = M(\psi_0)$, $V = V(\psi_0)$ *and* $\bar{Q}_{yyy}(.,.,.) = 1/\sigma^{16} Q_{yyy}(.,.,.)$, *viz. normalizing the trispectrum with respect to the parameter* σ^2 .

We conjecture but we do not provide any proof that under the same assumptions of Theorem 14,

$$T^{1/2}(\bar{\psi}_{2,T} - \psi_0) \to_d \mathcal{N}_p \left(0, M_2^{-1} V_2 M_2^{-1}\right) , \tag{3.36}$$

where $M_2 = M_2(\psi_0)$, $V_2 = V_2(\psi_0)$ and

$$\begin{aligned} M_2(\psi) &= plim_{T\to\infty} \frac{\partial^2 \bar{Q}_{2,T}(\psi)}{\partial\psi\partial\psi'} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} D'_{\hat{f}'_z}(\omega;\psi) \left(f'^{-1}_z(\omega;\psi) \otimes f_z^{-1}(\omega;\psi)\right) D_{\hat{f}_z}(\omega;\psi)\, d\omega\,, \end{aligned} \tag{3.37}$$

$$\begin{aligned} V_2(\psi) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} D'_{K'_z}(\omega;\psi) \left(K'^{-1}_z(\omega;\psi) \otimes K_z^{-1}(\omega;\psi)\right) D_{K_z}(\omega;\psi)\, d\omega \\ &+ \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} P(\omega_1;\psi)\, \Phi_{I,I}(-\omega_1,\omega_2,-\omega_2) P'(\omega_2;\psi)\, d\omega_1\, d\omega_2\,, \end{aligned} \tag{3.38}$$

$$D_{\hat{f}_z}(\omega;\psi) = plim_{T\to\infty}\frac{\partial vec \hat{f}_z(\omega;\psi)}{\partial\psi}\,,\ \hat{f}_z(\omega;\psi) = \begin{pmatrix} \hat{\tau}^2(\psi) K_x(\omega;\psi) & \hat{\tau}^3(\psi) K_{xy}(\omega;\psi) \\ \hat{\tau}^3(\psi) K_{yx}(\omega;\psi) & \hat{\tau}^4(\psi) K_y(\omega;\psi) \end{pmatrix},$$

$$P(\omega;\psi) = D'_{K'_z}(\omega;\psi) \left(K_z^{-1\prime}(\omega;\psi) \otimes K_z^{-1}(\omega;\psi)\right),$$

$$\Phi_{I,I}(\omega_1,\omega_2,\omega_3) = \begin{pmatrix} \phi_{1111} & \phi_{1121} & \phi_{1112} & \phi_{1122} \\ \phi_{2111} & \phi_{2121} & \phi_{2112} & \phi_{2122} \\ \phi_{1211} & \phi_{1221} & \phi_{2112} & \phi_{2122} \\ \phi_{2211} & \phi_{2221} & \phi_{2212} & \phi_{2222} \end{pmatrix} (\omega_1,\omega_2,\omega_3)\,,$$

and $\phi_{a,b,c,d}(\omega_1,\omega_2,\omega_3)$ denotes the trispectrum in the four variates $\{(\frac{x_t}{\sigma^2})^a, (\frac{x_t}{\sigma^2})^b, (\frac{x_t}{\sigma^2})^c, (\frac{x_t}{\sigma^2})^d\}$, with $a,b,c,d \in \{1,2\}$.

The non-Gaussian part of the asymptotic covariance matrix, (3.38), is also equivalent to the following

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} Q(\omega_1;\psi)\, \Phi_{I',I'}(-\omega_1,\omega_2,-\omega_2) Q'(\omega_2;\psi)\, d\omega_1\, d\omega_2\,, \tag{3.39}$$

where defining the permutation matrix [3]:

$$H_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

we get

$$Q'(.) = H_4 P'(.)\,,\; \Phi_{I',I'}(.,.,.) = H_4 \Phi_{I,I}(.,.,.) H_4'\,.$$

Also (3.38) is equal to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} P(\omega_1;\psi)\,\Phi_{I',I}(-\omega_1,\omega_2,-\omega_2) Q'(\omega_2;\psi)\,d\omega_1\,d\omega_2\,, \tag{3.40}$$

with

$$\Phi_{I',I}(.,.,.) = \Phi_{I',I'}(.,.,.) H_4'\,,$$

where the fact that the x_t are real implies that the matrix $\Phi_{I,I}$ is hermitian, that is

$$\phi_{a,b,c,d}(\lambda,\mu,\nu) = \bar{\phi}_{a,b,c,d}(-\lambda,-\mu,-\nu).$$

Remark: The non-Gaussianity of our problem induces a very complicated form of the asymptotic covariance matrix. The frequency domain approach for the PMLE delivers as a by-product the typical structure of the asymptotic covariance matrix, with a Gaussian part and a non-Gaussian part. The latter involves a double integral in the trispectrum of the observable process which, for the two PMLEs here considered, appear very difficult to make a comparison with, in order to obtain some relative efficiency result.

Let us consider the asymptotically equivalent 'time domain' expressions.

3.4.2 Asymptotic distribution of the PMLEs with 'time domain' Gaussian likelihood and efficiency comparison

In the coming two sections we will consider 'time domain' expressions for the Gaussian likelihood. This will involve a slightly different parameterization which nevertheless has no consequences in term of the efficiency comparison. For simplicity let us set $\mu = 0$ and $\psi = (\rho, \theta')'$.

[3] Which will interchange the position of the second and third row of the matrix which follows it.

The (univariate) PMLE case

By Assumption B'

$$E_{t-1}y_t = \tau h_{t-1}^2, \ \ var_{t-1}y_t = 2\tau^2 h_{t-1}^4, \ \ \tau = \sigma^2,$$

where $E_{t-1}(.)$ and $var_{t-1}(.)$ denote the expectation and variance respectively, conditional on $\mathcal{F}_{t-1}$. Replacing the α_i with the $\alpha_i(\theta)$ in (3.11), thus obtaining

$$\bar{h}_{t-1}(\psi) = \rho + \sum_{i=1}^{\infty} \alpha_i(\theta)\bar{\epsilon}_{t-i} = \rho + \sum_{i=1}^{t-1} \alpha_i(\theta) \frac{x_{t-i}}{\bar{h}_{t-i-1}(\psi)}, \ t = 2, \ldots, T, \ \ \bar{h}_0(\psi) = \rho,$$

the 'time domain' Gaussian pseudo likelihood for our model is

$$exp\, L(\tau;\psi) = \frac{1}{(4\pi\tau^2)^{T/2}} \prod_{t=1}^{T} (\bar{h}_{t-1}^2(\psi))^{-1} exp\left(-\frac{1}{4\tau^2}\sum_{t=1}^{T} \frac{(y_t - \tau\bar{h}_{t-1}^2(\psi))}{\bar{h}_{t-1}^4(\psi)}\right).$$

Taking the logarithm, differentiating with respect to τ from

$$\frac{\partial L(\hat{\tau}(\psi), \psi)}{\partial \tau} = 0,$$

one obtains the second order equation in $\hat{\tau}(\psi)$

$$2\hat{\tau}^2(\psi) + a_2(\psi)\hat{\tau}(\psi) - a_4(\psi) = 0\,, \tag{3.41}$$

with

$$a_i(\psi) = \frac{1}{T}\sum_{t=1}^{T} \frac{x_t^i}{\bar{h}_{t-1}^i(\psi)}.$$

Considering only the positive solution we get

$$\hat{\tau}(\psi) = \frac{1}{4}[-a_2(\psi) + ((a_2(\psi))^2 + 8a_4(\psi))^{1/2}].$$

For later use we will consider the FOC with respect to $\hat{\tau}(\psi)$ which can be written as

$$T = \frac{Ta_2(\psi)}{2\hat{\tau}(\psi)} - \frac{Ta_4(\psi)}{2\hat{\tau}^2(\psi)}. \tag{3.42}$$

Then from the concentrated (pseudo) loglikelihood $L(\psi) = L(\hat{\tau}(\psi), \psi)$ we get:

$$\begin{aligned}
\frac{\partial L(\psi)}{\partial \psi} &= \frac{-T}{\hat{\tau}(\psi)}\frac{\partial \hat{\tau}(\psi)}{\partial \psi} - \sum_{t=1}^{T} \frac{\partial log \bar{h}_{t-1}^2(\psi)}{\partial \psi} \\
&+ \frac{Ta_4(\psi)}{2\hat{\tau}^3(\psi)}\frac{\partial \hat{\tau}(\psi)}{\partial \psi} + \frac{1}{2\,\hat{\tau}^2(\psi)}\sum_{t=1}^{T} \frac{y_t^2}{\bar{h}_{t-1}^2(\psi)} \frac{\partial log \bar{h}_{t-1}^2(\psi)}{\partial \psi} \\
&- \frac{Ta_2(\psi)}{2\hat{\tau}^2(\psi)}\frac{\partial \hat{\tau}(\psi)}{\partial \psi} - \frac{1}{2\,\hat{\tau}(\psi)}\sum_{t=1}^{T} \frac{y_t}{\bar{h}_{t-1}^2(\psi)} \frac{\partial log \bar{h}_{t-1}^2(\psi)}{\partial \psi}.
\end{aligned}$$

Then by using (3.42) we get:

$$\frac{\partial L(\psi)}{\partial \psi} = -\sum_{t=1}^{T} \frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial \psi}$$
$$+\frac{1}{2}\sum_{t=1}^{T}\left(\frac{y_t^2}{\hat{\tau}^2(\psi)\bar{h}_{t-1}^4(\psi)} - \frac{y_t}{\hat{\tau}(\psi)\bar{h}_{t-1}^2(\psi)}\right)\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial \psi}. \tag{3.43}$$

Also, we have:

$$B_T(\psi) = \frac{1}{T}\frac{\partial^2 L(\psi)}{\partial\psi\partial\psi'} = \frac{1}{2T}\sum_{t=1}^{T}\left(\frac{y_t^2}{\hat{\tau}^2(\psi)\bar{h}_{t-1}^4(\psi)} - \frac{y_t}{\hat{\tau}(\psi)\bar{h}_{t-1}^2(\psi)} - 2\right)\frac{\partial^2 log\bar{h}_{t-1}^2(\psi)}{\partial\psi\partial\psi'}$$
$$+\frac{1}{4T}\sum_{t=1}^{T}\left(\frac{-4y_t^2}{\hat{\tau}^2(\psi)\bar{h}_{t-1}^4(\psi)} + \frac{2y_t}{\hat{\tau}(\psi)\bar{h}_{t-1}^2(\psi)}\right)\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi'}$$
$$+\frac{1}{2T}\sum_{t=1}^{T}\left(\frac{-2y_t^2}{\hat{\tau}^2(\psi)\bar{h}_{t-1}^4(\psi)} + \frac{y_t}{\hat{\tau}(\psi)\bar{h}_{t-1}^2(\psi)}\right)\frac{\partial\hat{\tau}(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi'}.$$

Although not formally proved we conjecture that under Assumptions A', B' the following hold

$$\hat{\tau}(\psi_0) = \tau + o_p(1), \text{ as } T \to \infty, \ \bar{h}_{t-1}(\psi_0) = h_{t-1} + o_p(1), \text{ as } t \to \infty,$$

so that taking conditional expectation and the limit in probability we get that

$$plim_{T\to\infty}B_T(\psi_0) = B(\psi_0)$$
$$= -\frac{5}{2}\left[E(\frac{\partial logh_{t-1}^2(\psi_0)}{\partial\psi}\frac{\partial logh_{t-1}^2(\psi_0)}{\partial\psi'}) - E(\frac{\partial logh_{t-1}^2(\psi_0)}{\partial\psi})E(\frac{\partial logh_{t-1}^2(\psi_0)}{\partial\psi'})\right],$$

using

$$plim\frac{\partial\hat{\tau}(\psi_0)}{\partial\psi} = -\tau E(\frac{\partial logh_{t-1}^2(\psi_0)}{\partial\psi}).$$

Subsequently we get a consistent estimate of $B(\psi_0)$ by replacing expectations with sums, the h_t with the $\bar{h}_t$ and plugging in $\hat{\psi}$. Regarding the covariance matrix of the score, by replacing $\hat{\tau}(\psi_0)$ with τ, given that the asymptotic distribution would be equal, we have that

$$E(\frac{\partial L(\psi)}{\partial\psi}|_{\hat{\tau}(\psi)=\tau,\psi_0}) = 0$$

so that an estimate of

$$lim_{T\to\infty}\frac{1}{T}var\left(\frac{\partial L(\psi_0)}{\partial\psi}\right),$$

is obtained by plugging the PMLE $\hat{\psi}_T$ into

$$V_T(\psi) = \frac{1}{4T}\sum_{t=1}^{T}\left(\frac{y_t^2}{\hat{\tau}^2(\psi)\bar{h}_{t-1}^4(\psi)} - \frac{y_t}{\hat{\tau}(\psi)\bar{h}_{t-1}(\psi)} - 2\right)^2\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi'},$$

in turn asymptotically equivalent to

$$A_T(\psi) = \frac{74}{4}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi'}.$$

evaluated at $\hat{\psi}_T$. Combining the above results yields the 'time domain' asymptotic covariance matrix, ACM_1, for $T^{1/2}(\hat{\psi}_T - \psi_0)$ a consistent estimator of which is given by

$$ACM_1(\hat{\psi}_T) = B_T^{-1}(\hat{\psi}_T)A_T(\hat{\psi}_T)B_T^{-1}(\hat{\psi}_T). \tag{3.44}$$

The bivariate PMLE case

Under Assumption B'

$$E_{t-1}(z_t) = (0, \tau h_{t-1}^2)' = \mu_t, \tag{3.45}$$

$$E_{t-1}(z_t - \mu_t)(z_t - \mu_t)' = \begin{pmatrix} \tau h_{t-1}^2 & 0 \\ 0 & 2\tau^2 h_{t-1}^4 \end{pmatrix}. \tag{3.46}$$

Then, replacing the h_t with the $\bar{h}_t(\psi)$ the Gaussian pseudo loglikelihood for the bivariate process z_t is

$$L_2(\tau,\psi) = -Tlog(2^{3/2}\pi) - \frac{3T}{2}log\,\tau - \frac{3}{2}\sum_{t=1}^{T}log\bar{h}_{t-1}^2(\psi) - \frac{1}{4\tau^2}\sum_{t=1}^{T}\frac{y_t^2}{\bar{h}_{t-1}^4(\psi)} - \frac{T}{4}. \tag{3.47}$$

Concentrating out τ we get

$$\hat{\tau}^2(\psi) = \frac{1}{3T}\sum_{t=1}^{T}\frac{y_t^2}{\bar{h}_{t-1}^4(\psi)}. \tag{3.48}$$

so that the concentrated objective function becomes

$$L_2(\psi) = -Tlog(2^{3/2}\pi) - T - \frac{3T}{4}log\hat{\tau}^2(\psi) - \frac{3}{2}\sum_{t=1}^{T}log\bar{h}_{t-1}^2(\psi), \tag{3.49}$$

so that

$$\frac{\partial L_2(\psi)}{\partial\psi} = \frac{-3}{4\hat{\tau}^2(\psi)}\frac{\partial\hat{\tau}^2(\psi)}{\partial\psi} - \frac{3}{2T}\sum_{t=1}^{T}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}, \tag{3.50}$$

with

$$\frac{\partial\hat{\tau}^2(\psi)}{\partial\psi} = \frac{-2}{3T}\sum_{t=1}^{T}\frac{y_t^2}{\bar{h}_{t-1}^4(\psi)}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}. \tag{3.51}$$

Then it easily follows that a consistent estimator of the asymptotic variance of

$$T^{-1/2}\frac{\partial L_2(\psi_0)}{\partial\psi}$$

is obtained plugging $\hat{\psi}_{2,T}$ into

$$A_{2,T}(\psi) = \frac{87}{4T}\sum_{t=1}^{T} \frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}. \tag{3.52}$$

For the hessian we get

$$\frac{\partial^2 L_2(\psi)}{\partial\psi\partial\psi'} = \frac{-3}{4\hat{\tau}^2(\psi)}\frac{\partial^2\hat{\tau}^2(\psi)}{\partial\psi\partial\psi'} + \frac{3}{4}\frac{\partial log\hat{\tau}^2(\psi)}{\partial\psi}\frac{\partial log\hat{\tau}^2(\psi)}{\partial\psi'} - \frac{3}{2T}\sum_{t=1}^{T}\frac{\partial^2 log\bar{h}_{t-1}^2(\psi)}{\partial\psi\partial\psi'},$$

with

$$\frac{\partial^2\hat{\tau}^2(\psi)}{\partial\psi\partial\psi'} = \frac{-4}{3T}\sum_{t=1}^{T}\frac{y_t^2}{\bar{h}_{t-1}^4(\psi)}\left[\frac{1}{2}\frac{\partial^2 log\bar{h}_{t-1}^2(\psi)}{\partial\psi\partial\psi'} - \frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi'}\right].$$

Thus by the same arguments as in the previous section

$$plim_{T\to\infty}\frac{\partial^2 L_2(\psi_0)}{\partial\psi\partial\psi'}$$
$$= -3\left[E(\frac{\partial log h_{t-1}^2(\psi_0)}{\partial\psi}\frac{\partial log h_{t-1}^2(\psi_0)}{\partial\psi'}) - E(\frac{\partial log h_{t-1}^2(\psi_0)}{\partial\psi})E(\frac{\partial log h_{t-1}^2(\psi_0)}{\partial\psi'})\right].$$

A consistent estimate of the limit of the hessian is given by

$$B_{2,T}(\psi) = \frac{-3}{T}\left[\sum_{t=1}^{T}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi'} - \sum_{t=1}^{T}\frac{\partial log\bar{h}_{t-1}^2(\psi)}{\partial\psi}\sum_{s=1}^{T}\frac{\partial log\bar{h}_{s-1}^2(\psi)}{\partial\psi'}\right], \tag{3.53}$$

evaluated at $\hat{\psi}_{2,T}$.

Combining the above results we finally get a consistent estimator of the 'time domain' covariance matrix, ACM_2, of $T^{1/2}(\hat{\psi}_{2,T} - \psi_0)$ given by

$$ACM_2(\hat{\psi}_{2,T}) = B_{2,T}(\hat{\psi}_{2,T})^{-1}A_{2,T}(\hat{\psi}_{2,T})B_{2,T}(\hat{\psi}_{2,T})^{-1}. \tag{3.54}$$

Efficiency comparison

The result follows by confronting (3.44) with (3.54) which yields

$$ACM_2 = \delta ACM_1$$

where

$$\delta = (87 \times 25)/(36 \times 74) < 1.$$

We conclude that there is an efficiency gain by considering the information gained from the x_t and the y_t rather than from the y_t only.

We conjecture that we could generalize this result hence getting a further efficiency improvement by considering the multivariate process $z_{p,t} = (x_t, x_t^2, \ldots, x_t^p)'$ for some integer $p > 2$ and obtaining the p-variate PMLE $\hat{\psi}_{p,T}$ as

$$\bar{\psi}_{p,T} \quad = argmin_{\psi \in \Psi} \bar{Q}_{p,T}(\psi). \tag{3.55}$$

$$\hat{\psi}_{p,T} \quad = argmin_{\psi \in \Psi} \hat{Q}_{p,T}(\psi), \tag{3.56}$$

where $\bar{Q}_{p,T}(\psi)$, $\hat{Q}_{p,T}(\psi)$ denote the Gaussian pseudo likelihoods for the $z_{p,t}$ in the integral and discretized version respectively. Thus denoting by ACM_p the asymptotic covariance matrix of the p-variate PMLE we should obtain

$$ACM_p \leq ACM_{p-1} \leq \ldots \leq ACM_1,$$

where any of the inequalities above means that the difference between the right hand side and the left hand side gives a semi-positive definite matrix, in particular equal to an identity matrix multiplied by a positive scalar constant. A very challenging problem to study would also be the conditions under which as $p \to \infty$ one would get the $\hat{\psi}_{p,T}$ to be fully efficient or, in other words, to coincide asymptotically with the exact MLE.

Chapter 4

The ‘one-shock’ model: asymptotic theory of the PMLE

In this chapter we will establish the asymptotic properties of the (univariate) Pseudo Maximum Likelihood Estimator proposed in Chapter 3. We will show that the PMLE is strongly $T^{1/2}$ consistent and asymptotic normal so that standard inference can be performed.

4.1 Strong Law of Large Numbers of the PMLE

We will adapt the Hannan (1973) methodology. The conditions required can be summarized as follows:

Ergodicity.

Pure non-deterministicity.

$(f_y(\lambda;\psi)+a)^{-1}$ to be continuous for any $-\pi \le \lambda < \pi$,$\psi \in \Psi$ and a constant $a > 0$.

Identification, viz. $f_y(\lambda;\psi) \neq f_y(\lambda;\bar{\psi})$ iff $\psi \neq \bar{\psi}$.

We will show that these conditions do indeed hold under the assumptions made through the next lemmas.

We cannot apply the Hannan (1973) result directly as Fox and Taqqu (1986) and Dahlhaus (1989) due to the non standard parameterization of our problem. In fact it does not seem that we can factorize $f_y(\lambda;\psi)$ (Hannan 1970, Theorem 10 p.6 3) as

$$f_y(\lambda;\psi) = \frac{m^2(\psi_1)}{2\pi} \mid \sum_{k=0}^{\infty} n_k(\psi_2) e^{ik\lambda} \mid^2, \tag{4.1}$$

with $\psi_1 \cup \psi_2 = \sigma^2 \cup \psi$ and $\psi_1 \cap \psi_2 = \emptyset$ and where the first coefficient of the $n_j(\psi_2)$, $j = 0, 1, \ldots$ is such that $n_0(\psi_2) = 1$.

Robinson (1978a) considered various examples where one might not be able to obtain such a factorization of the spectral density, differently from the linearly regular case, and provided an asymptotic theory for the Gaussian PMLE for such cases where just the knowledge of the functional form of the power spectrum is available. Even if of great generality this approach cannot be used in our framework due to the strong smoothness assumption made on the power spectrum which rules out the long memory case.

Henceforth for any function $h(\lambda;\phi)$, $-\pi \le \lambda < \pi$, ϕ a $q \times 1$ vector , $\phi \in \Phi \subset R^q$, integrable over $[-\pi,\pi]$, $\gamma_h(u;\phi)$, $u = 0, \pm 1, \ldots$ denote the sequence of its Fourier coefficients. $I(\mathcal{A})$ denotes the indicator function which takes value one when the event $\mathcal{A}$ is true and zero otherwise.

Lemma 1 *Under Assumption B the process y_t is ergodic and strictly stationary.*

Proof: Let us write

$$x_t = x(\epsilon_{t-j}; j \ge 0)\,.$$

Then defining the sequence of measurable functions $x^{(k)}$, $k = 1, 2, \ldots$ by

$$x^{(k)} = \mu + \epsilon_t(\rho + \sum_{i=1}^{k} \alpha_i \epsilon_{t-i})\,,$$

we get that

$$x_t = \lim_{k \to \infty} x^{(k)}\,.$$

Thus we have a process which is a measurable function of an i.i.d. process and so is ergodic and stationary by Royden (1968, Chapter 3 Theorem 20) and Stout (1974, Theorem 3.5.8).

$\square$

Lemma 2 *Under Assumptions A'_1, B' the process is purely non-deterministic.*

Proof: From Hannan (1970) a sufficient condition for pure non-deterministicity is given by

$$\int_{-\pi}^{\pi} log\, f_y(\omega) d\omega \quad < \infty .$$

But from

$$\mid log\, f_y(\omega) \mid \leq max \left(\mid f_y(\omega) - 1 \mid, \mid 1 - \frac{1}{f_y(\omega} \mid \right) , \text{ for any } \omega,$$

the result follows from integrability and the continuity of the inverse of the spectrum by Lemma 3.

$\square$

Lemma 3 *Under Assumption B' we obtain*

(i) For any $\psi \in \Psi$ and $0 \leq \lambda \leq \pi$, $f_y(\lambda; \psi) > 0$.

(ii) When also Assumption A'_1 holds, for any $\psi \in \Psi$ such that $0 < d(\theta) < 1/2$, we can find two positive constants δ and $C(\delta, \bar{\rho}^2)$ (where $C(\delta, \bar{\rho}^2)$ is not a function of neither θ and $\bar{\mu}$ but linear in $\bar{\rho}^2$) such that as $\lambda \to 0^+$

$$f_y(\lambda; \psi) \geq C(\delta, \bar{\rho}^2) \mid \lambda \mid^{1-4d(\theta)\ -\delta} .$$

Proof:

(i) For $\lambda \to 0+$ the result comes directly from Theorem 10.

We will consider now the case $0 < \lambda \leq \pi$. The result follows from the fact that we can decompose $f_y(\lambda; \psi)$ as the sum of a non-negative function and of a strictly positive quantity constant in λ. In turn this result is simply due to the nonlinear structure of the process which entail a sort of discontinuity (a positive jump) a the value $u = 0$ of the autocovariance function.

From Corollary 5 substituting the α_i by the $\alpha_i(\theta)$, for any $l \neq 0$

$$\gamma_{f_y}(l;\psi) = \sigma^8 \left(2\alpha^2_{|l|}(\theta)(\delta_{\alpha\alpha}(0;\theta) + \bar{\rho}^2) + 2\delta^2_{\alpha\alpha}(l;\theta) + 4\bar{\rho}^2\delta_{\alpha\alpha}(l;\theta) + 4\bar{\mu}\alpha_{|l|}(\theta)(\bar{\rho}^2 + \delta_{\alpha\alpha}(l;\theta))\right).$$

On the other hand, from Theorem 3,

$$var(y_t) = \sigma^8 \left(2\bar{\rho}^4 + 8\bar{\rho}^2\delta_{\alpha\alpha}(0) + 6\delta^2_{\alpha\alpha}(0) + 4\bar{\mu}^2(\bar{\rho}^2 + \delta_{\alpha\alpha}(0))\right) + \gamma_y(0),$$

and after simple manipulation

$$var(y_t) = \bar{\gamma}(0;\psi_0) + 2\sigma^8(\bar{\rho}^2 + \delta_{\alpha\alpha}(0))^2,$$

setting

$$\bar{\gamma}(0;\psi) = \sigma^8 \left(4(\bar{\rho}^2 + \delta_{\alpha\alpha}(0;\theta))(\bar{\mu}^2 + \delta_{\alpha\alpha}(0;\theta))\right) + \sigma^8 \left(4\bar{\rho}^2\delta_{\alpha\alpha}(0;\theta) + 2\delta^2_{\alpha\alpha}(0;\theta)\right). \tag{4.2}$$

Now if we define another sequence $\bar{\gamma}(l;\psi)$ such that $\bar{\gamma}(l;\psi) = \gamma_{f_y}(l;\psi), l \neq 0$ with $\bar{\gamma}(0;\psi)$ defined as in (4.2), we just need to show that $\bar{\gamma}(l;\psi)$, $l = 0, \pm 1, \dots$ can be thought of as an autocovariance function by its positive (semi) definiteness and so its Fourier transform is non-negative (Rozanov 1963, Theorem 5.1).

In fact, the only thing we need to show is that for any $l = \pm 1, \pm 2, \dots$

$$|\ \bar{\gamma}(l;\psi)\ | \leq \bar{\gamma}(0;\psi).$$

Hence we have

$$\begin{aligned}
&|\ \bar{\gamma}(l;\psi)\ | \leq (i) + |\ (ii)\ |, \\
&\text{with} \\
&(i) = \sigma^8 \left(2\delta^2_{\alpha\alpha}(l;\theta) + 4\bar{\rho}^2\delta_{\alpha\alpha}(l;\theta)\right), \\
&(ii) = \sigma^8 \left(2\alpha^2_l(\theta)(\bar{\rho}^2 + \delta_{\alpha\alpha}(0;\theta)) + 4\bar{\mu}\alpha_l(\theta)(\bar{\rho}^2 + \delta_{\alpha\alpha}(l;\theta))\right).
\end{aligned}$$

Thus we need to show that

$$|\ (ii)\ | \leq (iii),$$

with

$$(iii) = \sigma^8 \left(4(\bar{\rho}^2 + \delta_{\alpha\alpha}(0;\theta))(\bar{\mu}^2 + \delta_{\alpha\alpha}(0;\theta))\right),$$

because the second term on the right hand side of (4.2) bounds (i).

Considering the right part of $-(iii) \leq (ii) \leq (iii)$, it can be written as a second order inequality in $\bar{\mu}$

$$\bar{\mu}^2 - \alpha_l(\theta)\frac{(\bar{\rho}^2 + \delta_{\alpha\alpha}(l;\theta))^2}{(\bar{\rho}^2 + \delta_{\alpha\alpha}(0;\theta))^2}\bar{\mu} + (\delta_{\alpha\alpha}(0;\theta) - \frac{1}{2}\alpha_l^2(\theta)) > 0,$$

which is violated for some values of $\bar{\mu}$ when

$$\alpha_l^2(\theta)\frac{(\bar{\rho}^2 + \delta_{\alpha\alpha}(l;\theta))^2}{(\bar{\rho}^2 + \delta_{\alpha\alpha}(0;\theta))^2} \geq 4\left(\delta_{\alpha\alpha}(0;\theta) - \frac{1}{2}\alpha_l^2(\theta)\right), \tag{4.3}$$

considering that the second order inequality $x^2 + cx + d \leq 0$ is satisfied for some real values of x if and only if $c^2 \geq 4d$. But

$$\alpha_l^2(\theta) < 4\left(\delta_{\alpha\alpha}(0;\theta) - \frac{1}{2}\alpha_l^2(\theta)\right),$$

for any l violating (4.3). The same condition yields $-(iii) \leq (ii)$.

The Fourier transform of the $\bar{\gamma}(l;\psi)$ differs from $f_y(\lambda;\psi)$ by $2\sigma^8(\bar{\rho}^2 + \delta_{\alpha\alpha}(0;\theta))^2$ which for any $d(\theta) > 0$ is strictly positive including the case $\bar{\rho}^2 = 0$.

(ii) By Lemma 11 we only need to consider what happens around the zero frequency. From Theorem 10 we see that as $\lambda \to 0^+$ the terms involving $\bar{\mu}$ are always dominated. Then for any value of $\bar{\rho}^2$, by the exact asymptotic rate result of Theorem 10 we can take any $\delta > 0$ such that the result follows.

□

The exact rate assumption on the $\alpha_j(\theta)$ allows us to identify the power spectrum within the parametric family described by ψ. We will identify the model by a contradiction argument.

Lemma 4 *Under Assumption A_1', A_2', A_7', B', from (2.27) where $f_y(\lambda;\psi) = \sigma^8 K_y(\lambda;\psi)$, for any $\psi, \psi^* \in \Psi$ such that $\psi \neq \psi^*$ we obtain*

$$K_y(\lambda;\psi) \neq K_y(\lambda;\psi^*), \tag{4.4}$$

for any $-\pi \leq \lambda < \pi$.

Proof: Let us choose two vectors of parameters ψ, ψ^* such that for some $(p+2) \times 1$ vector of constants $\eta = (\eta_1, \eta_2, \eta_3')' \neq (0, 0, 0')'$

$$\bar{\mu} = \bar{\mu}^* + \eta_1,$$

$$\bar{\rho}^2 = \bar{\rho}^{2*} + \eta_2\,,$$
$$\theta = \theta^* + \eta_3\,,$$

and for simplicity let us define any quantity defined at ψ by skipping the argument and at ψ^* by writing a star (*) so that[1] $\alpha_i(\theta) = \alpha_i$ and $\alpha_i(\theta^*) = \alpha_i^*$.

Let us start, contrary to (4.4), assuming that for any $\lambda \in [-\pi, +\pi)$,

$$K_y(\lambda;\psi) = K_y(\lambda;\psi^*)\,. \tag{4.5}$$

We will see that the condition $\eta \neq 0$ will lead to contradict the hypothesis of the Lemma.

By using the fact that the autocovariance function completely identifies the power spectrum, condition (4.5) implies for any $u = \pm 1, \pm 2, \ldots$,

$$\begin{aligned}
&2\left(\delta_{\alpha\alpha}^2(u) - \delta_{\alpha\alpha}^{*2}(u)\right) + 4\left(\bar{\mu}\delta_{\alpha\alpha}(u) - \bar{\mu}^*\delta_{\alpha\alpha}^*(u)\right) \\
&+2\left((\bar{\mu} + \delta_{\alpha\alpha}^2(0))\alpha_{|u|}^2 - (\bar{\mu}^* + \delta_{\alpha\alpha}^{*2}(0))\alpha_{|u|}^{*2}\right) + 4\left(\bar{\rho}^2\bar{\mu}\alpha_{|u|} - \bar{\rho}^{*2}\bar{\mu}^*\alpha_{|u|}^*\right) \\
&+4\left(\bar{\rho}^2\alpha_{|u|}\delta_{\alpha\alpha}(u) - \bar{\rho}^{*2}\alpha_{|u|}^*\delta_{\alpha\alpha}^*(u)\right) = 0\,.
\end{aligned}$$

Now we will use the mean value theorem and thus we shall need to evaluate each of the above expressions at a value $\tilde{\theta}$ such that for some $p \times p$ matrix R one can write $\tilde{\theta} = \theta + R(\theta^* - \theta) = \theta + R\eta_3$ such that $\|\tilde{\theta} - \theta\| \leq \|\theta^* - \theta\|$. We shall denote any quantity evaluated at a given $\tilde{\theta}$ by writing a tilde ($\tilde{}$) over it. Although we need to evaluate different functions of θ (and thus different $\tilde{\theta}$ will be used) for simplicity we will not stress this difference from case to case, viz. we shall use the same matrix R. Then the last expression can be written as

$$\begin{aligned}
&4\tilde{\delta}_{\alpha\alpha}(u)\frac{\partial\tilde{\delta}_{\alpha\alpha}(u)}{\partial\theta'}\eta_3 + 4(\bar{\mu} - \bar{\mu}^*)\delta_{\alpha\alpha}^*(u) + 4\bar{\mu}\frac{\partial\tilde{\delta}_{\alpha\alpha}(u)}{\partial\theta'}\eta_3 \\
&+2(\bar{\mu} - \bar{\mu}^*)\alpha_{|u|}^2 + 2\bar{\mu}^*\frac{\partial\tilde{\alpha}_{|u|}^2}{\partial\theta'}\eta_3 + 2\left(\delta_{\alpha\alpha}^2(0) - \delta_{\alpha\alpha}^*(0)\right)\alpha_{|u|}^{2*} + 2\delta_{\alpha\alpha}^*(0)\frac{\partial\tilde{\alpha}_{|u|}^2}{\partial\theta'}\eta_3 \\
&+4(\bar{\rho}^2 - \bar{\rho}^{2*})\bar{\mu}\alpha_{|u|} + 4\bar{\rho}^{*2}\left((\bar{\mu} - \bar{\mu}^*)\alpha_{|u|} + \bar{\mu}^*\frac{\partial\tilde{\alpha}_{|u|}}{\partial\theta'}\eta_3\right) \\
&+4(\bar{\rho}^2 - \bar{\rho}^{*2})\alpha_{|u|}\delta_{\alpha\alpha}(u) + 4\bar{\rho}^{*2}\left(\frac{\partial\tilde{\alpha}_{|u|}}{\partial\theta'}\eta_3\delta_{\alpha\alpha}(u) + \alpha_{|u|}^*\frac{\partial\delta_{\alpha\alpha}(u)}{\partial\theta'}\eta_3\right) = 0\,.
\end{aligned}$$

[1] This is just a notational convenience not to be confused with the coefficients evaluated at the true parameter value ψ_0 as in Chapter 2.

As $u \to \infty$, considering the asymptotic behaviour of the $\alpha_i's$, it follows that the only way equation (4.5) can be satisfied is when $\eta = 0$, which is contrary to our assumption yielding $\psi = \psi^*$.

□

We can then establish the LLN for the PMLE of σ^8.

Lemma 5 *When $\psi \in \Psi$, under Assumptions A_1', A_2', B'*

$$\lim_{T\to\infty} \bar{\sigma^8}_T(\psi) = \lim_{T\to\infty} \hat{\sigma^8}_T(\psi) = \frac{\sigma^8}{2\pi}\int_{-\pi}^{\pi} \frac{K_y(\lambda)}{K_y(\lambda;\psi)} d\lambda \ a.s. \tag{4.6}$$

and the convergence is uniform in ψ.

Proof: Following the idea of Hannan (1973, Lemma 1) the result follows by Lemma 11 and Lemma 3 and by the results of Lemmas 1-2-4. In our model the power spectrum is always strictly positive for any $\psi \in \Psi$ so we will obtain a uniform convergence without restricting the parameter space as in Hannan (1973). Setting by $P_M(\lambda;\psi)$ the Cesaro sum of the Fourier series of $K_y^{-1}(\lambda;\psi)$ taken to M terms and with $\gamma_{K_y^{-1}}(u;\psi)$, $u = 0, \pm1, \pm2, \ldots$ the corresponding Fourier coefficients, we choose M large enough so that $| P_M(\lambda;\psi) - K_y^{-1}(\lambda;\psi) | < \epsilon$ for some $\epsilon > 0$ and uniformly on any Ψ_δ by the uniform convergence of the Cesaro sum in $(\lambda, \psi) \in [-\pi, \pi) \times \Psi$. Then setting

$$c_y(u) = (1/T)\sum_{i=1}^{T-u}(y_i - \bar{y})(y_{i+u} - \bar{y}) \ , u = 0, \pm1, \ldots ,$$

we obtain

$$| \hat{\sigma}_T^8(\psi) - (1/T)\sum_j I(\lambda_j)\, P_M(\lambda_j;\psi) | \le \epsilon (1/T)\sum_j I(\lambda_j) = \epsilon c_y(0) .$$

By the ergodic theorem (Hannan 1970) for any fixed $u = \pm1, \pm2, \ldots$

$$c_y(u) \to_{a.s.} \gamma_y(u) , \ \ c_y(0) \to_{a.s.} var(y_t),$$

which allows us to consider the quantity

$$(1/T)\sum_j I(\lambda_j) P_M(\lambda_j;\psi) = (1/2\pi)\sum_{u=-M}^{M} \gamma_{K_y^{-1}}(u;\psi)(1 - \frac{|u|}{M})[c_y(u) + c(u \pm T)] , \tag{4.7}$$

where

$$c(u \pm T) = \begin{cases} 0 & , u = 0 \,, \\ c_y(T-u) & , u > 0 \,, \\ c_y(T+u) & , u < 0 \,, \end{cases}$$

due to the fact that $\sum_{j>-T/2}^{[T/2]} e^{i\omega\lambda_j}$ is non zero for the case $\lambda = 0$ and for $\lambda = T$ which justifies $c_y(u)$ and $c(T \pm u)$ respectively.

Each $c(u \pm T)$ is made at the most of M terms of the form $\frac{1}{T}(y_i - \bar{y})(y_{T-u+i} - \bar{y})\,, u > 0\,, i > 0$. For each $c(u \pm T)$ being stationary and with $E \mid c(u \pm T) \mid < \infty$ we get that as $T \to \infty$

$$c(u \pm T) \to_{a.s.} 0\,.$$

This yields the convergence of (4.7) for fixed (but large) M to

$$(1/2\pi) \sum_{u=-M}^{M} \gamma_{K_y^{-1}}(u;\psi)(1 - \frac{\mid u \mid}{M})\,((1 - \delta(u,0))\gamma_y(u) + \delta(u,0)var(y_t))$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{-M}^{M} \gamma_{K_y^{-1}}(u;\psi)(1 - \frac{\mid u \mid}{M})e^{iu\lambda} f_y(\lambda) d\lambda = \frac{\sigma^8}{2\pi} \int_{-\pi}^{\pi} P_M(\lambda;\psi) K_y(\lambda) d\lambda\,.$$

The last expression differs by at most $\epsilon\, var(y_t)$ from the right hand term in (4.6).

$\square$

Lemma 6 *When $\psi \in \Psi$, under Assumptions A_1', A_2', B',*

$$\lim_{T\to\infty} \bar{Q}_T(\psi) = \lim_{T\to\infty} \hat{Q}_T(\psi) = Q(\psi) \;\; a.s.,$$

$$Q(\psi) = log \left(\frac{\sigma^8}{2\pi} \int_{-\pi}^{\pi} (\frac{K_y(\lambda)}{K_y(\lambda;\psi)}) d\lambda \right) + \frac{1}{2\pi} \int_{-\pi}^{\pi} log\, K_y(\lambda;\psi) + 1\,,$$

uniformly in Ψ.

Proof: Let us consider just the case of $\widehat{Q}_T(\psi)$. Considering the first element in equation (3.19), the result follows directly from Lemma 5 and the continuity of the logarithmic function. The second element requires more attention but the result follows by using a truncation argument and (Robinson 1994b, Lemma 4).

Let j_ϵ be an integer such that $\frac{1}{T}2\pi j_\epsilon \to \epsilon > 0$ for an arbitrary $0 < \epsilon < 1/2$ as $T \to \infty$. Also let m be an integer such that $\frac{1}{m} + \frac{m}{T} \to 0$ as $T \to \infty$ where we can always make it large enough so that $\frac{2\pi m}{T} < \epsilon$.

Then, remembering that the case $\lambda_0 = 0$ is excluded and by the evenness of the power spectrum one needs just to consider:

$$\frac{\sum_{j=1}^{[T/2]} log\ K_y(\lambda_j;\psi)}{T} =$$
$$\frac{\sum_{j=j_\epsilon+1}^{[T/2]} log\ K_y(\lambda_j;\psi)}{T} + \frac{\sum_{j=m}^{j_\epsilon} log\ K_y(\lambda_j;\psi)}{T} + \frac{\sum_{j=1}^{m-1} log\ K_y\ (\lambda_j;\psi)}{T} =$$
$$= (i)\ + (ii)\ + (iii).$$

Now, by the continuity of the spectrum away from the zero frequency, as $T \to \infty$

$$(i) \to \frac{\int_\epsilon^\pi log\ K_y(\lambda;\psi)d\lambda}{2\pi}.$$

In the second place, using the fact that:

$$1 - \frac{1}{X} \leq log\ X \leq X - 1 \ \text{ for } \ X > 0,$$

one obtains $(iii) = o(1)$ as $T \to \infty$.

In fact

$$(iii)' = \frac{m}{T} - (1/T)\sum_{j=1}^{m-1} \frac{1}{K_y(\lambda_j;\psi)} \leq (iii) \leq -\frac{m}{T} + \frac{\sum_{j=1}^{m-1} K_y\ (\lambda_j;\psi)}{T} = (iii)''.$$

As $T \to \infty$ the left-hand side of the latter expression is $O(\frac{m}{T}) = o(1)$ by the continuity of the inverse of the spectrum, and the right-hand-side is $O((\lambda_m)^{2\delta}) = o(1)$, by (Robinson 1994b, Lemma 4) (setting the slowly varying function equal to a constant) given the behaviour of the spectral density of the process at zero frequency and that by A_1' we can always choose $\delta > 0$ such that $max_{\psi \in \Psi} d(\theta) < 1/2 - \delta < 1/2$.

Concerning (ii), we have to show that

$$\lim_{T\to\infty} \frac{\sum_{j=m}^{j_\epsilon} log\ K_y(\lambda_j;\psi)}{T} = \frac{\int_0^\epsilon log\ K_y(\lambda;\psi)d\lambda}{2\pi}.$$

Now we obtain

$$| (ii) - \frac{1}{2\pi} \int_0^{\epsilon} log\, K_y(\omega; \psi) d\omega \,| \leq | \,(ii)' \,| + | \,(ii)'' \,|,$$

where $(ii)' = \left(\frac{1}{T} \sum_{j=m}^{j_\epsilon} log\, K_y(\lambda_j; \psi) - \frac{1}{2\pi} \int_{\lambda_m}^{\epsilon} log\; K_y(\lambda; \psi) d\lambda\right) = o(1)$, as $T \to \infty$, and secondly that $(ii)'' = \frac{1}{2\pi} \int_0^{\lambda_m} log\, K_y(\lambda; \psi) d\lambda = O(\lambda_m log\,(\lambda_m))$, by a direct solving of the integral.

In fact, we can bound $| \,(ii)' \,|$ as:

$$\begin{aligned} | \,(ii)' \,| &= \left| \frac{log\,(K_y(\lambda_m; \psi))}{T} + \frac{1}{2\pi} \sum_{j=m}^{(j_\epsilon)-1} \int_{\lambda_j}^{\lambda_{j+1}} log\,(\frac{K_y(\lambda_{j+1}; \psi)}{K_y(\lambda; \psi)}) d\lambda \right| \\ &\leq \left| \frac{log\,(K_y(\lambda_m; \psi))}{T} \right| + \left| \frac{1}{2\pi} \sum_{j=m}^{(j_\epsilon)-1} \int_{\lambda_j}^{\lambda_{j+1}} log\,(\frac{K_y(\lambda_{j+1}; \psi)}{K_y(\lambda; \psi)}) d\lambda \right| \\ &= | \,(ii.1)' \,| + | \,(ii.2)'' \,| \,. \end{aligned}$$

Now as $T \to \infty$

$$| \,(ii.1)' \,| = O(\frac{log\, T}{T}) \,.$$

For any a, b with $ab > 0$, from

$$1 - \frac{b}{a} \leq log\,(\frac{a}{b}) \leq \frac{a}{b} - 1 \,, \tag{4.8}$$

one obtains

$$1 - \frac{\lambda}{\lambda_{j+1}} \leq log\,(\frac{\lambda_{j+1}}{\lambda}) \leq \frac{\lambda_{j+1}}{\lambda} - 1 \,,$$

so that by integrating from both sides on the interval $[\lambda_j, \lambda_{j+1}]$ one gets

$$\frac{2\pi}{T} - \frac{\lambda_{j+1}^2 - \lambda_j^2}{2\lambda_{j+1}} \leq \int_{\lambda_j}^{\lambda_{j+1}} log\,(\frac{\lambda_{j+1}}{\lambda}) d\lambda \leq \lambda_{j+1} log\,(\frac{\lambda_{j+1}}{\lambda_j}) - \frac{2\pi}{T} \,. \tag{4.9}$$

After some algebra and from $log\,(X) \leq X - 1\,, X \geq 1$ the above double inequality implies

$$\frac{\pi}{T(j+1)} \leq \int_{\lambda_j}^{\lambda_{j+1}} log\,(\frac{\lambda_{j+1}}{\lambda}) d\lambda \leq \frac{2\pi}{Tj} \,.$$

We can then bound $(ii.2)''$ by

$$| \,(ii.2)'' \,| = O(\frac{log\; j_\epsilon}{T}) \,,$$

where all the bounds do not depend on ψ, thus justifying the uniform convergence.

□

Lemma 7 *Under Assumptions A_1' A_2', A_7' for any $\psi \in \Psi$*

$$Q(\psi) = log\,(\sigma^8) + log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{K_y(\lambda)}{K_y(\lambda;\psi)}d\lambda\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log\, K_y(\lambda;\psi) + 1$$
$$\geq Q(\psi_0) = log\,(\sigma^8) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log\, K_y(\lambda) + 1\,,$$

so that the minimum is attained for $\psi = \psi_0$.

Proof: By Jensen's inequality for integrals, for any positive integrable function $h(.)$

$$log\,(\frac{\int_{-\pi}^{\pi} h(x)dx}{2\pi}) \geq \frac{\int_{-\pi}^{\pi} log\; h(x)dx}{2\pi}\,,$$

where the equality is obtained for $h(x)$ being the constant function, the result follows as

$$log\,(\sigma^8) + log\,(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{K_y(\lambda)}{K_y(\lambda;\psi)}d\lambda) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log\; K_y(\lambda;\psi) + 1 =$$
$$= log\,(\sigma^8) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log\; K_y(\lambda)d\lambda + 1$$
$$+log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{K_y(\lambda)}{K_y(\lambda;\psi)}d\lambda\right) - \frac{1}{2\pi}\int_{-\pi}^{\pi} log\,\frac{K_y(\lambda)}{K_y(\lambda;\psi)}d\lambda \geq Q(\psi_0)\,.$$

□

Lemma 6 and Lemma 7 ensure the asymptotical global identifiability of the model. We can then prove the LLN for the PMLE of ψ.

Theorem 13 *Under Assumptions A_1', A_2', A_7', B',*

$$\lim_{T\to\infty}\bar{\psi}_T = \lim_{T\to\infty}\hat{\psi}_T = \psi_0 \; a.s.,$$
$$\lim_{T\to\infty}\bar{\sigma}_T^8(\bar{\psi}_T) = \lim_{T\to\infty}\hat{\sigma}_T^8(\widehat{\psi}_T) = \sigma^8 \; a.s..$$

Proof: We consider the proof for $\hat{\psi}_T$. Given this, the result follows for $\hat{\sigma}^8_T(\hat{\psi}_T)$ by its uniform convergence property (from Lemma 5 above).

Following (Hannan 1973, Theorem 1), if $\hat{\psi}_T$ does not converge to ψ_0 there is a subsequence converging to $\psi' \in \Psi$, $\psi' \neq \psi_0$. Let this subsequence be $\hat{\psi}_n$ where n is a subsequence of $T = 1, 2, \ldots$. So for any $\eta > 0$, by Lemma 6 and Lemma 7,

$$\begin{aligned}
\liminf_{n\to\infty} \hat{Q}_T(\hat{\psi}_n) &\geq \liminf_{n\to\infty} \left\{ log \left(\frac{1}{T} \sum_j \left(\frac{I(\lambda_j)}{K_y(\lambda_j; \hat{\psi}_n) + \eta} \right) \right) \right. \\
&+ \left. \frac{1}{T} \sum_j \left(log\,[K_y(\lambda_j; \hat{\psi}_n) + \eta] - \eta \right) + 1 \right\} \\
&= log\,(\sigma^8) + log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{K_y(\lambda)}{K_y(\lambda; \psi') + \eta} d\lambda \right) \\
&+ \frac{\int_{-\pi}^{\pi} (log\,[K_y(\lambda; \psi') + \eta] - \eta)\, d\lambda}{2\pi} + 1\,(a.s.) \\
&> Q(\psi_0)\,,
\end{aligned}$$

for a sufficiently small η.

But, for any $\psi \in \Psi$

$$\begin{aligned}
&\limsup_{n\to\infty} \hat{Q}_T(\hat{\psi}_n) \leq \limsup_{n\to\infty} \hat{Q}_T(\psi) = Q(\psi) = \\
&= log\,(\sigma^8) + log \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{K_y(\lambda)}{K_y(\lambda; \psi)} \right) d\lambda \right) \\
&+ \frac{\int_{-\pi}^{\pi} log\, K_y(\lambda; \psi) d\lambda}{2\pi} + 1\,,
\end{aligned}$$

so that

$$\limsup_{n\to\infty} \hat{Q}_T(\hat{\psi}_n) \leq \inf_{\psi\in\Psi} Q(\psi) = Q(\psi_0)\,.$$

So by contradiction, $\psi' = \psi_0$.

□

Remarks : (1) We exploited the ergodicity property for the a.s. convergence of the sample autocovariance. Alternatively we could use the fourth order cumulant expression (Theorem 12) in order to derive the convergence in probability of $c_y(u)$ given that as $T \to \infty$

$$var(c_y(u)) = O(T^{4d-2}).$$

(2) The result in Hannan (1973) is valid even in the case where the power spectrum has one or more zeros. In our case, as it will be clear in the CLT context, the full proof requires a strictly positive spectrum at all frequencies, due to the non standard parameterization of our model (cf. Robinson (1978a)) which implies a different object function. In fact, Fox and Taqqu (1986) who consider a standard parameterization case as well as Gaussianity can use Hannan (1973) directly.

(3) An alternative proof of the type of convergence displayed in Lemma 5 is in Hosoya and Taniguchi (1982, Lemma A.3.3) . Unfortunately they impose stricter conditions on the model ruling out the long memory case assuming that the power spectrum is $Lip(\alpha), \alpha > 1/2$ (Zygmund 1977) at all frequencies.

4.2 The Central Limit Theorem for the PMLE

Let us consider $\bar{Q}_T(\psi)$. We will also show formally that considering $\hat{Q}_T(\psi)$ instead yields asymptotically equivalent inference.

Differentiating with respect to ψ, by Lemma 13 below, yields

$$\frac{\partial}{\partial\psi}\bar{Q}_T(\psi) = \frac{1}{2\pi\bar{\sigma}_T^8(\psi)}\int_{-\pi}^{\pi}\frac{\partial}{\partial\psi}K_y^{-1}(\omega;\psi)I(\omega)d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\psi}log\, K_y(\omega;\psi)d\omega. \tag{4.10}$$

Assuming that the true value ψ_0 lies in the interior of Ψ, given the consistency of the PMLE $\bar{\psi}_T$, by the delta method, we obtain

$$\frac{\partial}{\partial\psi}\bar{Q}_T(\bar{\psi}_T) = 0 = \frac{\partial}{\partial\psi}\bar{Q}_T(\psi_0) + \frac{\partial^2}{\partial\psi\partial\psi'}\bar{Q}_T(\tilde{\psi})(\bar{\psi}_T - \psi_0),$$

where $\tilde{\psi} = \psi_0 + R(\bar{\psi}_T - \psi_0)$, for some $p+2 \times p+2$ matrix R such that $\| \tilde{\psi} - \psi_0 \| \leq \| \bar{\psi}_T - \psi_0 \|$, $\| . \|$ denoting the Euclidean norm.

Hence finding the asymptotic distribution of $\bar{\psi}_T$ entails finding the asymptotic distribution for the score evaluated as ψ_0. In turn, from (4.10) it means that one needs to find the asymptotic distribution of the quadratic form

$$\sum_{t,s=1}^{T} \gamma_g(t-s;\psi_0)(y_t-\bar{y})(y_s-\bar{y}), \tag{4.11}$$

with $\gamma_g(t-s;\psi)$ being the Fourier coefficients of $g(\lambda;\psi) = \frac{\partial}{\partial\psi}K_y^{-1}(\lambda;\psi)$.

In a short memory framework, the weak convergence for such a quadratic form was established when y_t is non-Gaussian for tapered (Dahlhaus 1983) and non-tapered data (Bentkus 1976). These results rely either on absolute summability of autocovariances and all higher order cumulants or on fourth integrability of the k-th order cumulant spectrum for any $k \geq 2$. In a long memory framework, weak convergence results were established when y_t is Gaussian (Fox and Taqqu 1987) or an instantaneous transformation of a Gaussian process (Fox and Taqqu 1985). These results have been generalized to the case where y_t is a linear process in i.i.d. white noise (Giraitis and Surgailis 1990a) or martingale difference sequence (Heyde and Gay 1993).

The fact that in out case y_t is non-Gaussian by its nonnegativity and displays long memory implies that we can not use such limit theorems to derive the asymptotic distribution of (4.11). Indeed, our approach is based on finding the asymptotic rate of the joint cumulants for (4.11) and on showing that for any order greater than 2 they converge to zero as the sample size goes to infinity. A similar idea was used by Robinson (1995b) in a semiparametric framework where again the combination long memory and nonlinearity (of a semiparametric nature) brought a great many technical problems ruling out the possibility of using existing limit laws.

We need to strengthen Assumption B' as follows:

Assumption B" : *The ϵ_t are Gaussian and satisfy Assumption B'.*

As was pointed out, we need to differentiate the objective function twice with respect to the parameters. Before doing that, we will establish a number of properties of $f_y(\lambda;\psi)$.

Lemma 8 *Under Assumptions A_1', A_2', A_3', A_4', B' for any $\psi \in \Psi$, as $\lambda \to 0^+$*

$$\frac{\partial f_y(\lambda;\psi)}{\partial\theta_j} \sim \begin{cases} K \mid\lambda\mid^{1-4d(\theta)} log(\frac{1}{\lambda}), & \bar{\rho}^2 = 0, \quad 1/4 < d(\theta) < 1/2, \\ K \mid\lambda\mid^{-2d(\theta)} log(\frac{1}{\lambda}), & \bar{\rho}^2 \neq 0, \quad 0 < d(\theta) < 1/2, \end{cases} \tag{4.12}$$

for any $j = 1, 2, \ldots, p$.

Proof: Using the same notation as in Theorem 7, by Assumptions A_1 and A_3, Lemma 27 and Lemma 29 we obtain, as $l \to \infty$

$$\begin{aligned} &\text{for } \frac{\partial f_1(\nu;\theta)}{\partial\theta_j}, \ \frac{\partial\delta^2_{\alpha\alpha}(l;\theta)}{\partial\theta_j} \sim K\, l^{4d(\theta)-2}\, log(l)\,, \\ &\text{for } \frac{\partial f_2(\nu;\theta,\bar{\rho}^2)}{\partial\theta_j}, \ \frac{\partial\delta_{\alpha\alpha}(l;\theta)}{\partial\theta_j} \sim K\, l^{2d(\theta)-1}\, log(l)\,, \\ &\text{for } \frac{\partial f_4(\nu;\theta,\mu,\bar{\rho}^2)}{\partial\theta_j}, \ \frac{\partial\alpha^2_{|l|}(\theta)}{\partial\theta_j} \sim K\, l^{2d(\theta)-2}\, log(l)\,, \\ &\text{for } \frac{\partial f_5(\nu;\theta)}{\partial\theta_j}, \ \frac{\partial\alpha_{|l|}(\theta)}{\partial\theta_j} \sim K\, l^{d(\theta)-1}\, log(l)\,, \\ &\text{for } \frac{\partial f_6(\nu;\theta,\mu)}{\partial\theta_j}, \ \frac{\partial\left(\alpha_{|l|}(\theta)\delta_{\alpha\alpha}(l;\theta)\right)}{\partial\theta_j} \sim K\, l^{3d(\theta)-2}\, log(l)\,. \end{aligned}$$

for $j = 1, 2, \ldots, p$.

Then, given also Assumptions A_2 and A_4 and Lemma 30, by using Proposition 2 to each of the $\frac{\partial}{\partial\theta_j} f_i(.;\theta,.)$, $i = 1, 2, 4, 5, 6$, we get as $\nu \to 0^+$

$$\begin{aligned} &\frac{\partial f_1(\nu;\theta)}{\partial\theta_j} \sim K \mid\nu\mid^{1-4d(\theta)} \ log(\frac{1}{\nu}), \text{ for } 1/4 < d(\theta) < 1/2\,, \\ &\frac{\partial f_2(\nu;\theta,\bar{\rho}^2)}{\partial\theta_j} \sim K \mid\nu\mid^{-2d(\theta)} \ log(\frac{1}{\nu}), \text{ for } 0 < d(\theta) < 1/2\,, \\ &\frac{\partial f_5(\nu;\theta,\mu,\bar{\rho}^2)}{\partial\theta_j} \sim K \mid\nu\mid^{-d(\theta)} \ log(\frac{1}{\nu}), \text{ for } 0 < d(\theta) < 1\,, \\ &\frac{\partial f_6(\nu;\theta,\mu)}{\partial\theta_j} \sim K \mid\nu\mid^{1-3d(\theta)} \ log(\frac{1}{\nu}), \text{ for } 1/3 < d(\theta) < 2/3\,. \end{aligned}$$

□.

Remark: In this case with respect to Proposition 2 a slowly varying function, the log function, (Yong 1974) appears.

More in general we can write a general expression for the multiple derivative.

Lemma 9 *Under assumptions* A_1', A_2', $A_5(s)'$, $A_6(s)'$, B' *for* $s = 1, 2, ...b$ *for any* $\psi \in \Psi$ *, as* $\lambda \to 0^+$

$$\frac{\partial^b f_y(\lambda;\psi)}{\partial\theta_{i_1}...\partial\theta_{i_b}} \sim \begin{cases} K \mid \lambda \mid^{1-4d(\theta)} log^b(\frac{1}{\lambda}), & \bar{\rho}^2 = 0, \quad 1/4 < d(\theta) < 1/2, \\ K \mid \lambda \mid^{-2d(\theta)} log^b(\frac{1}{\lambda}), & \bar{\rho}^2 \neq 0, \quad 0 < d(\theta) < 1/2, \end{cases} \tag{4.13}$$

for $i_j \in \{1, 2, \ldots, p\}, j = 1, 2, \ldots, b$.

Proof: Following the line of Lemma 8 the result follows directly.

□

Then with respect to the scalar parameters $\bar{\mu}$, $\bar{\rho}^2$ we get the following:

Lemma 10 *Under Assumptions* A_1, A_2, B', *for any* $\psi \in \Psi$ *, as* $\lambda \to 0^+$

$$\frac{\partial f_y(\lambda;\psi)}{\partial\bar{\mu}} \sim \begin{cases} K \mid \lambda \mid^{1-3d(\theta)}, & \bar{\rho}^2 = 0, \quad 1/3 < d(\theta) < 1/2, \\ K \mid \lambda \mid^{-d(\theta)}, & \bar{\rho}^2 \neq 0, \quad 0 < d(\theta) < 1/2, \end{cases}$$

and

$$\frac{\partial f_y(\lambda;\psi)}{\partial\bar{\rho}^2} \sim K \mid \lambda \mid^{-2d(\theta)}, 0 < d(\theta) < 1/2.$$

Proof: Using the same notation as in Theorem 7, we get

$$\frac{\partial f_y(\lambda;\psi)}{\partial\mu} = f_5(\lambda;\theta,\mu = 1,\bar{\rho}^2) + f_6(\lambda;\theta,\mu = 1),$$
$$\frac{\partial f_y(\lambda;\psi)}{\partial\bar{\rho}^2} = f_2(\lambda;\theta,\mu,\bar{\rho}^2 = 1) + f_5(\lambda;\theta,\mu,\bar{\rho}^2 = 1).$$

Then the result follows by considering the asymptotic behaviour of each of the $f_i(.,.)$, $i = 2, 5, 6$ as obtained in Theorem 10.

□

The following result will be used, which gives the conditions such that a function is approximately (due to the slowly varying function) Lipschitz continuous (Zygmund 1977) away from the zero frequency.

Proposition 3 *(Robinson 1994a, Lemma 8)*
Let u_t be a stochastic process with autocovariance function $\gamma_u(l)$ and power spectrum $f_u(\lambda)$. Let $L(x)$ be a slowly varying function. Then if

as $\lambda \to 0+$, $0 < d < 1/2$,

$$f_u(\lambda) \sim L(\frac{1}{\lambda})\lambda^{-2d},$$

and if $\gamma_u(l)$ is quasi-monotonically convergent to zero (QMC),

$$min(\lambda, \mid \lambda - \mu \mid) \mid f_u(\lambda) - f_u(\lambda - \mu) \mid = O\left(\mid \mu \mid^{1-2d} L(\frac{1}{\mid \mu \mid})\right) \quad as\ \mu \to 0.$$

uniformly in $\lambda \in (0, \pi)$.

Then concerning the behaviour of the $f_y(\lambda; \psi)$ away from the origin, one obtains

Lemma 11 *Under Assumptions A_1', A_2', B' $f_y(\lambda; \psi)$ is continuous for all (ψ, λ) such that $\lambda \neq 0$ and $\frac{I(\bar{\rho}^2=0)}{4} < d(\theta) < \frac{1}{2}$.*

Proof: Under Assumptions A_1, A_2 from Theorem 10 and Lemma 30 in Appendix A, the Assumptions of Proposition 3 hold. For the case $\bar{\rho}^2 = 0$ the conditions of Proposition 3 are still valid where the asymptotic rate of the power spectrum at the origin being equal to $1 - 4d(\theta)$.

□

Lemma 12 *Under Assumptions $A_1', A_2', A_5(s)', A_6(s)', B'$ for $s = 1, 2, \ldots, S$ the function*

$$\frac{\partial^S f_y(\lambda; \psi)}{\partial\psi_{i_1} \cdots \psi_{i_S}}$$

is continuous for all (ψ, λ) *such that* $\lambda \neq 0$.

Proof: Following the same line of Lemma 11 and by using Lemma 8-9-17 and Lemma 30.

□

Lemma 13 *Under Assumptions* $A_1', A_2', A_3', A_4', A_5'(2), A_6'(2), B'$

$$\int_{-\pi}^{\pi} log\, K_y(\omega, \psi) d\omega$$

can be differentiated twice under the integral sign.

Proof: We will follow Fox and Taqqu (1986). Let us skip the second differentiation case given its total similarity with the first differentiation case.

Denoting by e_j the $j-th$ unit vector in R^{p+2}, for some arbitrary $\epsilon > 0$ we write

$$(1/\epsilon)\left(\int_{-\pi}^{\pi} log\, K_y(\omega; \psi + e_j\epsilon) d\omega - \int_{-\pi}^{\pi} log\, K_y(\omega; \psi) d\omega\right) =$$
$$(1/\epsilon)\int_{-\pi}^{\pi} \left(log\, K_y(\omega; \psi + e_j\epsilon) - log\, K_y(\omega; \psi)\right) d\omega\,,$$

Thus by the mean value theorem the integrand can be majorized by

$$\mid \frac{1}{\partial \psi_k} log\, K_y(\omega; \tilde{\psi}) \mid$$

where $\|\tilde{\psi} - \psi\| < \epsilon$.

Then by writing $d_l = min_{\psi \in \Psi} d(\theta)$ and $d_u = max_{\psi \in \Psi} d(\theta)$ we can in turn bound the last integral by

$$K \int_{-\pi}^{\pi} \mid \omega \mid^{(2+2\, I(\bar{\rho}^2=0))(d_l - d_u) - \delta} d\omega < \infty\,,$$

which is bounded because

$$(d_l - d_u) > -1/2 + I(\bar{\rho}^2 = 0)/4$$

and

$$(-1/2 + I(\bar{\rho}^2 = 0)/4)(2 + 2\, I(\bar{\rho}^2 = 0)) = -1$$

where δ is arbitrary . Thus the dominated convergence theorem yields the result.

□

Lemma 14 *Under Assumptions A_1', A_2', A_7', B' the Fourier coefficients of $1/K_y(\omega;\psi)$ are QMC.*

Proof: The continuity and thus the integrability of the inverse of the spectrum, by the Riemann-Lebesgue Theorem yields the convergence to zero of its Fourier coefficients. Then the Lipschitz type of continuity given by Lemma 11 allows us to use Zygmund (1977, Theorem 4.7 (i) pp.46) and thus to bound the Fourier coefficients of $1/K_Y(\lambda;\psi)$ as $u \to \infty$ by

$$\int_{-\pi}^{\pi} 1/K_y(\lambda;\psi)e^{i\lambda u}d\lambda = O(|\, u \,|^{-\beta}),$$

for some $0 < \beta < 1$. These two results by Yong (1974, Lemma I-1 p.4) justify the QMC property.

$\square$

Lemma 15 *Under Assumptions $A_1', A_2', A_3', A_4', A_7', B'$, the Fourier coefficients of $g(\omega;\psi)$ are QMC and as $u \to \infty$*

$$\gamma_g(u;\psi) \sim \begin{cases} Ku^{-4d(\theta)}log\,(u), & \bar{\rho}^2 = 0 \\ Ku^{-1-2d(\theta)}log\,(u), & \bar{\rho}^2 \neq 0, \end{cases}$$

for any $\psi \in \Psi$.

Proof: The function $g(\lambda;\psi)$ is obtained as the product of three functions, viz. 2 times $1/K_y(\lambda;\psi)$ multiplied by $-\frac{\partial}{\partial\psi}K_y(\lambda;\psi)$. Thus each $\gamma_g(u;\psi)$ is equivalent to the (multiple) convolution of three series of QMC coefficients, by Lemma 30 and 14, and so by Lemma 27 it is QMC itself. Then [2] Proposition 2 and using Theorem 10 8 and 17 the second part of the Lemma follows.

$\square$

We now prove the asymptotic equivalence of $\bar{Q}_T(\psi)$ and $\hat{Q}_T(\psi)$.

[2]More rigorously we use one of the extensions of Proposition 2 not reported here for simplicity which allows the Fourier transform to have a zero at the origin as Yong (1974, Theorem III-33).

Lemma 16 *Under the Assumptions $A_1', A_2', A_3', A_4', B'$, for any $\psi \in \Psi$ as $T \to \infty$*

$$\frac{\partial \hat{Q}_T(\psi)}{\partial \psi} - \frac{\partial \bar{Q}_T(\psi)}{\partial \psi} = o_p(T^{-1/2}),$$

Proof: Setting

$$w(\lambda;\psi) = \frac{1}{K_y(\lambda;\psi)} \frac{\partial K_y(\lambda;\psi)}{\partial \psi},$$

we can write

$$\begin{aligned}
&\left| \frac{\partial \hat{Q}_T(\psi)}{\partial \psi} - \frac{\partial \bar{Q}_T(\psi)}{\partial \psi} \right| \leq \\
&\left| \frac{1}{T} \sum_j w(\lambda_j;\psi) - \frac{1}{2\pi} \int_{-\pi}^{\pi} w(\lambda;\psi) d\lambda \right| + \\
&+ \frac{1}{2\pi T} \sum_{a,b=1}^{T} \left| (y_a - \bar{y})(y_b - \bar{y}) \right| \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda;\psi) e^{i(a-b)\lambda} d\lambda - \frac{1}{T} \sum_j g(\lambda_j;\psi) e^{i(a-b)\lambda_j} \right| \\
&= |\, (i) \,| \; + \; |\, (ii) \,| \; .
\end{aligned}$$

Thus following the line of Lemma 6 we shall obtain that as $T \to \infty$

$$T^{1/2}(i) = o(1),$$

where, in order to get the result we need to impose stricter conditions on the integer m, which we can always do, such as $T \to \infty$

$$\frac{1}{m} + \frac{\log T\, m}{T^{1/2}} + \frac{m^{2-4d(\theta)}}{T^{3/2-4d(\theta)}} \to 0,$$

for any $\psi \in \Psi$, and we use the fact that the gradient is continuous off the zero frequency (Lemma 12) together with the asymptotic behaviour at zero frequency (Theorem 8-17).

Concerning (ii) writing by stationarity $E(y_t - \bar{y})^2 = K$ for some finite positive constant K and setting $u = a - b$ we obtain

$$E(|\,(ii)\,|) \leq |\,(ii.1)\,| + |\,(ii.2)\,|,$$

with

$$(ii.1) = \frac{K}{4\pi^2 T} \sum_{a,b=1}^{T} \left(\sum_j \int_{\lambda_{j-1}}^{\lambda_j} \left(g(\omega;\psi) - g(\lambda_j;\psi)\right) d\omega \cos(u\lambda_j) \right),$$

$$(ii.2) = \frac{K}{4\pi^2 T} \sum_{a,b=1}^{T} \left(\sum_j \int_{\lambda_{j-1}}^{\lambda_j} g(\omega;\psi) \left(\cos(u\lambda_j) - \cos(u\omega)\right) d\omega \right).$$

Denoting by $D_n(x)$ the Dirichlet Kernel, viz.

$$D_n(x) = 1/2 + \sum_{j=1}^{n} cos(jx) \,,$$

by Abel transformation we obtain a bound for $(ii.1)$ as

$$| \, (ii.1) \, | \leq \frac{2K}{4\pi^2 T} \sum_{a,b=1}^{T} \left(| \sum_{j=1}^{[T/2]} \left\{ \int_{\lambda_{j-1}}^{\lambda_j} (g(\omega;\psi) - g(\lambda_j;\psi))d\omega - \int_{\lambda_j}^{\lambda_{j+1}} (g(\omega;\psi) - g(\lambda_{j+1};\psi))d\omega \right\} \times \right.$$
$$\left. \times (D_j(2u\pi/T) - 1/2) \, | + | \int_{\lambda_{[T/2]-1}}^{\lambda_{[T/2]}} (g(\omega;\psi) - g(\lambda_{[T/2]};\psi))d\omega (D_{[T/2]}(2u\pi/T) - 1/2) \, | \right) .$$

Then from $| \, D_n(x) - 1/2 \, | \leq \pi/x \,, 0 < x \leq \pi$, we obtain

$$| \, (ii.1) \, | \leq \frac{2K}{4\pi^2 T} \sum_{a,b=1}^{T} \frac{T}{2 \, | \, a-b \, |} \left(| \int_0^{\lambda_1} (g(\omega;\psi) - g(\lambda_1;\psi))d\omega \right.$$
$$\left. - \int_{\lambda_{[T/2]}}^{\lambda_{[T/2]+1}} (g(\omega;\psi) - g(\lambda_{[T/2]+1};\psi))d\omega \, | + | \int_{\lambda_{[T/2]}}^{\lambda_{[T/2]+1}} (g(\omega;\psi) - g(\lambda_{[T/2]+1};\psi))d\omega \, | \right) .$$

Thus, by the mean value theorem, we can write for some $0 \leq \lambda' \leq \lambda_1$ and $\lambda_{[T/2]} \leq \lambda'' \leq \lambda_{[T/2]+1}$,

$$\int_0^{\lambda_1} (g(\omega;\psi) - g(\lambda_1;\psi))d\omega = \frac{2\pi}{T}(g(\lambda';\psi) - g(\lambda_1;\psi)) \,,$$
$$\int_{\lambda_{[T/2]}}^{\lambda_{[T/2]+1}} (g(\omega;\psi) - g(\lambda_{[T/2]+1};\psi))d\omega = \frac{2\pi}{T}(g(\lambda'';\psi) - g(\lambda_{[T/2]+1};\psi)) \,.$$

Thus from the (approximate) Lipschitz condition we obtain for some $\beta > 0$ as $T \to \infty$

$$| \, (ii.1) \, | = O(T^{-\beta} \sum_{u=-T+1}^{T-1} (1 - \frac{|\,u\,|}{T}) \frac{1}{|\,u\,|}) = O(T^{-\beta} log \, T) \,.$$

Considering now $(ii.2)$ and using the result that for any continuous function with integrable derivative $e(x), x \in [-\pi, \pi]$ we obtain (Brillinger 1975, ex.17.14 p.15) as $T \to \infty$

$$\frac{2\pi}{T} \sum_{j=-[T/2]+1}^{[T/2]} e(\lambda_j) = \int_{-\pi}^{\pi} e(\omega)d\omega \; (1 \, + \, O(1/T)) \,,$$

we can write as $T \to \infty$

$$| \, (ii.2) \, | = O\left(\sum_{u=-T+1}^{T-1} (1 - \frac{|\,u\,|}{T}) \left[\gamma_g(u;\psi) - \gamma_g(u;\psi)(1 + 1/T) \right] \right) .$$

By Lemma 15 as $u \to \infty, \gamma_g(u;\psi) = O(u^{-4d(\theta)+\delta})$, thus yielding

$$|\,(ii.2)\,| = O(T^{-4d(\theta)+\delta}) .$$

We can then conclude that $T^{1/2}E\,|\,(ii)\,| = o(1)$ as $T \to \infty$ concluding the proof.

□

Given the last result, we can consider the objective function expressed in integral form

$$\bar{Q}_T(\theta) = log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I(\lambda)}{K_y(\lambda;\psi)}d\lambda\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log\,(K_y(\lambda;\psi))\,d\lambda + 1 .$$

Theorem 14 *Under Assumptions $A_1', A_2', A_3', A_4', A_5('2), A_6'(2), A_7', B''$ one obtains*

$$T^{1/2}(\bar{\psi}_T - \psi_0) \to_d \mathcal{N}_p\left(0, M^{-1}VM^{-1}\right) ,$$

where

$$M(\psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} N(\lambda;\psi)N(\lambda;\psi)'d\lambda , \tag{4.14}$$

with

$$N(\lambda;\psi) = \frac{\partial log\,(K_y(\lambda;\psi))}{\partial\psi} - \left[\int_{-\pi}^{\pi}\frac{\partial log\,(K_y(\omega;\psi))}{\partial\psi}d\omega\right] , \tag{4.15}$$

and

$$\begin{aligned} V(\psi) &= 4\pi\int_{-\pi}^{\pi} g(\omega;\psi)g(\omega;\psi)'K_y^2(\omega)d\omega \\ &+ 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\omega_1;\psi)g(\omega_2;\psi)'\bar{Q}_{yyy}(-\omega_1,\omega_2,-\omega_2)d\omega_1 d\omega_2 , \end{aligned}$$

and where $M = M(\psi_0)$, $V = V(\psi_0)$ and $\bar{Q}_{yyy}(.,.,.) = 1/\sigma^{16}Q_{yyy}(.,.,.)$, viz. normalizing the trispectrum with respect to the parameter σ^2 .

Proof: The result is proved within the following set of lemmas (Lemma 17 to Lemma 22) where we show that

$$(i)\ \bar{M}_T(\psi) = \frac{\partial^2\bar{Q}_T(\psi)}{\partial\psi\partial\psi'} \to M(\psi) ,\ \text{a.s.} \tag{4.16}$$

uniformly for $\psi \in \Psi$,

$$(ii)\, T^{1/2} \frac{\partial \bar{Q}_T(\psi_0)}{\partial \psi} \rightarrow_d \mathcal{N}(0, V)\,. \tag{4.17}$$

Given the mean value theorem expansion together with (i) and (ii), the result will follows.

□

In order to prove (ii) we will take two steps.

First, there is the need to define a third kind of objective function by

$$\bar{Q^*}_T(\psi) = log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I^*(\lambda)}{K_y(\lambda;\psi)}d\lambda\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi} log\,\left(K_y(\lambda;\psi)\right)d\lambda + 1, \tag{4.18}$$

where

$$I^*(\lambda) = \frac{\mid \sum_{t=1}^{T}(y_t - \zeta)e^{it\omega}\mid^2}{2\pi T},$$

with

$$E(y_t) = \zeta\,.$$

We know from (2.11) that $\zeta = \mu^2 + \sigma^2(\rho^2 + \sigma^2\delta_{\alpha\alpha}(0;\theta_0))$ but we will not exploit this structure.

Thus we need to prove that as $T \rightarrow \infty$

$$(iii)\, T^{1/2} \frac{\partial \bar{Q^*}_T(\psi_0)}{\partial \psi} \rightarrow_d \mathcal{N}(0, V)\,,$$

and

$$(iv) T^{1/2} E \mid \frac{\partial \bar{Q}_T(\psi_0)}{\partial \psi} - \frac{\partial \bar{Q^*}_T(\psi_0)}{\partial \psi} \mid = o(1)\,.$$

so that (iii) and (iv) will give (ii).

First, let us consider (iii).

Lemma 17 *Under Assumptions* $A_1', A_2', A_3', A_4', B''$

$$T^{1/2} \frac{\partial \bar{Q^*}_T(\psi_0)}{\partial \psi} \rightarrow_d \mathcal{N}_p(0, V)\,.$$

Proof: We will establish (iii) by using two Lemmas. First, by using the method of cumulants, we will show in Lemma 18 that

$$T^{1/2}\left[\frac{\partial \bar{Q}^{*}{}_{T}(\psi_0)}{\partial \psi} - E\left(\frac{\partial \bar{Q}^{*}{}_{T}(\psi_0)}{\partial \psi}\right)\right] \rightarrow_d \mathcal{N}(0, V)\,. \tag{4.19}$$

Secondly in Lemma 19, we will show that

$$ET^{1/2}\left(\frac{\partial \bar{Q}^{*}{}_{T}(\psi_0))}{\partial \psi}\right) = o(1)\,. \tag{4.20}$$

So, by (4.19) and (4.20), the result follows.

□

Now, considering that

$$\frac{\partial \bar{Q}^{*}{}_{T}(\psi_0)}{\partial \psi} = \tag{4.21}$$
$$\frac{1}{\bar{\sigma}_T^8(\psi_0)}\frac{\int_{-\pi}^{\pi} g(\lambda;\psi_0)I^{*}(\lambda)d\lambda}{2\pi} + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial log\, K_y(\lambda;\psi_0)}{\partial \psi}d\lambda\,,$$

from Lemma 5, in view of the (strong) consistency result for $\bar{\sigma}_T^8(\psi_0)$, the following statistic, given by replacing the latter with σ^8, by using standard argument on convergence in distribution (Chow and Teicher 1978, Corollary 2 p.249), will have the same asymptotic distribution as (4.21) ,

$$\frac{\partial \bar{Q}_T^{\dagger}(\psi_0))}{\partial \psi} = \frac{1}{2\pi\sigma^8}\int_{-\pi}^{\pi} g(\lambda;\psi_0)I^{*}(\lambda)d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial log\, K_y(\lambda;\psi_0)}{\partial \psi}d\lambda + 1.$$

Remark: We need to use the latter type of objective function because it allows us to exploit the linearity of the cumulant operator.

The following result will be used:

Proposition 4 *(Fox and Taqqu 1987, Theorem 1)*
Suppose that f and g each satisfy the regularity condition. Suppose in addition that there exists $\alpha < 1$ and $\beta < 1$ such that for each $\delta > 0$

$$| f(x) | = O(| x |^{-\alpha-\delta}) \; as\, x \to 0 \, ,$$

and

$$| g(x) | = O(| x |^{-\beta-\delta}) \; as\, x \to 0 \, ,$$

Then

(a) if $p(\alpha+\beta) < 1$,

$$\lim_{N \to \infty} \frac{tr(R_N A_N)^p}{N} = (2\pi)^{2p-1} \int_{-\pi}^{\pi} [f(x)g(x)]^p dx \, , \tag{4.22}$$

and

(b) if $p(\alpha+\beta) \geq 1$,

$$tr(R_N A_N)^p = o(N^{p(\alpha+\beta)+\epsilon}) \, ,$$

for every $\epsilon > 0$ where R_N and A_N are respectively the symmetric Toeplitz matrices in the Fourier coefficients of $f(.)$ and $g(.)$ and $tr(.)$ denotes the trace operator.

Remarks: (1) A function $f(.)$ satisfies the *regularity condition* in Proposition 4 if the discontinuities of $f(.)$ have Lebesgue measure 0 and $f(.)$ is bounded on the interval $[\delta, \pi]$ for all the $\delta > 0$.
(2) Due to the non Gaussianity of the problem, we cannot use Proposition 4 directly but nevertheless we will use its results to derive some asymptotic results for traces of Toeplitz matrices.

Lemma 18 *Under Assumptions $A_1', A_2', A_3', A_4', B''$ as $T \to \infty$*

$$cum_k\left(A(\psi_0)_T\right) = cum_k\left(\underbrace{A(\psi_0)_T,)\ldots,A(\psi_0)_T}_{k\ times}\right) = O(T^{-(k-1)}) \;\; for\; k \geq 2 \, ,$$

where

$$A(\psi_0)_T = \frac{\int_{-\pi}^{\pi} g(\lambda;\psi_0) I^*(\lambda) d\lambda}{2\pi}.$$

Proof : Let us set

$$\hat{y}_t \stackrel{def}{=} y_t - E(y_t) = \sum_{u,v=1}^{\infty} \alpha_u \alpha_v \left(\epsilon_t^2 \epsilon_{t-u} \epsilon_{t-v} - \sigma^4 \delta(u,v)\right),$$

$y = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T)'$,

$g(\omega) = g(\omega;\psi_0)$, $\gamma_g(u) = \gamma_g(u;\psi_0)$,

$G_T = G_T(\psi_0) = \{\gamma_g(a-b)\}$, $a, b = 1, \dots, T$ a $T \times T$ symmetric Toeplitz matrix,

I_T the $T \times T$ identity matrix ,

$1_a = \underbrace{(1, \dots, 1)'}_{atimes}$ a $a \times 1$ vector of ones .

where we set both $\bar{\mu}$ and $\bar{\rho}^2$ equal to zero, given that the terms involving them are included within the analysis of the bilinear form $\hat{y}_t$.

Then we can express $A(\psi_0)_T$ as

$$A(\psi_0)_T = \frac{y' G_T(\psi_0) y}{(2\pi)^2 T} = \frac{1}{(2\pi)^2 T} \sum_{t_1, s_1 = 1}^{T} \gamma_g(t_1 - s_1) \hat{y}_{t_1} \hat{y}_{s_1}.$$

Then, by Brillinger (1975, (i), (iv), (v) p.19) we can write

$$cum_k\left(A(\psi_0)_T\right) = \tag{4.23}$$

$$\frac{1}{(2\pi)^{2k} T^k} \sum_{s_1, t_1 = 1}^{T} \sum_{s_2, t_2 = 1}^{T} \cdots \sum_{s_k, t_k = 1}^{T} \gamma_g(t_1 - s_1) \dots \gamma_g(t_k - s_k)\, cum_k(\hat{y}_{t_1} \hat{y}_{s_1}, \dots, \hat{y}_{t_k} \hat{y}_{s_k}).$$

Now, the nonlinearity of the model compels us to apply twice the result of Proposition 1; first, to reduce the generalized cumulants into the elemental ones (Barndorff-Nielsen and Cox 1989), for the process $\{\hat{y}_t\}$.

For this, we have to consider all the indecomposable partitions (Brillinger 1975, p.20) for the $k \times 2$ array

$$\begin{array}{cc} \hat{y}_{t_1} & \hat{y}_{s_1} \\ \hat{y}_{t_2} & \hat{y}_{s_2} \\ \dots & \dots \\ \hat{y}_{t_k} & \hat{y}_{s_k}, \end{array}$$

which leads to the expression for (B.1) as

$$cum_k\left(A(\psi_0)_T\right) = \tag{4.24}$$

$$= \frac{\sum_{s_1,t_1=1}^{T}\sum_{s_2,t_2=1}^{T}\cdots\sum_{s_k,t_k=1}^{T}}{(2\pi)^{2k}T^k}\gamma_g(s_1-t_1)\ldots\gamma_g(s_k-t_k)\ \times$$

$$\left[\sum_{\substack{D=D_{1_q}+\ldots+D_{q_q}\\ 1\le q\le k\ dim(D_{i_q})\ge 2}}^{*} C_y(D_{1_q})\ldots C_y(D_{q_q})\right], \tag{4.25}$$

where D_{i_q} denotes the $i-th$ element, (i.e the subset) over q of them, for one of the indecomposable partition of the set $D \equiv \{t_1, s_1, \ldots, t_k, s_k\}$, which can be expressed say as

$$D_{i_q} = \{\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_{\mu_{i_q}}\},$$

and

$$C_y(D_{i_q}) = cum_{\mu_{i_q}}\left(\hat{y}_{\tilde{r}_1}, \hat{y}_{\tilde{r}_2}, \ldots, \hat{y}_{\tilde{r}_{\mu_{i_q}}}\right),$$

with μ_{i_q} being the dimension of the subset D_{i_q} and where the symbol $dim(\mathcal{A})$ expresses the number of elements in the set $\mathcal{A}$.

The second step involves expanding the $\hat{y}_t$'s, in terms of the coefficients α_j 's and the ϵ_t 's and applying again Proposition 1. For example, expanding the term $C_y(D_{i_q})$ leads one to consider the $\mu_{i_q} \times 4$ array:

$$\begin{array}{cccc}
\epsilon_{t_{\tilde{r}_1}} & \epsilon_{t_{\tilde{r}_1}} & \epsilon_{t_{\tilde{r}_1-u_1}} & \epsilon_{t_{\tilde{r}_1-v_1}} \\
\epsilon_{t_{\tilde{r}_2}} & \epsilon_{t_{\tilde{r}_2}} & \epsilon_{t_{\tilde{r}_2-u_2}} & \epsilon_{t_{\tilde{r}_2-v_2}} \\
\vdots & \vdots & \vdots & \vdots \\
\epsilon_{t_{\tilde{r}_{\mu_{i_q}}}} & \epsilon_{t_{\tilde{r}_{\mu_{i_q}}}} & \epsilon_{t_{\tilde{r}_{\mu_{i_q}}-u_{\mu_{i_q}}}} & \epsilon_{t_{\tilde{r}_{\mu_{i_q}}-v_{\mu_{i_q}}}}
\end{array}.$$

For every elements of the third and fourth column there is a corresponding coefficient α_j. Then, when the elements of the third and fourth columns, for different rows (and either for different or same columns) , do hook, say $\epsilon_{\tilde{r}_i-u_i}$ and $\epsilon_{\tilde{r}_j-v_j}$ $j \neq i$ then a term equal to $\sum_{u=1}^{\infty}\alpha_u\alpha_{u+|\tilde{r}_j-\tilde{r}_i|}$ appears which, as $|\ \tilde{r}_j - \tilde{r}_i\ | \rightarrow \infty$ is asymptotically of order $|\ \tilde{r}_j - \tilde{r}_i\ |^{2d-1}$, given A_1' .

Alternatively, when one element from the first or second column hooks with one from the third and fourth column, say $\epsilon_{\tilde{r}_i}$ and $\epsilon_{\tilde{r}_j-u_j}$, obviously for different rows, then a term equal to $\alpha_{|\tilde{r}_i-\tilde{r}_j|}$ appears which as $|\ \tilde{r}_i - \tilde{r}_j\ | \rightarrow \infty$ is of order $|\ \tilde{r}_i - \tilde{r}_j\ |^{d-1}$. As we can see, different coefficients arise from different partitions but henceforth we will always denote by R_T the $T \times T$ Toeplitz matrix made by any of these sets of coefficients.

All the other combinations would lead either to a zero term (say when terms of the same row but from column one or two and from column three of four hook) or to a constant (say when terms of the same row and column three and four hook), with respect to the indexes (and permutation of) $\{s_1, t_1, \ldots, s_k, t_k\}$.

So the final expression for (B.1) is given by

$$cum_k\,(A(\psi_0)_T) = \tag{4.26}$$

$$= \frac{\sum_{s_1,t_1=1}^{T}\sum_{s_2,t_2=1}^{T}\cdots\sum_{s_k,t_k=1}^{T}}{(2\pi)^{2k}T^k}\gamma_g(s_1-t_1)\ldots\gamma_g(s_k-t_k)\times \tag{4.27}$$

$$\left[\sum_{\substack{D=D_{1_q}+\ldots+D_{q_q}\\ 1\le q\le k\, dim(D_{i_q})\ge 2}}^{*}\left(\sum_{\substack{D_{1_q}=E_{1,1_q}+E_{2,1_q}+\ldots+E_{\mu_{1_q},1_q}\\ dimE_{i,1_q}=2}}^{**} C_\epsilon(E_{1,1_q})C_\epsilon(E_{2,1_q})\ldots C_\epsilon(E_{\mu_{1_q},1_q})\right)\times\right.$$

$$\left.\ldots\left(\sum_{\substack{D_{q_q}=E_{1,q_q}+E_{2,q_q}+\ldots+E_{\mu_{q_q},q_q}\\ dimE_{i,q_q}=2}}^{**} C_\epsilon(E_{1,q_q})C_\epsilon(E_{2,q_q})\ldots C_\epsilon(E_{\mu_{q_q},q_q})\right)\right] \tag{4.28}$$

where

$$E_{i,j_q} \equiv \left\{(\alpha_n)^{\gamma_n}\epsilon_{\tilde{r}-n},\,(\alpha_m)^{\gamma_m}\epsilon_{\tilde{s}-m}\right\},1\le i\le q\,,1\le j_q\le q\,,$$
$$C_\epsilon(E_{i,j_q}) = (\alpha_n)^{\gamma_n}\,(\alpha_m)^{\gamma_m}cum_2(\epsilon_{\tilde{r}-n},\epsilon_{\tilde{s}-m}) = (\alpha_n)^{\gamma_n}\,(\alpha_n)^{\gamma_n}\sigma^2 I(\tilde{r}-n=\tilde{s}-m)\,,$$

with the two ϵ's entering into the expression for either one or two (different) y's of the element D_{j_q} of one of the indecomposable partitions and $\gamma_i \in \{0,1\}$.

Given the definition of *regularity condition* in Proposition 4, the results of Theorem 10 - 8 - 17 and Lemma 11 - 12 imply that both $f_y(\lambda;\psi)$ and $g(\lambda;\psi)$ satisfy them. In the following part of the Lemma and in Lemma 22 we will always write the asymptotic value (i.e. in terms of the $d's$) instead of the exact value (i.e. in terms of the α 's) for all the coefficients involved so that for example we will write $\mid u \mid^{2d-1}$ instead of $\sum_{i=1}^{\infty}\alpha_i\alpha_{i+u}$ etc. .

Let us consider the cumulant of order k=3. The set of indexes involved are $\{s_1, t_1, s_2, t_2, s_3, t_3\}$. The admissible combinations can be classified in two cases, case (A) made of two subsets of dimension two and four respectively and case (B) made of two subsets both of dimension three.

Remarks: (1) Depending on the partition considered, the exponent in any of the terms (and product of) arising from the cumulants evaluation will vary from case to case. The possible values for the exponent are $s(d-1)$, $s(2d-1)$ for $s = 1, 2, \ldots$ so that the maximum (negative) possible exponent is $(2d-1)$. We define a partition 'narrow' when it considers the least number of links. Obviously, the larger the partition, the easier it is to bound the cumulant expression in that the exponent will be bigger in absolute value (they are always negative) and/or there will be more terms 'linking' the indexes s_a, t_a, $a = 1, 2, 3$.

(2) In any case, by assumptions A_1, A_2 all the coefficients involved, function of the $\alpha' s$, are QMC coefficients and thus Proposition 2 applies yielding the asymptotic behaviour at the zero frequency of their fourier transforms. This result allows us to use Proposition 4 involving $g(\lambda; \psi) = \frac{\partial}{\partial \psi} K^{-1}(\lambda; \psi)$, evaluated at the true parameter value ψ_0 on the one hand and any of the functions by-product of the cumulants' evaluation on the other, the Toeplitz matrices for the latter denoted by R_T (we defined by G_T the Toeplitz matrix corresponding to $g(\lambda)$). Finally, in terms of the notation of Proposition 4, we set $\beta = 1 - 4d$ whereas α will vary from partition to partition.

(3) In general, from a set of k rows, we must consider at least (k-1) 'links'. (4) We can always take the sequence $\gamma_g(u)$ to be positive. In fact from $\gamma_g(u) = \gamma_g(u)^+ - \gamma_g(u)^-$ with $\gamma_g(u)^-, \gamma_g(u)^+ \geq 0$, we can bound $\mid \gamma_g(u) \mid$ by $\gamma_g(u)^+ + \gamma_g(u)^-$, following Dahlhaus (1989) and using Graybill (1983, Theorem 12.2.3(3)) .

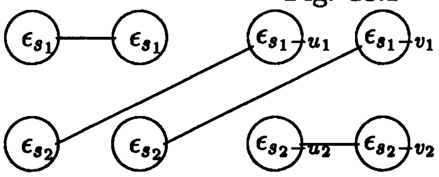
(5) The choice of the indexes in the two partitions is arbitrary.

(6) All the indexes t_a or s_b for $a, b = 1, 2, 3$ in the summation below run from 1 to T so that reference to the range of summation will always be skipped.

Case (A).

We must consider the 'least' indecomposable partitions. In particular we consider two subcases, case A.1 and case A.2 . In both cases, the partition of dimension two is the same and is showed in Fig. 18.1 , assuming without loss of generality that this subset of indexes is made of $\{s_1, s_2\}$

Fig. 18.1



Then we obtain that the dimension two partition generates a term of order

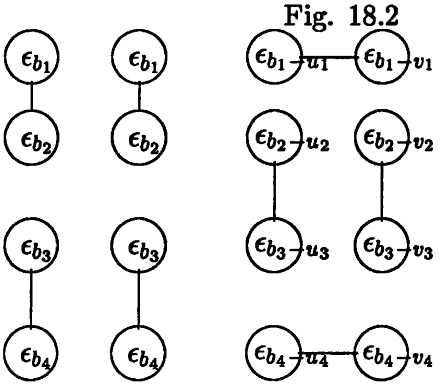
$$\{|s_1 - s_2|^{4d-2}\}.$$

Then let us consider the two possible subcases. The subset of indexes for the companion partition (of dimension four) is always given by $\{t_1, t_2, t_3, s_3\}$.

Subcase (A.1) .

The dimension four partition is displayed in Fig. 18.2 .

Fig. 18.2



Thus two terms are generated within this partition of order

$$\{|b_1 - b|^{2d-1}\}, \{|b_2 - b|^{2d-1}\}$$

where $\{b_1, b_2, b_3, b_4\} \equiv \{t_1, t_2, t_3, s_3\}$ and $b_i = b_j \; i \neq j$ for one couple of indexes only.

Now, setting $\nu = 1 - 2d$, we list the six possible cases as

1. $b = t_1 = t_2 \;\; , |t_3 - t|^{-\nu} \, |s_3 - t|^{-\nu},$
2. $b = t_1 = t_3 \;\; , |t_2 - t|^{-\nu} \, |s_3 - t|^{-\nu},$
3. $b = t_2 = s_3 \;\; , |t_2 - t|^{-\nu} \, |t_3 - t|^{-\nu},$
4. $b = t_2 = t_3 \;\; , |t_1 - t|^{-\nu} \, |s_3 - t|^{-\nu},$
5. $b = t_2 = s_3 \;\; , |t_1 - t|^{-\nu} \, |t_3 - t|^{-\nu},$
6. $b = t_3 = s_3 \;\; , |t_1 - t|^{-\nu} \, |t_2 - t|^{-\nu}.$

Let us consider case 1., viz. $b = t_1 = t_2$. Then the summation relative at this partition is bounded by

$$\begin{aligned}
&\sum_{t,s_1,s_2,s_3,t_3} \gamma_g(t-s_1)|s_1-s_2|^{-\nu}\gamma_g(s_2-t)\,|t-t_3|^{-\nu}\gamma_g(t_3-s_3)|s_3-t|^{-\nu} \\
&\sum_{s_1,s_2,s_3,t_3} |s_1-s_2|^{-\nu}\gamma_g(s_3-t_3) \times \\
&\times[\sum_t |t-t_3|^{-\nu}\gamma_g(s_1-t)|s_3-t|^{-\nu}\gamma_g(s_2-t)\,]. \qquad (4.29)
\end{aligned}$$

Now, by Schwarz's inequality and by Jensen's inequality for sums (Hardy, Littlewood, and Polya 1964, Theorem 19) we can bound the term in square brackets as follows,

$$|\sum a\,b\,c\,d\,| \leq (\sum a^2\,b^2)^{1/2}(\sum c^2 d^2)^{1/2} \leq [\sum a\,b]\,[\sum c\,d], \qquad (4.30)$$

where the summation runs with respect to a single index, viz. t , and we identify

$$a = \gamma_g(t-s_1)\,, b = |t-t_3|^{-\nu}\,, \gamma_g(t-s_2)\,, d = |t-s_3|^{-\nu}\,.$$

So (4.29) is bounded by

$$\sum_{s_1,t_1,s_2,t_2,s_3,t_3} \gamma_g(t_1-s_1)|s_1-s_2|^{-\nu}\gamma_g(s_2-t_2)|t_2-s_3|^{-\nu}\gamma_g(s_3-t_3)|t_3-t_1|^{-\nu},$$

which, by setting $-\nu = (2d-1)$ and using the matrix notation and

result (a) in Proposition 4 (setting $\alpha = 2d$ thus yielding $\alpha + \beta = 1 - 2d < 1/2$), as $T \to \infty$, behaves as

$$tr(G_T R_T G_T R_T G_T R_T) = O(T)\,,$$

Likewise, for the cases 2. – 5. , e.g. for 2. one should take, in (4.30),

$$a = \gamma_g(t - s_1), \;\; b = |t - s_3|^{-\nu}, \;\; c = \gamma_g(t - s_1), \;\; d = |t - t_2|^{-\nu}.$$

For case 6., instead, we have to consider a different bound, viz.

$$\sum_{s_1, t_1, s_2, t_2, t} \gamma_g(t_1 - s_1)|s_1 - s_2|^{-\nu}\gamma_g(s_2 - t_2)|t_2 - t|^{-\nu}|t - t_1|^{-\nu},$$

which, in matrix notation, can be written as

$$tr(G_T R_T G_T R_T^2) = tr(R_T G_T R_T G_T R_T)\,. \tag{4.31}$$

Henceforth I_T and 1_T denote the identity matrix and a vector of ones respectively both of dimension T. Now, by the continuity of the inverse of the spectrum we obtain

$$I_T \leq K G_T\,,$$

for a positive constant K big enough (the matrix G_T being positive definite) and so we can bound (4.31) as $T \to \infty$ by

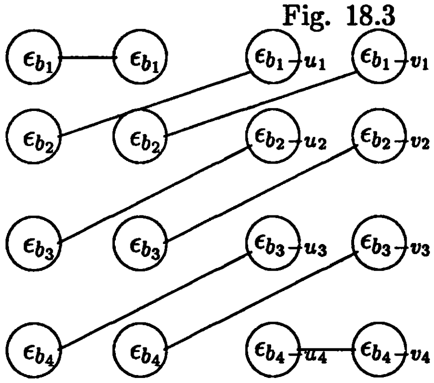
$$tr(G_T R_T G_T R_T G_T R_T) = O(T)\,.$$

Subcase (A.2) .

In this case the dimension four partition is displayed in Fig. 18.3 and yields the following set of coefficients

$$|b_1 - b_2|^{2d-2}, \;\; |b_2 - b_3|^{2d-2}, \;\; |b_3 - b_4|^{2d-2}.$$

Fig. 18.3



Let us set $\nu = 2 - 2d$. The possible cases are therefore

1. $|t_3 - t_1|^{-\nu}$, $|t_1 - t_2|^{-\nu}$, $|t_2 - s_3|^{-\nu}$,

2. $|s_3 - t_1|^{-\nu}$, $|t_1 - t_2|^{-\nu}$, $|t_2 - t_3|^{-\nu}$,

3. $|t_2 - t_1|^{-\nu}$, $|t_1 - t_3|^{-\nu}$, $|t_3 - s_3|^{-\nu}$,

4. $|s_3 - t_1|^{-\nu}$, $|t_1 - t_3|^{-\nu}$, $|t_3 - t_2|^{-\nu}$,

5. $|t_2 - t_1|^{-\nu}$, $|t_1 - s_3|^{-\nu}$, $|s_3 - t_3|^{-\nu}$,

6. $|t_3 - t_1|^{-\nu}$, $|t_1 - s_3|^{-\nu}$, $|s_3 - t_2|^{-\nu}$,

7. $|t_1 - t_2|^{-\nu}$, $|t_2 - t_3|^{-\nu}$, $|t_3 - s_3|^{-\nu}$,

8. $|s_3 - t_2|^{-\nu}$, $|t_2 - t_3|^{-\nu}$, $|t_3 - t_1|^{-\nu}$,

9. $|t_1 - t_2|^{-\nu}$, $|t_2 - s_3|^{-\nu}$, $|s_3 - t_3|^{-\nu}$,

10. $|t_3 - t_2|^{-\nu}$, $|t_2 - s_3|^{-\nu}$, $|s_3 - t_2|^{-\nu}$,

11. $|t_1 - t_3|^{-\nu}$, $|t_3 - s_3|^{-\nu}$, $|s_3 - t_2|^{-\nu}$,

12. $|t_2 - t_3|^{-\nu}$, $|t_3 - s_3|^{-\nu}$, $|s_3 - t_1|^{-\nu}$.

Now, the possible bounds are, as above, either of the form

$$tr(G_T R_T G_T R_T G_T R_T)\,,$$

or of the form

$$1_T' G_T R_T G_T R_T G_T 1_T \,. \tag{4.32}$$

In fact taking the case (5.) we obtain

$$\sum_{t_1\, s_1, t_2, s_2, t_3, s_3} (\gamma_g(t_3 - s_3)|s_3 - t_1|^{-\nu}\gamma_g(t_1 - s_1)|s_1 - s_2|^{-\nu}\gamma_g(t_2 - s_2)|t_2 - t_1|^{-\nu}|t_3 - s_3|^{-\nu})$$
$$= O\left(\sum_{t_1\, s_1, t_2, s_2, t_3, s_3} (\gamma_g(t_3 - s_3)|s_3 - t_1|^{-\nu}\gamma_g(t_1 - s_1)|s_1 - s_2|^{-\nu}\gamma_g(t_2 - s_2))\right)$$
$$= O(1_T' G_T R_T G_T R_T G_T 1_T),$$

for a positive constant K. But then, defining the Toepliz matrix of the Fourier coefficients of $1/g(\lambda)$ as

$$H_T = H_T(g^{-1}) \,,$$

we can write from Graybill (1983, Theorem 12.2.3)

$$\begin{aligned}
&1_T' G_T R_T G_T R_T G_T 1_T = 1_T' H_T^{-1/2} H_T^{1/2} G_T R_T G_T R_T G_T H_T^{1/2} H_T^{-1/2} 1_T \leq \\
&1_T' H_T^{-1} 1_T |H_T^{1/2} G_T R_T G_T R_T G_T H_T^{1/2}| = \\
&1_T' H_T^{-1} 1_T (tr(G_T R_T G_T R_T G_T H_T G_T R_T G_T R_T G_T H_T))^{1/2} \leq \\
&1_T' H_T^{-1} 1_T (tr(G_T H_T G_T H_T G_T H_T G_T H_T G_T H_T G_T H_T))^{1/2} \,.
\end{aligned} \tag{4.33}$$

From Adenstedt (1974)

$$1_T' H_T^{-1} 1_T \leq K T^{2-4d+\delta}$$

for any $\delta > 0$, whereas for the second factor in (4.33), by case (a) of Proposition 4 (setting $\alpha = 4d - 1$) as $T \to \infty$

$$tr(G_T H_T G_T H_T G_T H_T G_T H_T G_T H_T G_T H_T) = O(T) \,.$$

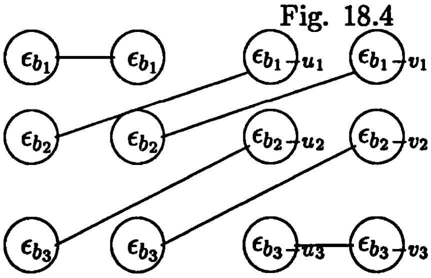
Case (B).

Let us now consider the possible partitions in two subpartitions both of dimension three. Again we have to distinguish two subcases, subcase (B.1) and subcase (B.2).

For subcase (B.1), displayed in Fig. 18.4 , we obtain the following links

$$|b_1 - b_2|^{2d-2}, \ |b_2 - b_3|^{2d-2}.$$

Fig. 18.4

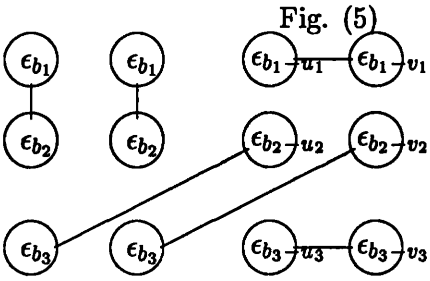


and for subcase (B.2), displayed in Fig. (5), we get instead the single link:

$$|b - b_3|^{2d-2},$$

with $b = b_1 = b_2$.

Fig. (5)



In both subcases the exponent will always be equal to $\nu = 2d - 2$.

All the possibilities are described if we consider two groups of subsets of indexes, namely $\{\{s_1, t_1, t_2\}, \{s_2, t_3, s_3\}\}$ on one hand, and $\{\{s_1, s_2, s_3\}, \{t_1, t_2, t_3\}\}$ on the other.

Now we have to apply the two subcases (B.1) and (B.2) to each of the four subsets of indexes, thus yielding six different cases that can be represented with obvious notation as

$$B.1(\{s_1, t_1, t_2\}) \text{ and } B.1(\{s_2, t_3, s_3\}),$$
$$B.1(\{s_1, t_1, t_2\}) \text{ and } B.2(\{s_2, t_3, s_3\}),$$

$$B.2(\{s_1, t_1, t_2\}) \text{ and } B.2(\{s_2, t_3, s_3\}),$$
$$B.1(\{s_1, s_2, s_3\}) \text{ and } B.1(\{t_1, t_2, t_3\}),$$
$$B.1(\{s_1, s_2, s_3\}) \text{ and } B.2(\{t_1, t_2, t_3\}),$$
$$B.2(\{s_1, s_2, s_3\}) \text{ and } B.2(\{t_1, t_2, t_3\}).$$

where we denote by $B.j(\{a, b, c\})$ the subcase $(B.j)$ applied to the subset of indexes $\{a, b, c\}$.

In particular, the following links are generated. Starting with $B.1(\{s_1, t_1, t_2\})$ we obtain the following :

1. $|t_1 - s_1|^{-\nu}, \ |s_1 - t_2|^{-\nu}$,
2. $|t_1 - t_2|^{-\nu}, \ |t_2 - s_1|^{-\nu}$,
3. $|s_1 - t_1|^{-\nu}, \ |t_1 - t_2|^{-\nu}$.

With respect to $B.2(\{s_1, t_1, t_2\})$ we obtain

1. $b = t_1 = s_1, \ |b - t_2|^{-\nu}$,
2. $b = t_1 = t_2, \ |b - s_1|^{-\nu}$,
3. $b = s_1 = t_2, \ |b - t_1|^{-\nu}$.

With respect to $B.1(\{s_2, t_3, s_3\})$ one gets

1. $|t_3 - s_3|^{-\nu}, \ |s_3 - s_2|^{-\nu}$,
2. $|t_3 - s_2|^{-\nu}, \ |s_2 - s_3|^{-\nu}$,
3. $|s_3 - t_3|^{-\nu}, \ |t_3 - s_2|^{-\nu}$.

and with respect to $B.2(\{s_2, t_3, s_3\})$,

1. $b = t_3 = s_3, \ |b - s_2|^{-\nu}$,

2. $b = t_3 = s_2, \ |b - s_3|^{-\nu}$,

3. $b = s_3 = s_2, \ |b - t_3|^{-\nu}$.

Equivalently with respect to the second group of subsets of indexes, $B.1(\{s_1, s_2, s_3\})$ gives

1. $|s_2 - s_1|^{-\nu}, \ |s_1 - s_3|^{-\nu}$,

2. $|s_2 - s_3|^{-\nu}, \ |s_3 - s_1|^{-\nu}$,

3. $|s_1 - s_2|^{-\nu}, \ |s_2 - s_3|^{-\nu}$.

With respect to $B.2(\{s_1, s_2, s_3\})$ we obtain

1. $b = s_2 = s_1, \ |b - s_3|^{-\nu}$,

2. $b = s_2 = s_3, \ |b - s_1|^{-\nu}$,

3. $b = s_1 = s_3, \ |b - s_2|^{-\nu}$.

With respect to $B.1(\{t_1, t_2, t_3\})$ one gets

1. $|t_2 - t_3|^{-\nu}, \ |t_3 - t_1|^{-\nu}$,

2. $|t_2 - t_1|^{-\nu}, \ |t_1 - t_3|^{-\nu}$,

3. $|t_3 - t_2|^{-\nu}, \ |t_2 - t_1|^{-\nu}$.

and with respect to $B.2(\{t_1, t_2, t_3\})$,

1. $b = t_2 = t_3, \ |b - t_1|^{-\nu}$,

2. $b = t_2 = t_1, \ |b - t_3|^{-\nu}$,

3. $b = t_3 = t_1, \ |b - t_2|^{-\nu}$.

As for case (A), we can bound the above expressions by either $tr((G_T R_T)^b)$ or $1_T' G_T (R_T G_T)^b 1_T$ for some integer number $b \geq 1$. The above results do not depend on the fact that we are using the cumulant of the 3rd order. In fact, in terms of Proposition 4 we are always able to find an α and a β such that $\alpha + \beta < 0$. This means that the result is generalisable to a cumulant of any order $k \geq 3$ yielding, as $T \to \infty$

$$T^{k/2}(cum_k\,(A(\psi_0)_T) = o(1)\,.$$

For the second order cumulant, denoting by $\sum^{**}$ the sum over all indecomposable partitions of the set of indexes $\{t_1, s_1, t_2, s_2\}$ (thus excluding the links t_1 with s_1 and t_2 with s_2), we obtain

$$cum_2(A(\psi_0)_T, A(\psi_0)_T) =$$
$$O(1/T^2 \sum_{t_1,s_1,t_2,s_2=1}^{T} \sum^{**} \gamma_g(t_1 - s_1)\gamma_g(t_2 - s_2) \mid \tilde{r}_1 - \tilde{r}_2 \mid^{4d-2} \mid \tilde{r}_3 - \tilde{r}_4 \mid^{4d-2}) = O(1/T)\,,$$

where the last equality is given by part (a) of Proposition 4 .

□

Lemma 19 *Under Assumption $A_1', A_2', A_3', A_4', B''$ as $T \to \infty$*

$$E(T^{1/2}\frac{\partial \bar{Q}_T^{\dagger}(\psi_0)}{\partial \psi}) = o(1)\,.$$

Proof : Setting $\gamma_g(u;\psi_0) = \gamma_g(u)$, from

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} w(\lambda;\psi)d\lambda = \frac{-1}{2\pi}\int_{-\pi}^{\pi} g(\lambda;\psi)K_y(\lambda;\psi)d\lambda\,,$$

by Parseval's relation we get

$$\frac{\partial \bar{Q}_T^{\dagger}(\psi_0)}{\partial \psi} = \tag{4.34}$$
$$\frac{1}{T(2\pi)^2\sigma^8}\sum_{i,j=1}^{T} \gamma_g(i-j)(y_i - \zeta)(y_j - \zeta) - \frac{1}{(2\pi)^2}\sum_{u=-\infty}^{\infty} \gamma_g(u)\gamma_y(u).$$

By taking expectations of (4.34) and rearranging terms one obtains

$$E\left(\frac{\partial \bar{Q}_T^{\dagger}(\psi_0)}{\partial \psi}\right) = -2\left(\sum_{u=T-1}^{\infty} \gamma_g(u)\,\gamma_y(u) + \sum_{u=1}^{T-1} \frac{u}{T}\gamma_g(u)\gamma_y(u)\right)\,. \tag{4.35}$$

Now, as $u \to \infty$,

$$\gamma_g(u)\gamma_y(u) \sim K\, u^{-2+\delta}$$

for any arbitrary $\delta > 0$ due to the logarithmic slowly varying function in $g(\lambda)$ so that both terms in (4.35) are of order $O(T^{-1+\delta})$ and we can always take a $\delta < 1/2$.

□

To show now the equivalence with respect to the mean-adjusting let us consider the following result.

Proposition 5 *(Fox and Taqqu 1987, Lemma 8.1 p.238)*
Let us suppose that f and g each satisfy the conditions in Proposition 4 with $\alpha + \beta < 1/2$. Let x_N be an N-dimensional vector of observations of a process $\{x_n\}$ with spectral density $f(.)$, with $E(\hat{x}_N\, \hat{x}'_N) = A_N(f)$, $\hat{x}_N = x_N - E(x_N)$.

$$lim_{N\to\infty} \frac{1}{N^{1/2}} E \mid x'_N R_N(g) x_N - \hat{x}'_N R_N(g)\hat{x}_N \mid = 0\,.$$

Lemma 20 *Under Assumptions $A'_1, A'_2, A'_3, A'_4, B''$ as $T \to \infty$*

$$ET^{1/2} \mid \frac{\partial \bar{Q}_T(\psi_0)}{\partial \psi} - \frac{\partial \bar{Q}^*{}_T(\psi_0)}{\partial \psi} \mid = o(1)\,.$$

Proof : The result follows by a direct use of Proposition 5. In fact under our assumptions both $K_y(\lambda)$ and $g(\lambda;\psi_0)$ satisfy the *regularity condition* with $\alpha + \beta = 0 < 1/2$ and thus all the hypotheses of Proposition 5 are valid. Note how the Gaussianity assumption in Fox and Taqqu (1987) is irrelevant for this result to hold, as only expressions in the second moments are involved.

□

Thus we have established the weak convergence result (ii) in Theorem 14. Now we need to show the (uniform) convergence result of the hessian matrix as expressed in (i) of Theorem 14.

Let us consider now the hessian matrix for the 'sum-form' objective function $\widehat{Q}_T$. The result follows much more easily for the 'integral-form' $\bar{Q}_T$.

Lemma 21 *Under Assumptions $A_1', A_2', A_3', A_4', A_5(2)', A_6(2)', B''$ as $T \to \infty$*

$$\widehat{M}_T(\psi) = \frac{\partial^2 \widehat{Q}_T(\psi)}{\partial \psi \partial \psi'} \to M(\psi) \ a.s.,$$

uniformly in $\psi \in \Psi$.

Proof : Now, $\widehat{M}_T(\psi)$ can be written as

$$\widehat{M}_T(\psi) = (a) + (b) + (c) + (d),$$

with

$$(a) = 1/\left(\frac{1}{T}\sum_j \frac{I(\lambda_j)}{K_y(\lambda_j;\psi)}\right) \frac{1}{T}\sum_j \frac{\partial^2 K_y^{-1}(\lambda_j;\psi)}{\partial\psi\partial\psi'} I(\lambda_j),$$

$$(b) = -1/\left(\frac{\sum_j \frac{I(\lambda_j)}{K_y(\lambda_j;\psi)}}{T}\right)^2 \left(\frac{1}{T}\sum_j \frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi} I(\lambda_j)\right)\left(\frac{1}{T}\sum_j \frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi'} I(\lambda_j)\right),$$

$$(c) = \frac{1}{T}\sum_j \frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial K_y(\lambda_j;\psi)}{\partial\psi'}, \ (d) = \frac{1}{T}\sum_j K_y^{-1}(\lambda_j;\psi)\frac{\partial^2 K_y(\lambda_j;\psi)}{\partial\psi\partial\psi'}.$$

Now, for (a) and (b), given the results in Lemma 12 and the argument (and the result) used in Lemma 5, it follows that

$$(a) \quad \to_{a.s.} \frac{1}{\left(\frac{\int_{-\pi}^{\pi}\frac{K_y(\lambda)}{K_y(\lambda;\psi)}d\lambda}{2\pi}\right)}\left(\frac{\int_{-\pi}^{\pi}\frac{\partial^2 K_y^{-1}(\lambda;\psi)}{\partial\psi\partial\psi'}K_y(\lambda)d\lambda}{2\pi}\right),$$

$$(b) \quad \to_{a.s} \frac{1}{\left(\frac{\int_{-\pi}^{\pi}\frac{K_y(\lambda)}{K_y(\lambda;\psi)}}{2\pi}\right)^2}\left(\frac{\int_{-\pi}^{\pi} g(\lambda;\psi)K_y(\lambda)d\lambda}{2\pi}\right)\left(\frac{\int_{-\pi}^{\pi} g(\lambda;\psi)K_y(\lambda)d\lambda}{2\pi}\right).$$

The last two terms, (c) and (d), even being not stochastic, need more attention due to the behaviour at the zero frequency of the power spectrum.

We can consider only the proof for (c), given that the results will follow for (d) just by considering that they have the same behaviour in the neighborhood of the zero frequency

(Lemma 12). Then, following Lemma 6, let us consider

$$\frac{1}{T}\sum_{j=1}^{|T/2|}\frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial K_y(\lambda_j;\psi)}{\partial\psi'} = (c.1)\,+\,(c.2)\,+(c.3)\,,$$

$$(c.1) = \frac{1}{T}\sum_{j=1}^{m-1}\frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial K_y(\lambda_j;\psi)}{\partial\psi'},$$

$$(c.2) = \frac{1}{T}\sum_{j=m}^{j^{\epsilon}}\frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial K_y(\lambda_j;\psi)}{\partial\psi'},$$

$$(c.3) = \frac{1}{T}\sum_{j=j_\epsilon}^{|T/2|}\frac{\partial K_y^{-1}(\lambda_j;\psi)}{\partial\psi}\frac{\partial K_y(\lambda_j;\psi)}{\partial\psi'}.$$

Now, by Lemma 12 and Lemma 3-(part (i)) as $T\to\infty$

$$(c.3)\to\frac{1}{2\pi}\int_{\epsilon}^{\pi}\frac{\partial K_y^{-1}(\lambda;\psi)}{\partial\psi}\frac{\partial K_y(\lambda;\psi)}{\partial\psi'}d\lambda\,.$$

Now from Theorem 10and Lemmas 8 17 one obtains that for any $i,j = 1,2,\ldots,(p+2)$ as $\lambda\to 0^+$

$$\frac{\partial K_y^{-1}(\lambda;\psi)}{\partial\psi_i}\frac{\partial K_y(\lambda;\psi)}{\partial\psi_j} = O(log^2(\frac{1}{\lambda}))\,.$$

This yields as $T\to\infty$

$$(c.1) = O(\frac{\sum_{j=1}^{m-1}log^2(\lambda_j)}{T}) = O(log^2(\lambda_m)\lambda_m) = o(1)\,.$$

Finally, given

$$(c.2)' = \frac{\sum_{j=m}^{j_\epsilon}log^2(\lambda_j)}{T} - \frac{\int_{\lambda_m}^{\epsilon}log^2(\lambda)d\lambda}{2\pi},$$

$$(c.2)'' = \int_0^{\lambda_m}log^2(\lambda)d\lambda\,,$$

we get

$$|\,(c.2)\,|\le|\,(c.2)'\,|\,+\,|\,(c.2)''\,|\,.$$

Thus we just need to prove that $|\,(c.2)'\,|= o(1)$ as $T\to\infty$ given that solving the integral $|\,(c.2)''\,|= O(\lambda_m\,log^2(\lambda_m))$.

Now

$$|\,(c.2)'\,|\le|\,\frac{log^2(\lambda_m)}{T}\,|+\frac{1}{2\pi}\sum_{j=m}^{j_\epsilon-1}\int_{\lambda_j}^{\lambda_{j+1}}\left(log^2(\lambda_{j+1}) - log^2(\lambda)\right)d\lambda\,|\,. \tag{4.36}$$

The second term can be bounded as follows.

$$|\sum_{j=m}^{j_\epsilon-1}\int_{\lambda_j}^{\lambda_{j+1}}\Big(log^2(\lambda_{j+1}) - log^2(\lambda)\Big)\,d\lambda\,|\leq|\,\frac{2\pi}{T}\sum_{j=m}^{j_\epsilon-1}\Big(log^2(\lambda_{j+1}) - log^2(\lambda_j)\Big)\,|$$
$$=\frac{2\pi}{T}\,|\,\Big(log^2(\lambda_{j_\epsilon-1}) - log^2(\lambda_m)\Big)\,|=\frac{2\pi}{T}\,|\,(log(j_\epsilon/m)log(\lambda_{j_\epsilon+m-1}))\,|=O(\frac{log\,(j_\epsilon)}{T}),$$

where all the bounds are independent of ψ.

$\square$

After some algebra the limit matrix $M = M(\psi_0)$ takes the expression (4.14) which is assumed not singular (it is certainly positive semi-definite by construction).

Proposition 6 *(Hosoya 1993, Lemma A4.2)*
For a fourth-order stationary process, with fourth order cumulant $Q_4(.,.,.)$ and trispectrum $\tilde{Q}_4(.,.,.)$ and where $h(.)$ denotes an even continuous function in $[-\pi,\pi]$ with Fourier coefficients $\gamma_h(u)$, under the following two conditions

F_1 $\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|h(\omega_1)\,h(\omega_2)Q_4(\omega_1,-\omega_2,\omega_2)|d\omega_1 d\omega_2 < \infty.$

F_2 *For any $\epsilon > 0$,*

$$\sup_{|\lambda_i|<\epsilon,i=1,2,3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|h(\omega_1)h(\omega_2)\,|\times$$
$$\Big[\tilde{Q}_4(\omega_1+\lambda_1,\omega_2+\lambda_2,\omega_3+\lambda_3)-\tilde{Q}_4(\omega_1,-\omega_2,\omega_2)\Big]\,|d\omega_1 d\omega_2 = O(\epsilon)\,,$$

we get

$$\lim_{T\to\infty}\frac{1}{T}\sum_{s_1,t_1,s_2,t_2}\gamma_h(t_1-s_1)\gamma_h(t_2-s_2)Q_4(s_1-t_1,t_2-s_1,s_2-t_1)=$$
$$=2\pi\int_{-\pi}^{\pi}h(\omega_1)h(\omega_2)\tilde{Q}_4(\omega_1,-\omega_2,\omega_2)d\omega_1 d\omega_2\,.$$

Lemma 22 *Under Assumptions $A_1', A_2', A_3', A_4', B''$*

$$\lim_{T\to\infty}\quad T\,var(A(\psi_0)_T)=4\pi\int_{-\pi}^{\pi}g(\omega;\psi_0)g(\omega;\psi_0)'(f_y(\omega))^2 d\omega$$
$$+\quad 2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}g(\omega_1;\psi_0)g(\omega_2;\psi_0)'Q_{yyy}(-\omega_1,\omega_2,-\omega_2)d\omega_1 d\omega_2\,.$$

Proof: Setting $\gamma_g(u;\psi_0) = \gamma_g(u)$ and $cum_{yyy}(.,.,.)$ the fourth order cumulant of y_t, evaluating the variance and given the non Gaussianity of the process, we obtain the covariance terms and the (fourth order) cumulant term:

$$
\begin{aligned}
& var(A(\psi_0)_T) = var(1/(2\pi)^2 \int_{-\pi}^{\pi} g(\omega;\psi_0) I^*(\omega) d\omega) \\
& = \frac{1}{T^2(2\pi)^2} \left[\sum_{t_1,s_1,t_2,s_2} \gamma_g(t_1 - s_1)\gamma_g(t_2 - s_2) \left[\gamma_y(t_1 - s_2)\gamma_y(t_2 - s_1) + \gamma_y(t_1 - t_2)\gamma_y(s_2 - s_1)\right] \right] \\
& + \frac{1}{T^2(2\pi)^2} \left[\sum_{t_1,s_1,t_2,s_2} \gamma_g(t_1 - s_1)\gamma_g(t_2 - s_2) cum_{yyy}(s_1 - t_1, t_2 - t_1, s_2 - t_1) \right] \\
& = (i) + (ii).
\end{aligned}
$$

We will use two different results to prove, in turn, the convergence of the terms involving the spectrum and the term involving the fourth cumulant.

With respect to (i), we use Proposition 4 part (a) directly by identifying $\alpha = 1 - 4d = -\beta$, and $p = 1$, so that

$$
\lim_{T\to\infty} (i) = 4\pi \int_{-\pi}^{\pi} g(\omega;\psi_0) g(\omega;\psi_0)' (f_y(\omega))^2 d\omega .
$$

With respect to the second term involving the fourth cumulant we will rely on the result of Proposition 6. We will show how conditions F_1 and F_2 of Proposition 6 are satisfied, therefore concluding the proof. With respect to condition F_1, unravelling the trispectrum in terms of the fourth-order cumulant we obtain

$$
\begin{aligned}
& \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\omega_1;\psi_0) g(\omega_2;\psi_0) Q_{yyy}(\omega_1, -\omega_2, \omega_2) d\omega_1 d\omega_2 \\
& = \frac{1}{(2\pi)^3} \sum_{\substack{j_h=-\infty \\ h=1,2,3}}^{\infty} \left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\omega_1;\psi_0) g(\omega_2;\psi_0) e^{-i(\omega_1 j_1 - \omega_2 j_2 + \omega_2 j_3)} d\omega_1 d\omega_2 \right) cum_{yyy}(j_1, j_2, j_3)) \\
& = \frac{1}{(2\pi)^3} \sum_{\substack{j_h=-\infty \\ h=1,2,3}}^{\infty} \gamma_g(j_1)\gamma_g(j_2 - j_3) cum_{yyy}(j_1, j_2, j_3) .
\end{aligned}
$$

In order to bound the last expression we need to evaluate the fourth-order cumulant. Thus we have to consider all the possible indecomposable partitions of the 4 × 4 array

$$\begin{array}{cccc}
\epsilon_t & \epsilon_t & \epsilon_{t-v} & \epsilon_{t-u} \\
\epsilon_{t+j_1} & \epsilon_{t+j_1} & \epsilon_{t+j_1-v_1} & \epsilon_{t+j_1-u_1} \\
\epsilon_{t+j_2} & \epsilon_{t+j_2} & \epsilon_{t+j_2-v_2} & \epsilon_{t+j_2-u_2} \\
\epsilon_{t+j_3} & \epsilon_{t+j_3} & \epsilon_{t+j_3-v_3} & \epsilon_{t+j_3-u_3} .
\end{array}$$

Following the procedure of Lemma 18, when elements of the third or the fourth column hook one with the other (for different rows) or hook with elements of the first and second column (again for different rows) terms like $\alpha_{|j_a-j_b|}$ or $\sum_{u=1}^{\infty}\alpha_u\alpha_{u+|j_a-j_b|}$ arise. In the sequel we will express the asymptotic expressions of these coefficients as $| j_a - j_b |^{d-1}$ and $| j_a - j_b |^{2d-1}$ instead of their exact expression.

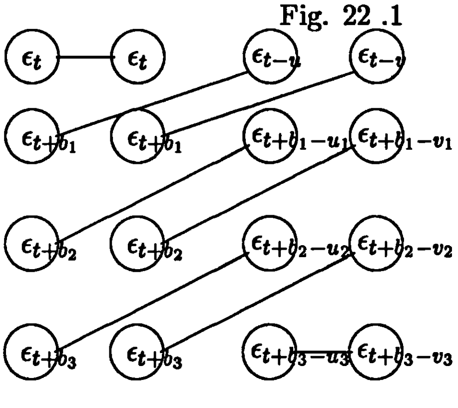
We will rely on the result in Proposition 2, given the QMC of the Fourier coefficients involved[3] and the property that the Fourier transform of a convolution of two functions is equal to the product of their Fourier transforms (Zygmund 1977, Theorem 1.5).

Now the possible 'least' indecomposable partitions can be summarized in three main categories denoted P_1, P_2, P_3 as follows :

P_1 (see Fig. 22.1)

$$|b_1|^{2d-2}, \ |b_1 - b_2|^{2d-2}, \ |b_2 - b_3|^{2d-2} .$$

Fig. 22 .1

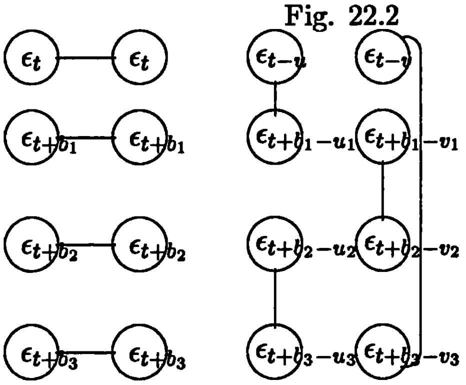


[3]We could instead make some regularity assumptions on the function involved as in Fox and Taqqu (1986, Lemma 4 ; Lemma 5).

P_2 (see Fig. 22.2)

$$b_1 = b_2 = b, \quad |b|^{4d-2}\,|b - b_3|^{2d-2}.$$

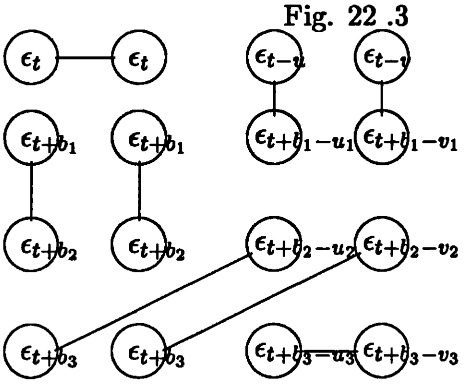
Fig. 22.2



P_3 (see Fig. 22.3)

$$|b_1|^{2d-1}, \quad |b_1 - b_2|^{2d-1}, \quad |b_2 - b_3|^{2d-1}, \quad |b_3|^{2d-1}.$$

Fig. 22 .3



For all the other partitions the result follows directly. For instance, taking the case of the partition for which $\epsilon_{t+j_i} = \epsilon_t$ for $i = 2, 3$ then condition F_1 holds.

Now starting with P_1, the possible combinations are given by:

1. $|j_1|^{2d-2}, \quad |j_1 - j_2|^{2d-2}, \quad |j_2 - j_3|^{2d-2},$

2. $|j_2|^{2d-2}, \quad |j_2 - j_1|^{2d-2}, \quad |j_1 - j_3|^{2d-2},$

3. $|j_3|^{2d-2}$, $|j_3 - j_2|^{2d-2}$, $|j_2 - j_1|^{2d-2}$.

With the index of the summations always running from 1 to ∞, we need to evaluate the following

$$\sum_{j_1, j_2, j_3} |j_1|^{-2d-2}|j_2 - j_3|^{-2d-2}|j_1 - j_2|^{2d-2}. \tag{4.37}$$

The following coefficients

$$\gamma_{1.1}(j) = \sum_u |u|^{-2d-2}|u - j|^{2d-2},$$

represent the Fourier coefficient of the function product of the Fourier transform of $|u|^{-2d-2}$ and of $|u|^{2d-2}$, say $f_1(\omega)$ and $f_2(\omega)$ respectively whose behaviour at zero frequency is given by as $\omega \to 0^+$

$$f_1(\omega) \sim K\omega^{2d+1},$$

$$f_2(\omega) \sim K\omega^{1-2d}.$$

Then the Fourier coefficients of the product function, which in turn behaves as $\omega \to 0^+$

$$f_1(\omega)f_2(\omega) \sim K\omega^2,$$

by Proposition 2 behaves asymptotically as $j \to \infty$

$$\gamma_{1.1}(j) \sim K|j|^{-3}.$$

Thus (4.37) can be bounded by

$$O(\sum_{j_1 j_2} |j_1|^{-3}|j_1 - j_2|^{-2d-2}) = O(\sum_j |j|^{2d-4}) = O(1).$$

The other two cases of P_1 follow by the same argument.

Taking now the case of P_2, the possible combinations are given by

1. $j_1 = j_2 = j$, $|j|^{4d-2}$, , $|j_3 - j|^{4d-2}$,

2. $j_1 = j_3 = j$, $|j|^{4d-2}$, $|j_2 - j|^{4d-2}$,

3. $j_2 = j_3 = j$, $|j|^{4d-2}$, $|j_1 - j|^{4d-2}$.

Considering the first case and following the same approach as for P_1 we obtain

$$\sum_{j,j_3} |j|^{-2}|j-j_3|^{-2} = O(\sum_j |j|^{-3}) = O(1),$$

and likewise for the other two terms.

With respect to P_3, we obtain

1. $|j_1|^{2d-1}$, $|j_2-j_1|^{2d-1}$, $|j_3-j_2|^{2d-1}$, $|j_3|^{2d-1}$,
2. $|j_1|^{2d-1}$, , $|j_1-j_3|^{2d-1}$, , $|j_2-j_3|^{2d-1}$, $|j_2|^{2d-1}$,
3. $|j_2|^{2d-1}$, $|j_2-j_1|^{2d-1}$, $|j_1-j_3|^{2d-1}$, $|j_3|^{2d-1}$,
4. $|j_2|^{2d-1}$, $|j_2-j_3|^{2d-1}$, $|j_3-j_1|^{2d-1}$, $|j_1|^{2d-1}$,
5. $|j_3|^{2d-1}$, $|j_3-j_2|^{2d-1}$, $|j_2-j_1|^{2d-1}$, $|j_1|^{2d-1}$,
6. $|j_3|^{2d-1}$, $|j_3-j_1|^{2d-1}$, $|j_1-j_2|^{2d-1}$, $|j_2|^{2d-1}$.

Considering the first case, we have to consider the order of magnitude of

$$O(\sum_{j_1,j_2,j_3} |j_1|^{-2d-1}|j_2-j_3|^{-2d-1}|j_1-j_2|^{2d-1}|j_3|^{2d-1}). \tag{4.38}$$

In particular we obtain that for any $\delta > 0$ the following coefficient

$$\gamma_{3,1}(u) = \sum_j |j|^{-2d-1}|j-u|^{2d-1-\delta},$$

bounds each of the two convolutions in (4.38).

Then defining by $f_3(.)$ and $f_3(.)$ the Fourier transforms of $|u|^{-2d-1}$ and $|u|^{2d-1-\delta}$ respectively, we get that $\gamma_{3,1}(u) \sim K|u|^{-1-\delta}$ as $u \to \infty$ so that we can bound (4.38) by

$$O(\sum_j |j|^{-2-2\delta}) = O(1).$$

For the other five subcases of P_3 the bound follows along the same lines, thus satisfying condition F_1.

Concerning condition F_2 for a given integer $M < \infty$, let us split the trispectrum as follows:

$$Q_{yyy}(\omega_1, \omega_2, \omega_3) = \frac{1}{(2\pi)^3} \sum_{j_i=-\infty;i=1,2,3}^{\infty} cum_{yyy}(j_1, j_2, j_3) e^{-i(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)}$$

$$= \frac{1}{(2\pi)^3} \sum_{j_i=-M;i=1,2,3}^{M} cum_{yyy}(j_1, j_2, j_3) e^{-i(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)}$$

$$+ \frac{1}{(2\pi)^3} \sum_{|j_i|>M;i=1,2,3} cum_{yyy}(j_1, j_2, j_3) e^{-i(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)}$$

$$+ \frac{1}{(2\pi)^3} \sum_{\substack{|j_a|>M,\, for\, at\, least\, one\, a=1,2,3 \\ |j_b|\leq M,\, for\, at\, least\, one\, b=1,2,3}} cum_{yyy}(j_1, j_2, j_3) e^{-i(\omega_1 j_1 + \omega_2 j_2 + \omega_3 j_3)}$$

$$= Q_M(\omega_1, \omega_2, \omega_3) + Q^{*}_M(\omega_1, \omega_2, \omega_3) + Q^{**}_M(\omega_1, \omega_2, \omega_3).$$

We can write

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |\, g(\omega_1; \psi_0)\, g(\omega_2; \psi_0)' \times$$

$$\times [Q_{yyy}(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3) - Q_{yyy}(\omega_1, -\omega_2, \omega_2)]\, |d\omega_1\, d\omega_2 =$$

$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |g(\omega_1; \psi_0)\, g(\omega_2; \psi_0)' \times$$

$$\times [Q_M(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3) - Q_M(\omega_1, -\omega_2, \omega_2) +$$

$$+ Q^{*}_M(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3) - Q^{*}_M(\omega_1, -\omega_2, \omega_2) +$$

$$+ Q^{**}_M(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3) - Q^{**}_M(\omega_1, -\omega_2, \omega_2)]\, |d\omega_1 d\omega_2 \leq$$

$$\leq A_1 + A_2 + A_3,$$

with

$$A_1 =$$

$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |g(\omega_1; \psi_0)\, g(\omega_2; \psi_0)\, [Q_M(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3) - Q_M(\omega_1, -\omega_2, \omega_2)]\, |d\omega_1 d\omega_2,$$

$$A_2 =$$

$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |g(\omega_1; \psi_0)\, g(\omega_2; \psi_0) \times$$

$$|\, (|\, Q^{**}_M(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3)\, | + |\, Q^{**}_M(\omega_1, -\omega_2, \omega_2)\, |)\, d\omega_1 d\omega_2,$$

$$A_3 =$$

$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |g(\omega_1; \psi_0)\, g(\omega_2; \psi_0) \times$$

$$|\, (|\, Q^{*}_M(\omega_1, -\omega_2, \omega_2)\, | + |\, Q^{*}_M(\omega_1 + \lambda_1, -\omega_2 + \lambda_2, \omega_2 + \lambda_3)\, |)\, d\omega_1 d\omega_2.$$

Now by choosing M large enough, A_2 and A_3 can be made arbitrary small by part (i) of the Lemma.

By Mitrinovic (1970, Theorem 3.8.22) for complex Z such that $|Z| < 1$ one obtains $|e^Z - 1| < |z|7/4$, we get

$$
\begin{aligned}
&A_1 = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |g(\omega_1;\psi_0)\, g(\omega_2;\psi_0) \times \\
&\times (Q_M(\omega_1+\lambda_1, -\omega_2+\lambda_2, \omega_2+\lambda_3) - Q_M(\omega_1, -\omega_2, \omega_2))\, |d\omega_1 d\omega_2 \leq \\
&\frac{1}{(2\pi)^3} \sum_{j_a=-M;a=1,2,3}^{M} |\, cum_{yyy}(j_1, j_2, j_3) \left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\omega_1)\, g(\omega_2) e^{-i(\omega_1 j_1 - \omega_2 j_2 + \omega_2 j_3)} d\omega_1 d\omega_2 \right) | \times \\
&\times |1 - e^{-i(\lambda_3 j_1 + \lambda_4 j_2 + \lambda_5 j_3)}| \,.
\end{aligned}
$$

Now we can always choose some $\epsilon' > 0$ and $\epsilon'' > 0$ such that $|\lambda_a| < \epsilon' \, a = 1,2,3$ and setting $z = -i(\lambda_3 j_1 + \lambda_4 j_2 + \lambda_5 j_3)$ where $|j_a| \leq M$, $a = 1,2,3$, such that $|z| < \epsilon''/2$. But from $|e^z - 1| < \epsilon''$, again by part (i) of the Lemma, one obtains that for some arbitrary $\epsilon > 0$

$$A_1 < \epsilon \,.$$

□

Chapter 5

The 'one-shock' model: small sample properties and applications of the PMLE

In this chapter we will estimate the 'one-shock' model both with simulated data as well as using the data set introduced in Chapter 1.

5.1 Small sample properties of the PMLE

5.1.1 The 'one-shock' model : simulation

There are many ways of simulating long memory processes in the Gaussian and linear case. A very successful algorithm to simulate stationary Gaussian time series is the frequency domain method of Davies and Harte (1987), which is computationally fast in view of the possibility to use the FFT. For a class of linear processes, namely the $ARFIMA(p, d, q)$, d not integer, Hosking (1984) shows how to use efficiently recursive expressions for the conditional mean and variance to simulate the process. Alternatively, for more general linear processes, the aggregation idea of Robinson (1978b) and Granger (1980) can be employed, particularly suited when dealing with very long sample sizes (Beran 1994).

Unfortunately, all of these methods are of no use for our problem. In fact our model is by construction nonlinear and non Gaussian, even if the simulated disturbances are drawn from a standard normal distribution (cf. section 3.2.1). For this reason we adopt a different approach which yields a simulated series exactly characterized by the theoretical correlation structure of the 'one-shock' model that is without employing some truncation.

The algorithm works as follows. For a given choice of the weights $\alpha_i(\theta),\ i = 1, 2, \ldots$ let us define the partial sum

$$\delta_i^2(\theta) = \begin{cases} \sum_{k=1}^{\infty} \alpha_k^2(\theta), & i = 1, \\ \delta_{i-1}^2(\theta) - \alpha_{i-1}^2(\theta), & i \geq 2. \end{cases} \tag{5.1}$$

Let us choose the values for μ, ρ, σ^2, the last one being strictly positive.

Then, to produce a simulated sample of dimension T, let us draw $T + 1$ independent realizations of a random variable with mean zero and variance σ^2, say from a $\mathcal{N}(0, \sigma^2)$, obtaining the $T + 1$ vector $(\eta, e_1, \ldots, e_T)'$. Thus the simulated vector of data $x = (x_1, \ldots, x_T)'$ is given by:

$$x_t = \mu + e_t \left(\rho + \delta_t(\theta)\eta + \sum_{k=1}^{t-1} \alpha_k(\theta) e_{t-k} \right), \quad t = 1, \ldots, T. \tag{5.2}$$

for a given vector of parameter values $\theta, \mu, \rho, \sigma^2$ and a given distribution for the $\eta, e_t,\ t = 1, \ldots, T$.

From a computational point of view, given that for each $t = 1, \ldots, T$ a convolution between the e_{t-k} and the α_k, $k = 1, \ldots, t-1$ is involved, we can use the fast Fourier transform in order to convolve them taking the product of the results and then deconvolving (cf. section 3.2.2 for another application of this idea). The desired convolution is the value of the deconvolved function at the Fourier frequency $2\pi(t-1)/(t-1)'$ with $t' = exp[log(t)/log(2)]$, $[.]$ denoting the integer part operator.

Finally, in order to make the importance of the initial value η as small as possible we generated $T+M$ realizations of the e_t with M=1,000 but so that the $t-th$ simulated observation

is
$$x_t = \mu + e_{t+M}\left(\rho + \delta_{t+M}(\theta)\eta + \sum_{k=1}^{t+M-1} \alpha_k(\theta)e_{t+M-i}\right), \quad t = 1, \dots, T. \tag{5.3}$$

5.1.2 Montecarlo results

In the sequel we set $\sigma^2 = 1$ and in turn $\rho \in \{0, 1, 10\}$. We set $\theta = d$ with

$$\alpha_i(d) = \prod_{j=1}^{i} \frac{j+d-1}{j}, \; 0 < d < 1/2, \;\; i = 1, 2, \dots,$$

and where in turn $d \in \{0.1, 0.2, 0.375, 0.45, 0.49\}$. In particular this implies that

$$\delta_1^2 = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} - 1.$$

where $\Gamma(n)$ denotes the Gamma function. We generated the η, e_t, $t = 1, \dots, T+M$ as randomly drawn from a standard normal using the GAUSS simulation routine, where T denotes the chosen sample size.

For each set of values of the parameters we generated $1,000$ replications of a sample of size T where $T \in \{128, 256, 512, 1024, 2048\}$.

Concerning the optimization, we report below the bias and the mean square error of the PMLE of the parameter d, setting all the other parameters equal to the true values used to generate the data in question.

The PMLE is obtained as the value that minimizes the objective function $\hat{Q}_T(d)$ given in (3.19), the discrete version of the Whittle function. The optimization is made of two parts. The first based on a double grid search over the interval $[\epsilon, 1/2 - \epsilon]$ for some small and positive ϵ. The second is based on 10 iterations of the GAUSS subroutine OPTMUM with the Polak-Ribiere-type option, starting from the estimates obtained by the grid search.

By the nature of the model, the implementation of PML estimation is very cumbersome from a computational point of view even if considerable gains can be achieved via the use of the FFT (cf. section 3.2.2). This explains the relatively small size of the Montecarlo experiments here presented.

Table 5.1: **PMLE : small sample properties** ($\rho = 0$)

d_0	0.1	0.2	0.375	0.45	0.49
sample size : 128					
$BIAS$	0.0809	- 0.0029	-0.1534	-0.1963	-0.1900
MSE	0.0240	0.0246	0.0215	0.0216	0.022
sample size : 256					
$BIAS$	0.0569	- 0.0233	-0.1557	-0.1917	-0.1776
MSE	0.0189	0.0179	0.0156	0.0144	0.015
sample size : 512					
$BIAS$	0.0417	- 0.0304	-0.1408	-0.1724	-0.1608
MSE	0.0145	0.0146	0.0108	0.0084	0.0096
sample size : 1024					
$BIAS$	0.0184	- 0.0302	-0.1141	-0.1438	-0.1329
MSE	0.0088	0.0103	0.0065	0.0041	0.0047
sample size : 2048					
$BIAS$	0.0088	- 0.0304	-0.0899	-0.1246	-0.1191
MSE	0.0056	0.0074	0.0038	0.0023	0.0025

For each sample size and parameter d value
1000 Montecarlo replications are performed .
The other parameter values equal to $\sigma^2 = 1$, $\rho = 0$.

The results are presented in Table 5.1 for $\rho = 0$, in Table 5.2 for $\rho = 1$ and in Table 5.3 for $\rho = 10$. We have also plotted the finite sample distribution of the PMLE for the 75 cases : from Fig. 5.1 to Fig. 5.5 for $\rho = 0$, from Fig. 5.6 to Fig. 5.10 for $\rho = 1$ and from Fig. 5.11 to Fig. 5.15 for $\rho = 10$.

The first clear pattern is that the asymptotic distribution is a valid approximation the smaller the parameter ρ is. The best approximation is obtained when $\rho = 0$. The second is that when ρ is non zero, small (true) values of d are very difficult to be obtained by estimation with a degree of imprecision increasing with ρ and decreasing with d. In fact the order of magnitude of the variance of the squared process increases as $O(\rho^2)$ independently on the value of d. Finally, as we might expect, given the pseudo nature of our estimator, a quite big sample is needed in order to avoid misleading inference, at least equal to 1024.

Table 5.2: **PMLE : small sample properties ($\rho = 1$)**

d_0	0.1	0.2	0.375	0.45	0.49
sample size : 128					
$BIAS$	0.3659	0.2285	-0.01233	-0.0786	-0.1017
MSE	0.0065	0.0095	0.0075	0.0067	0.0091
sample size : 256					
$BIAS$	0.3575	0.1892	-0.0370	-0.1069	-0.1236
MSE	0.0096	0.0121	0.0045	0.0045	0.0081
sample size : 512					
$BIAS$	0.3468	0.1632	-0.0439	-0.1125	-0.1276
MSE	0.0136	0.0120	0.0029	0.0032	0.0072
sample size : 1024					
$BIAS$	0.3074	0.1313	-0.0405	-0.1122	-0.1208
MSE	0.0246	0.0132	0.0024	0.0024	0.0055
sample size : 2048					
$BIAS$	0.2625	0.0937	-0.0391	-0.1009	-0.1098
MSE	0.0334	0.0131	0.0016	0.0018	0.0043

For each sample size and parameter d value

1000 Montecarlo replications are performed .

The other parameter values equal to $\sigma^2 = 1$, $\rho = 1$.

Table 5.3: **PMLE : small sample properties** ($\rho = 10$)

d_0	0.1	0.2	0.375	0.45	0.49
sample size : 128					
BIAS	0.3343	0.2252	0.0628	-0.0139	-0.0667
MSE	0.0071	0.0090	0.0070	0.0063	0.0062
sample size : 256					
BIAS	0.3235	0.2183	0.0486	-0.0345	-0.0739
MSE	0.0126	0.0106	0.0106	0.0099	0.0092
sample size : 512					
BIAS	0.2787	0.1979	0.0273	-0.0397	-0.0927
MSE	0.0230	0.0209	0.0165	0.0129	0.0136
sample size : 1024					
BIAS	0.2989	0.1531	0.0092	-0.0539	-0.0834
MSE	0.0241	0.02705	0.0208	0.0159	0.0114
sample size : 2048					
BIAS	0.2613	0.1698	0.0109	-0.0521	-0.0748
MSE	0.0348	0.0309	0.0221	0.0151	0.0101

For each sample size and parameter d value

1000 Montecarlo replications are performed .

The other parameter values equal to $\sigma^2 = 1$, $\rho = 10$.

Figure 5.1: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 128 . Parameter values : $\sigma^2 = 1$, $\rho = 0$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

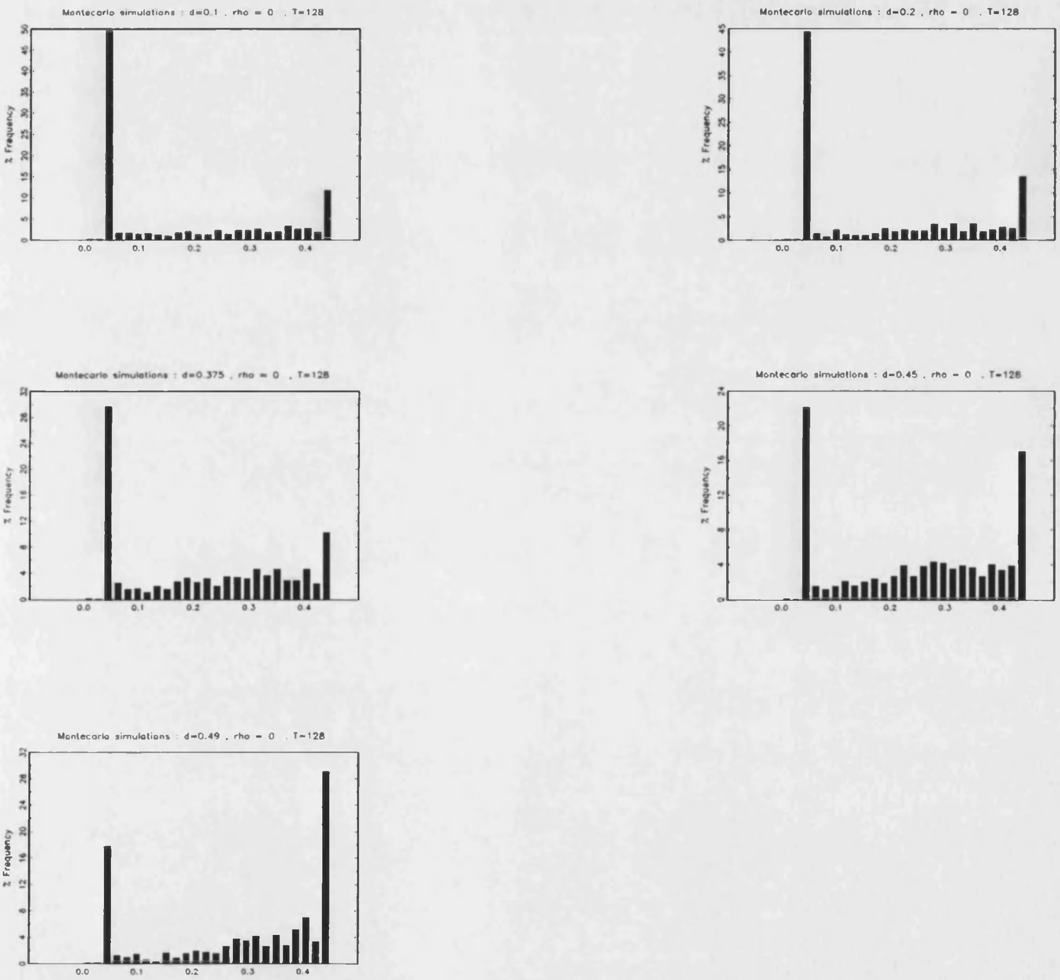


Figure 5.2: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 256 . Parameter values : $\sigma^2 = 1$, $\rho = 0$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

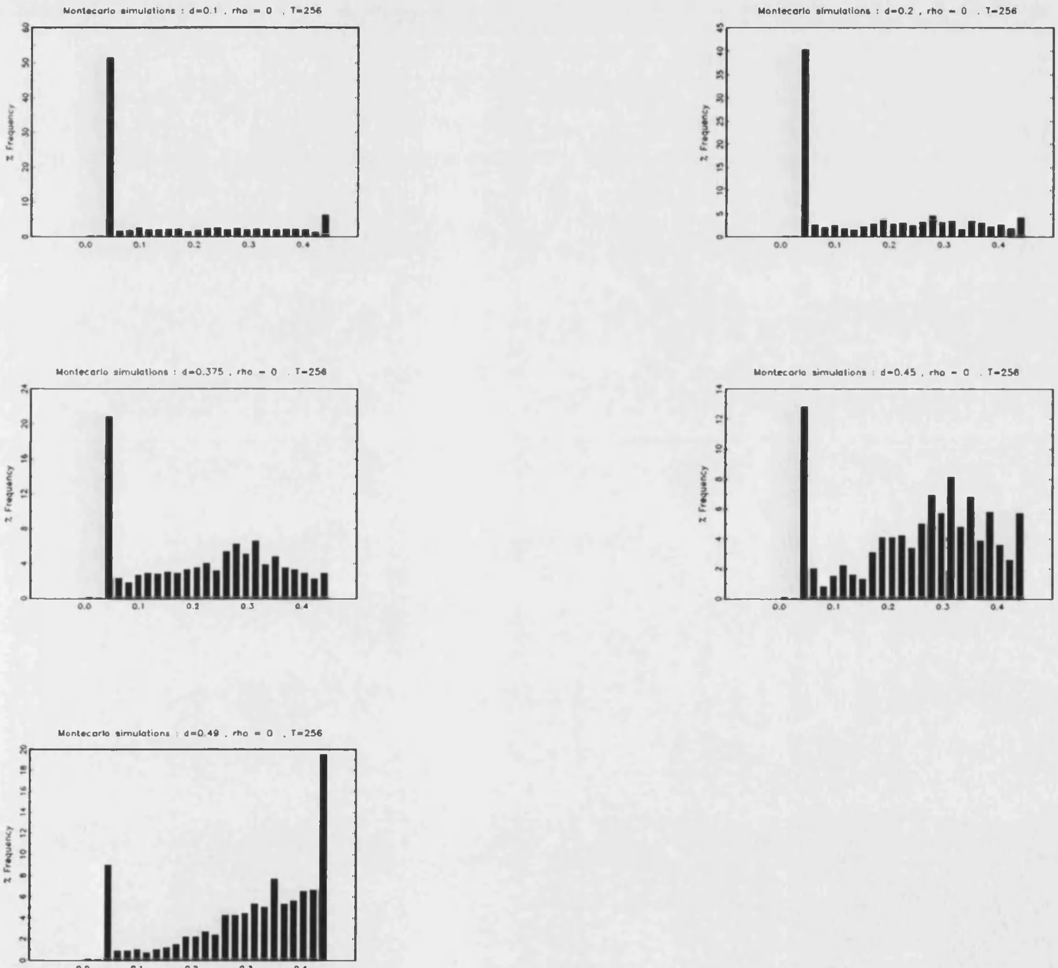


Figure 5.3: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 512 . Parameter values : $\sigma^2 = 1$, $\rho = 0$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

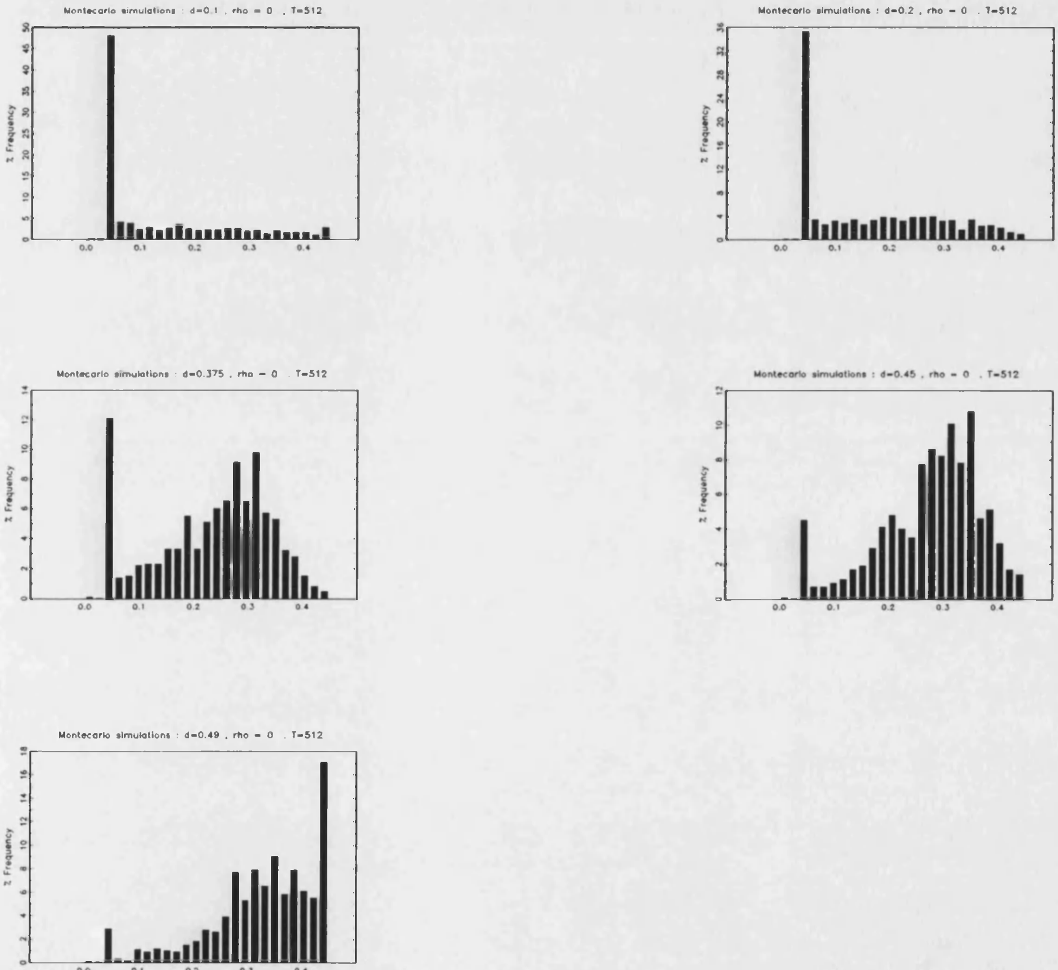


Figure 5.4: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 1024 . Parameter values : $\sigma^2 = 1$, $\rho = 0$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

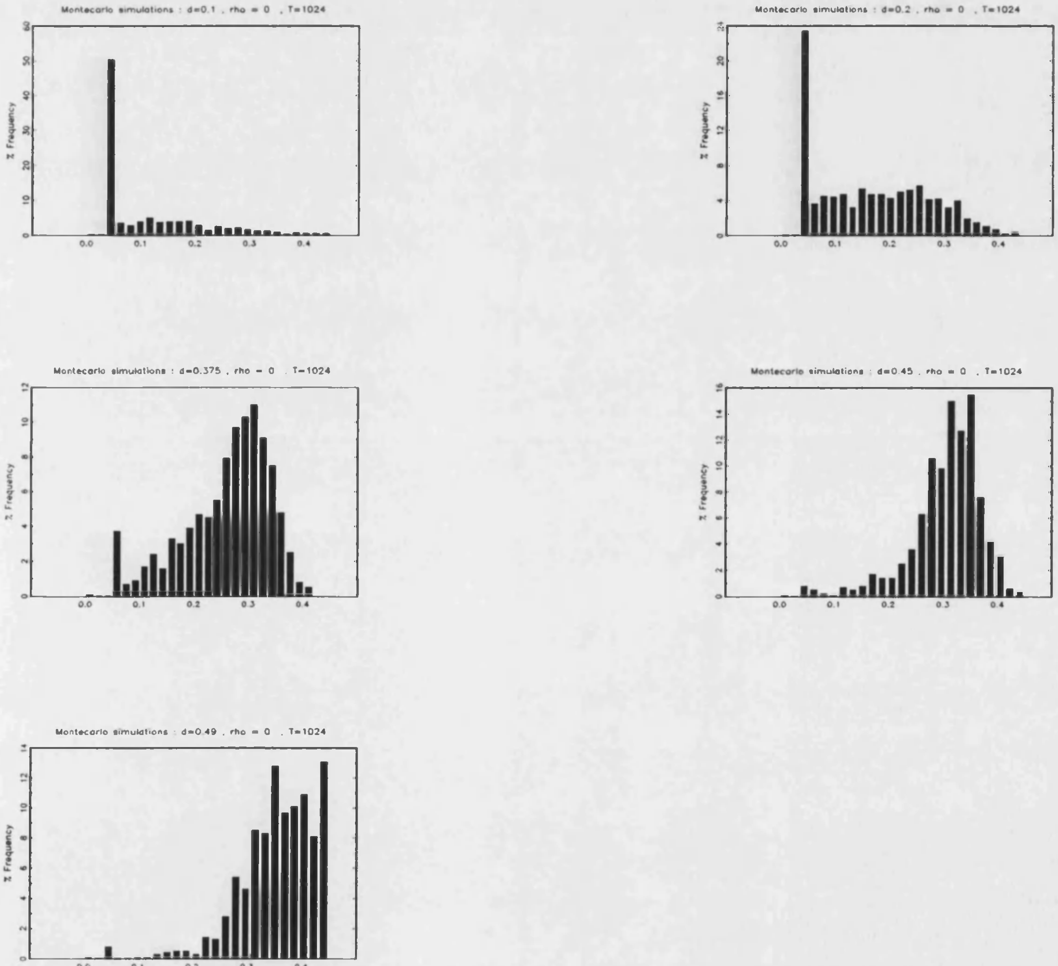


Figure 5.5: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 2048 . Parameter values : $\sigma^2 = 1$, $\rho = 0$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

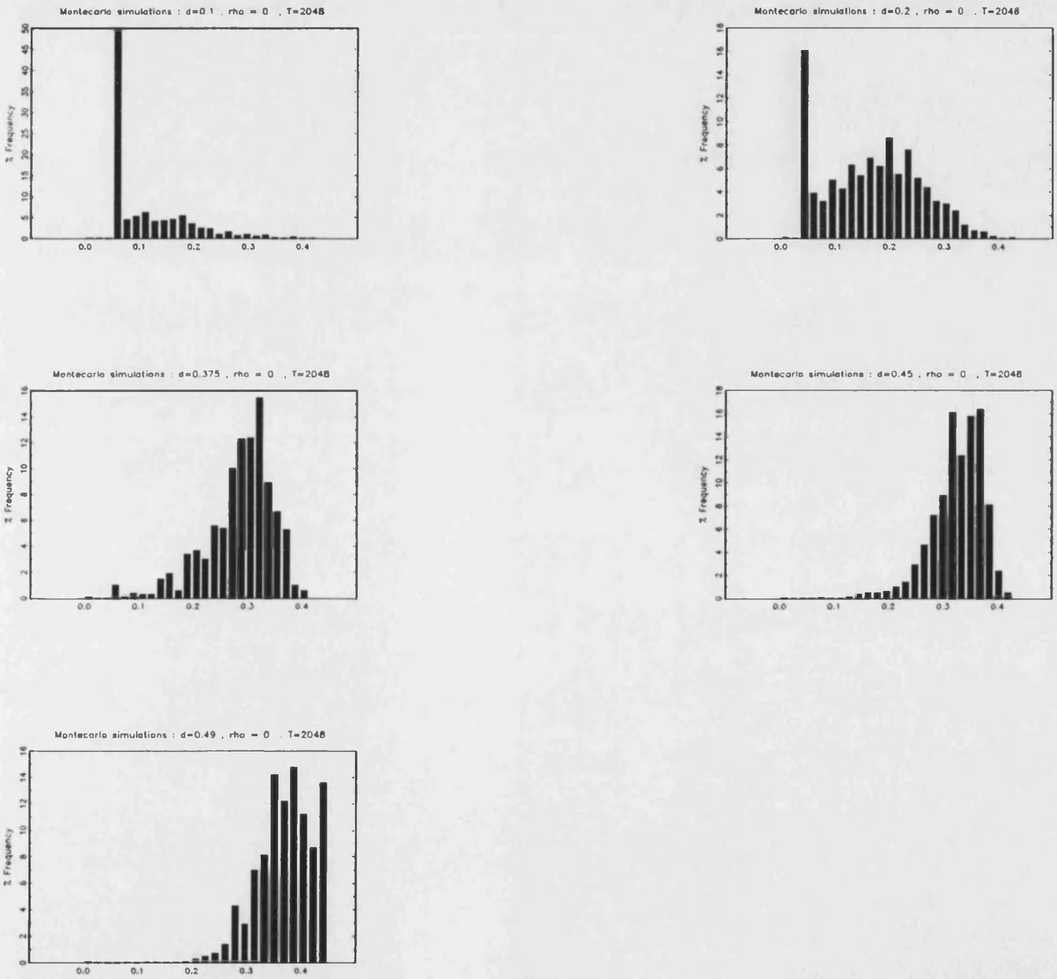


Figure 5.6: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 128 . Parameter values : $\sigma^2 = 1$, $\rho = 1$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

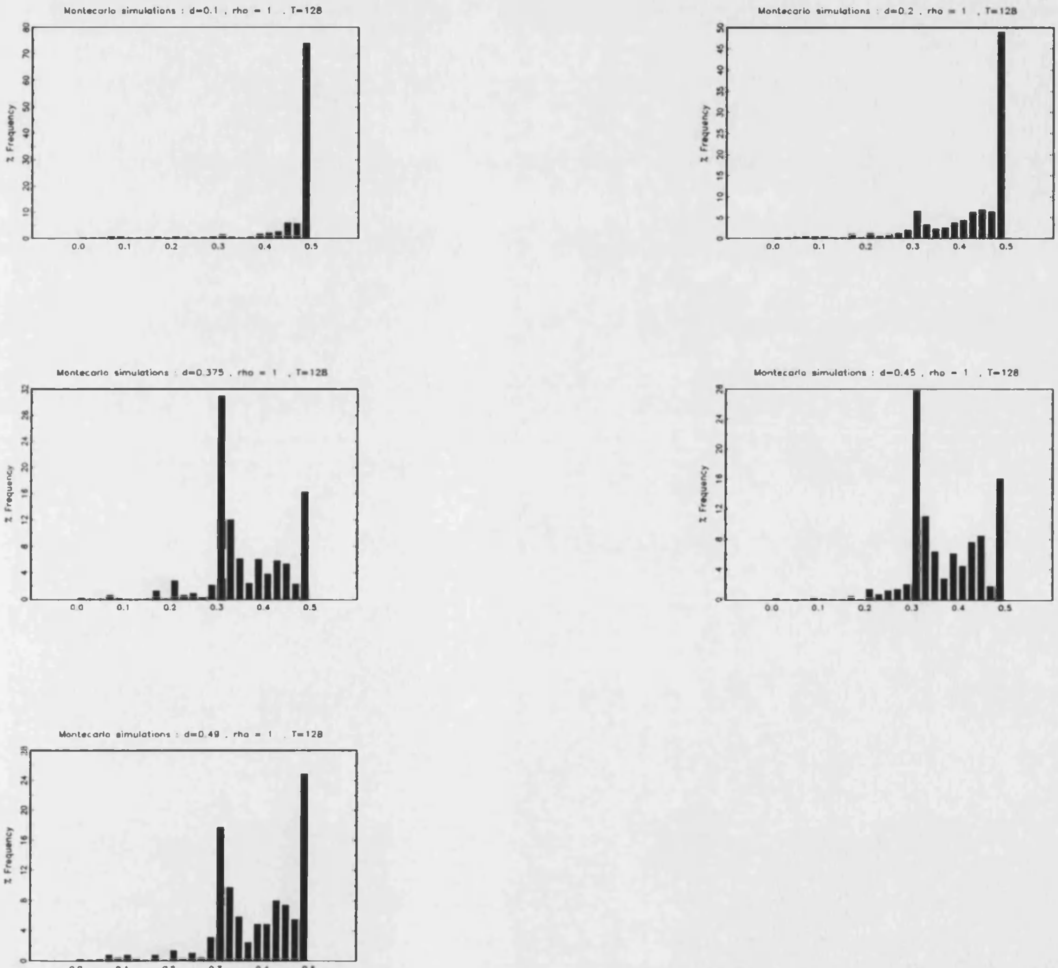


Figure 5.7: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 256. Parameter values : $\sigma^2 = 1$, $\rho = 1$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

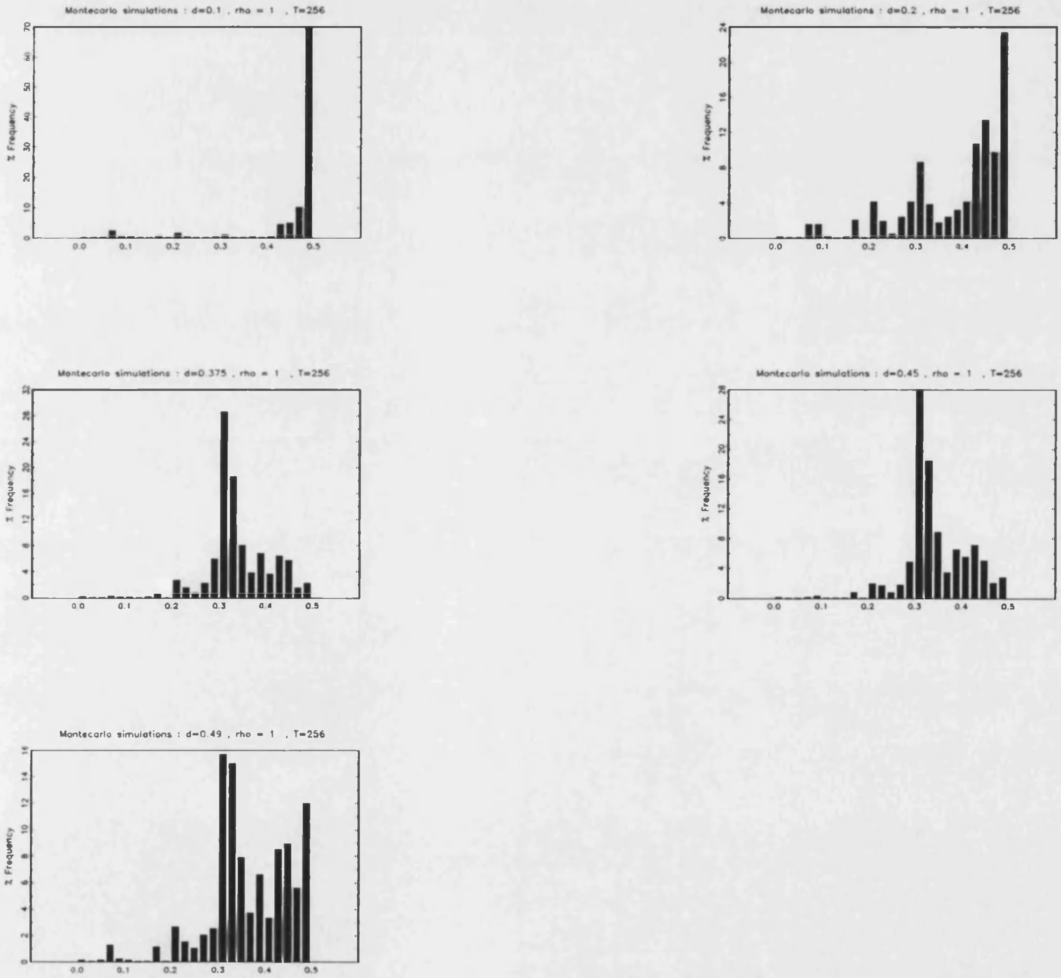


Figure 5.8: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 512 . Parameter values : $\sigma^2 = 1$, $\rho = 1$. The parameter d takes in turn the values {0.1, 0.2, 0.375, 0.45, 0.49} starting from the graph on the left-up corner and moving by row.**

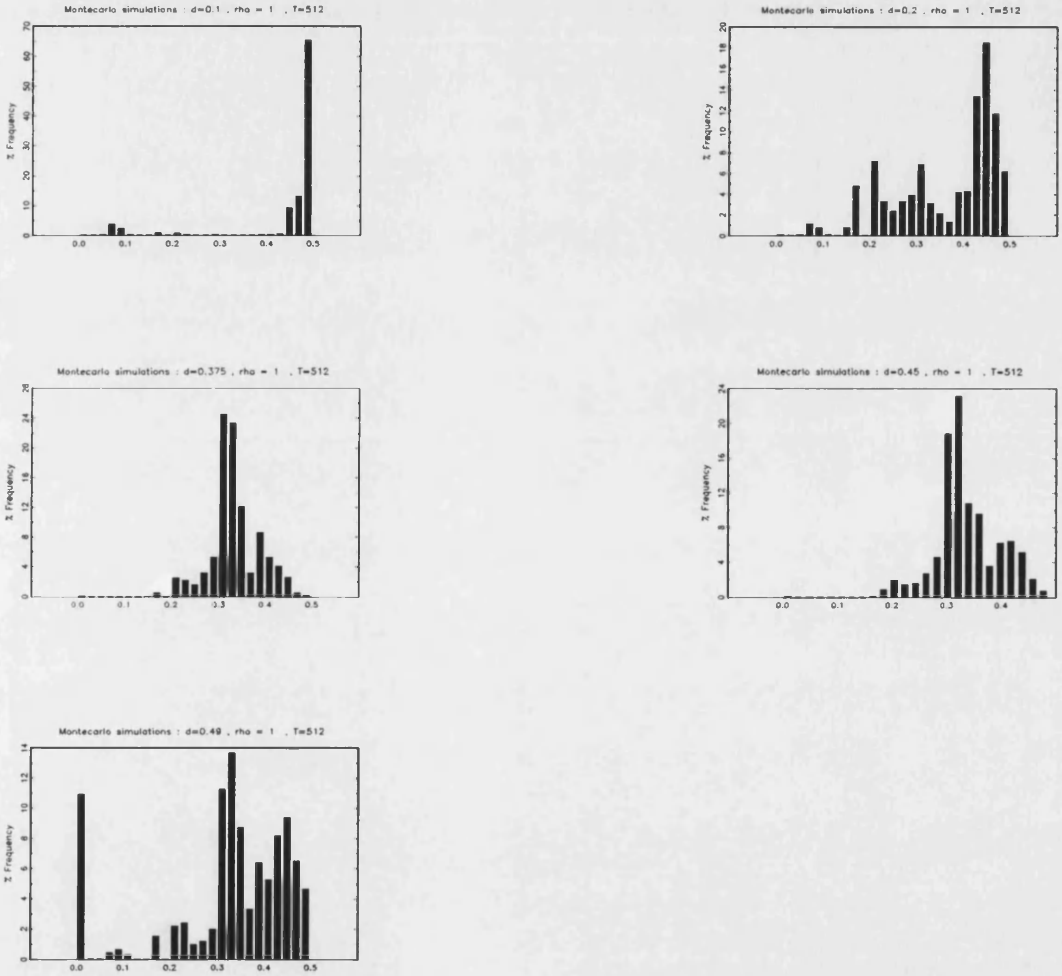


Figure 5.9: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 1024 . Parameter values : $\sigma^2 = 1$, $\rho = 1$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

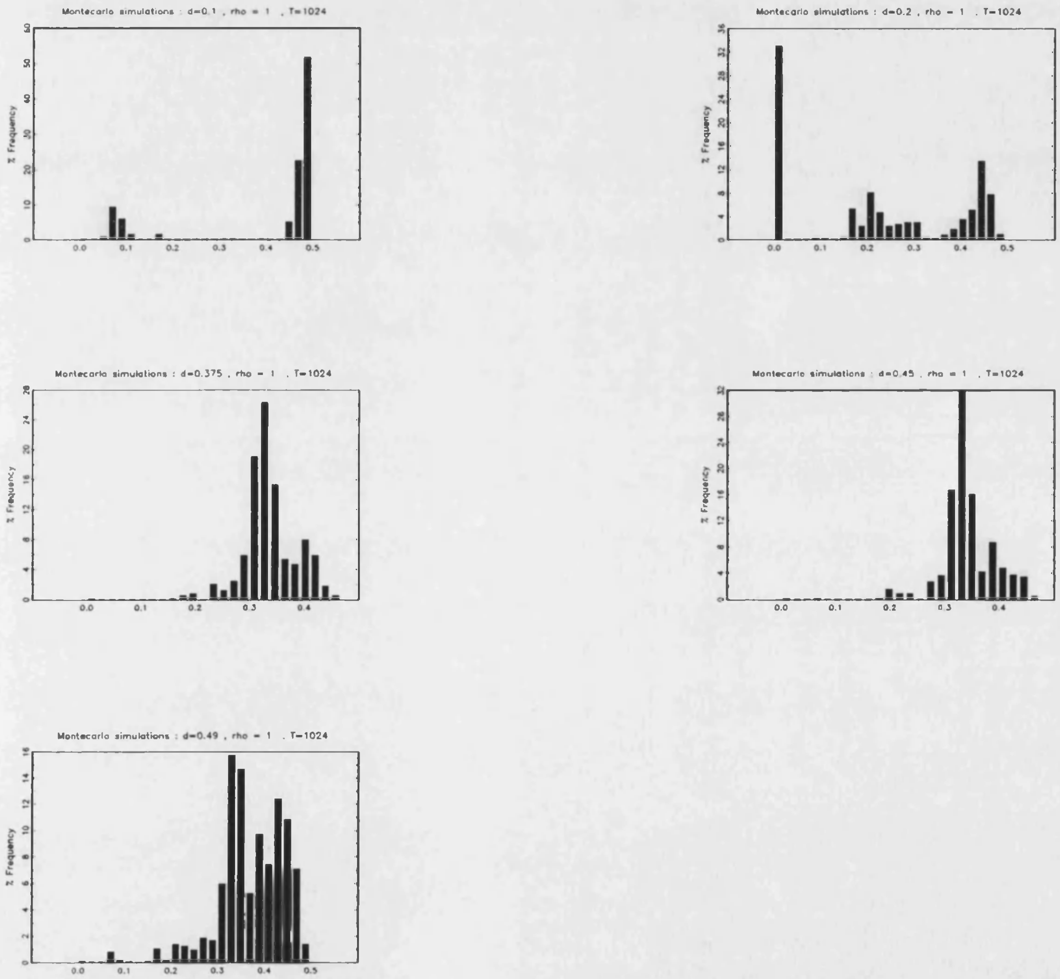


Figure 5.10: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 2048 . Parameter values : $\sigma^2 = 1$, $\rho = 1$. The parameter d takes in turn the values {0.1, 0.2, 0.375, 0.45, 0.49} starting from the graph on the left-up corner and moving by row.**

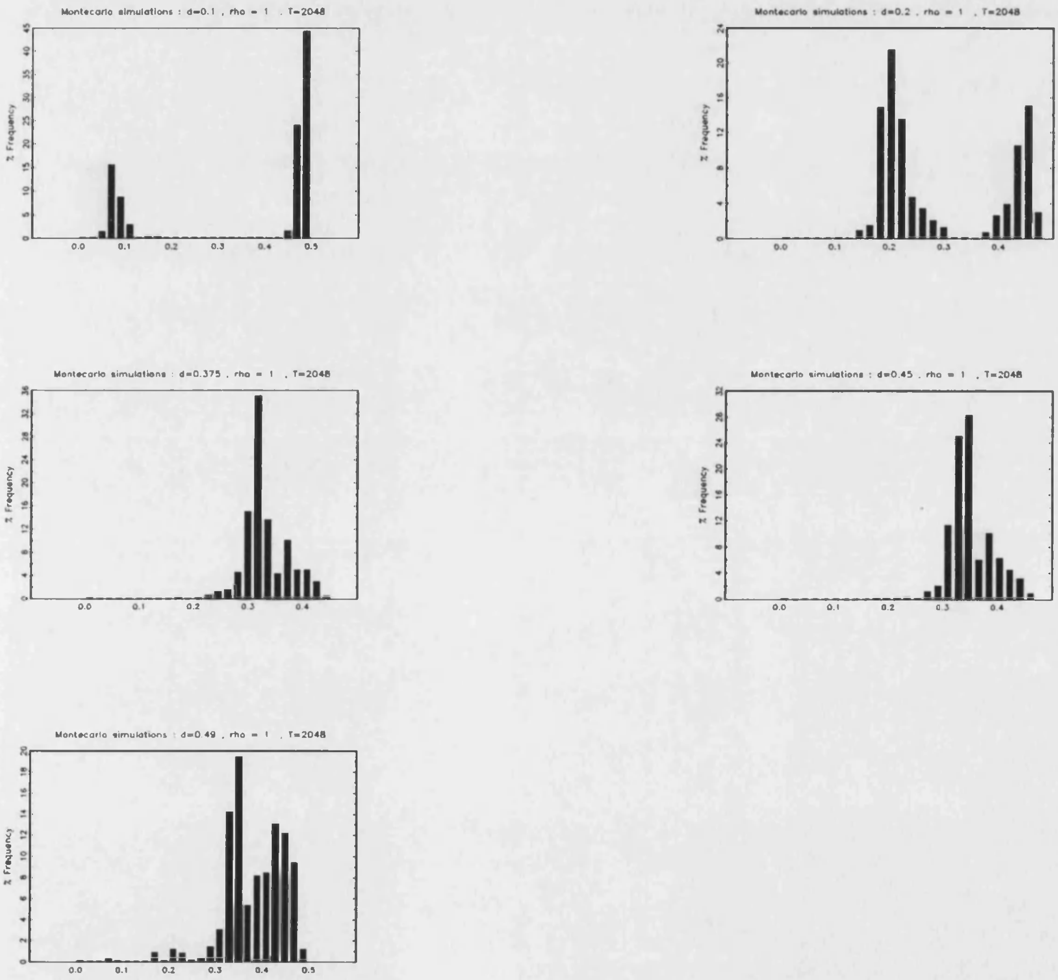


Figure 5.11: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 128 . Parameter values : $\sigma^2 = 1$, $\rho = 10$. The parameter d takes in turn the values {0.1, 0.2, 0.375, 0.45, 0.49} starting from the graph on the left-up corner and moving by row.**

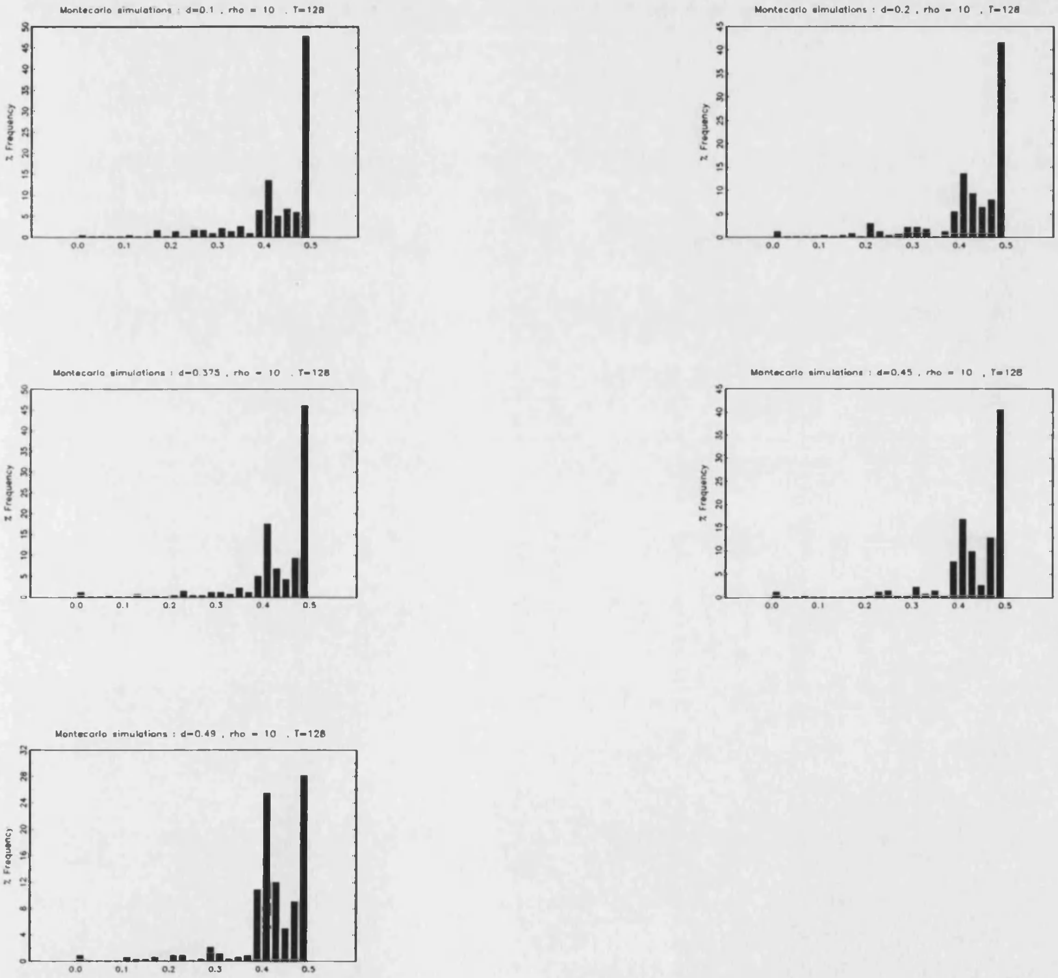


Figure 5.12: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 256. Parameter values : $\sigma^2 = 1$, $\rho = 10$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

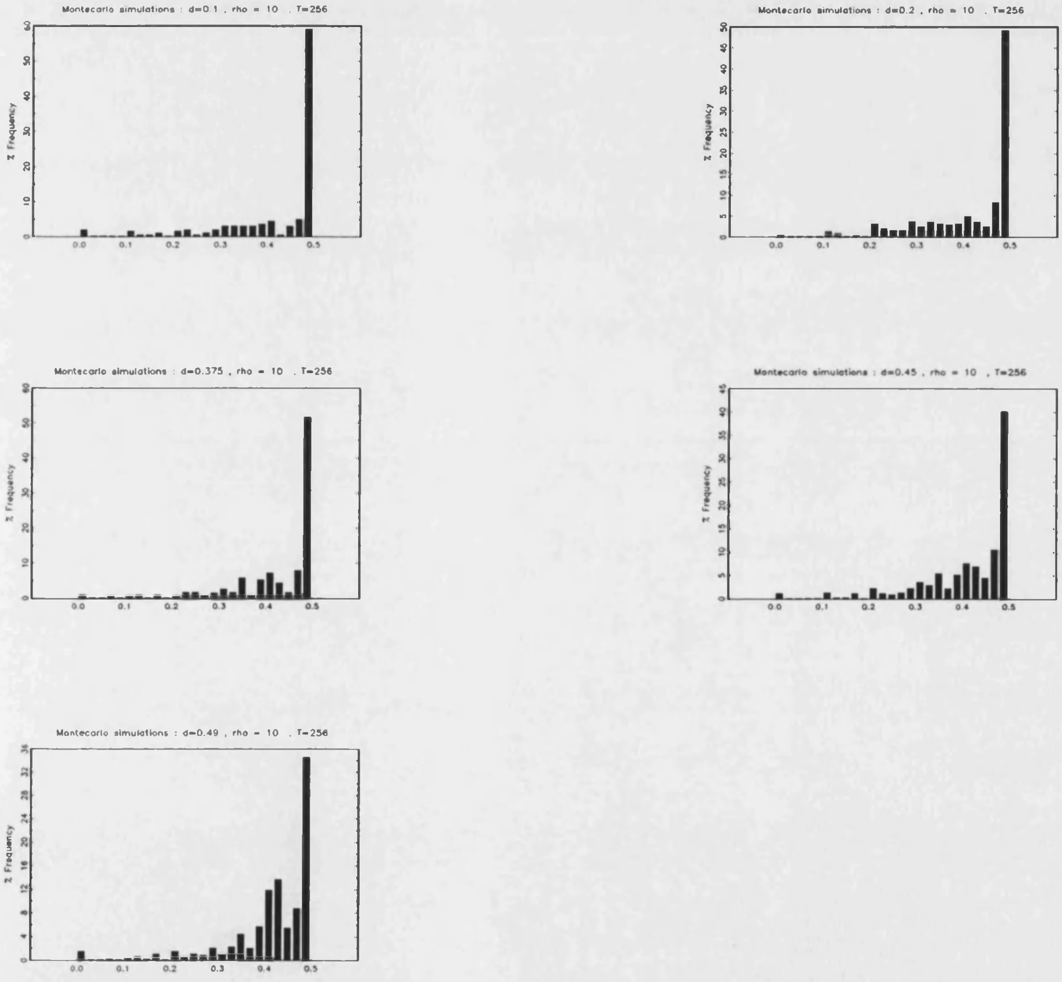


Figure 5.13: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 512 . Parameter values : $\sigma^2 = 1$, $\rho = 10$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**

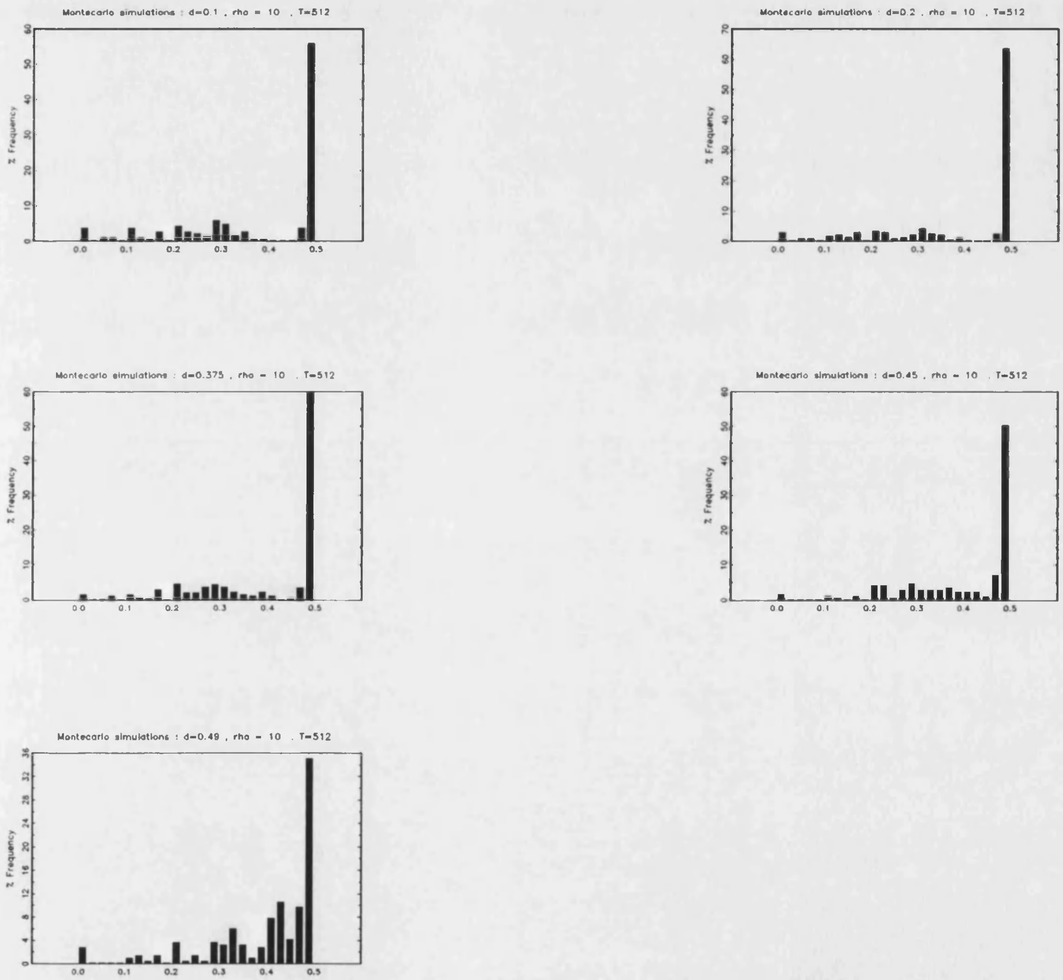


Figure 5.14: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 1024 . Parameter values : $\sigma^2 = 1$, $\rho = 10$. The parameter d takes in turn the values {0.1, 0.2, 0.375, 0.45, 0.49} starting from the graph on the left-up corner and moving by row.**

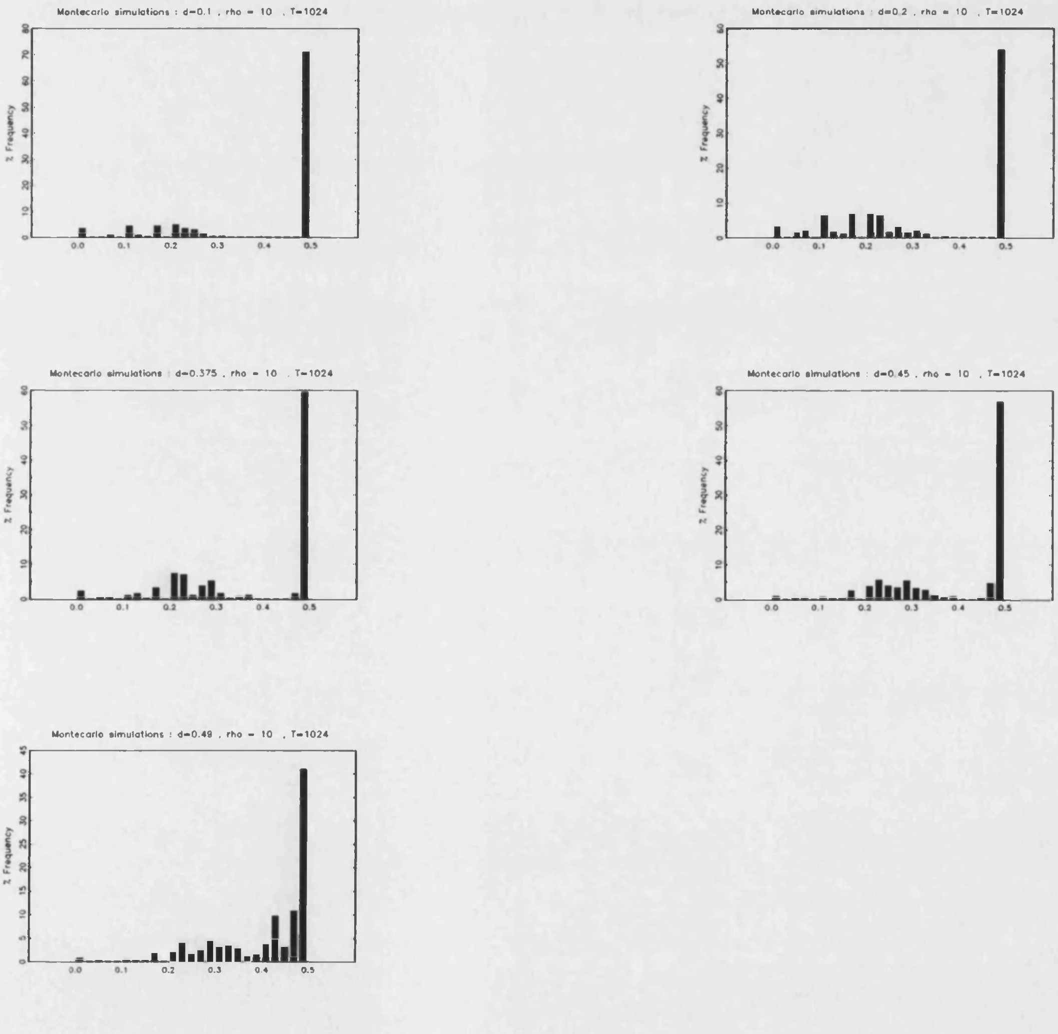
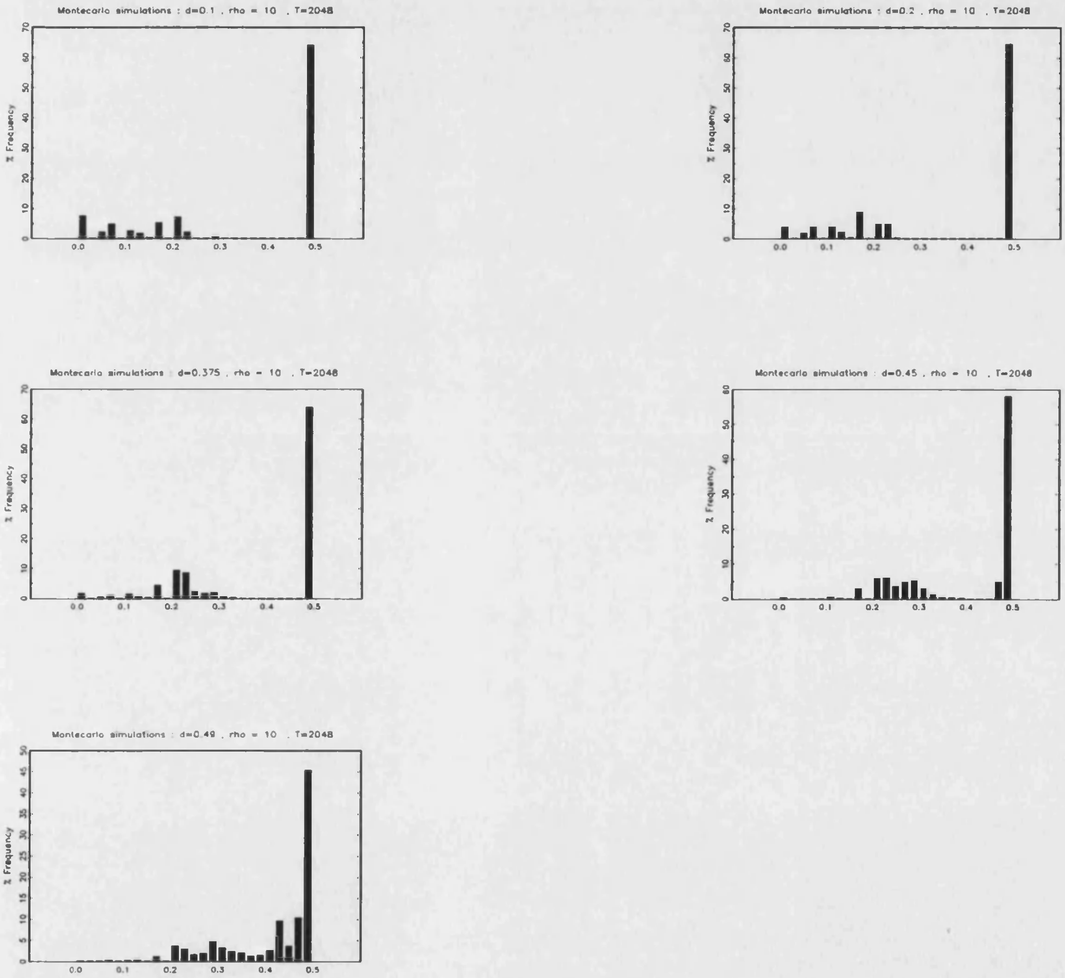


Figure 5.15: **Montecarlo simulations with 1000 replications. Each graph represents the small sample distribution, in terms of histograms, of the PMLE of the parameter d. Sample size = 2048 . Parameter values : $\sigma^2 = 1$, $\rho = 10$. The parameter d takes in turn the values $\{0.1, 0.2, 0.375, 0.45, 0.49\}$ starting from the graph on the left-up corner and moving by row.**



5.2 Empirical applications of the PMLE

We consider the data introduced in section 1.1.1. We choose the following parsimonious parameterization

$$\alpha_i(d) = \prod_{j=1}^{i} \frac{j+d-1}{j},\ 0 < d < 1/2,\ i = 1, 2, \dots,$$

so that

$$\psi = (\bar{\mu}, \bar{\rho}^2, d)'.$$

This means that $h_t - \rho$ has an ARFIMA(0,d,0) representation.

To optimize the pseudo likelihood for y_t we used the GAUSS subroutine OPTMUM with the Polak-Ribiere-type option, with 50 iterations from estimates obtained by a grid search. Standard errors and thus Student-t statistics use the estimates of the trispectrum for the squared data of Taniguchi (1982) and Keenan (1987) with a Fejer window.

The results are displayed in Table 5.4, the hatted quantities indicating parameter estimates with t-ratios in parentheses.

For each raw series the estimates of the normalized mean $\bar{\mu}$ are significantly different from zero for all but the $Yen/Dollar$ series. Things are much more interesting once we consider the estimates of the parameters of the nonlinear part of the model. In fact all the data display a strong degree of dependence in the squares, some of the d values being close to the boundary of the stationary region.

For all but the $Yen/Pound\,spot$ return, $\bar{\rho}^2$ is not significantly different from zero, so in view of Theorem 6, taking e as the memory parameter of the squares we have $\hat{e} = 2\hat{d} - 1/2$, from the relation $2e - 1 = 4d - 2$. For this reason we report the t-ratio for the null $H_0 : d = 1/4$. We see that in all the cases the test is significant against the alternative $H_1 : d > 1/4$. Finally, we observe how the biggest estimates of the $\bar{\rho}^2$ parameter characterize the series with the smallest coefficient of kurtosis, in agreement with the theoretical result of Theorem 7[1]

[1]Note that for the summary statistics in Table 1.1 only, we skipped the week commencing on Monday, 17^{th} October 1987. Otherwise the kurtosis figure for the return indexes were much larger.

Table 5.4: **PMLE : empirical applications**

Data	$\hat{\mu}$ ($t_{\hat{\mu}}$)	$\hat{\rho}^2$ ($t_{\hat{\rho}^2}$)	$\hat{d}$ (t_d) ($t_{d-1/4}$)	$\hat{\sigma}^2$	*loglikelihood*
sYP	1.433 (1.540)	9.027 (2.502)	0.465 (47.551) (22.010)	0.0018	-19.507
fYP	1.398 (1.052)	13.458 (1.541)	0.465 (16.991) (7.864)	0.0017	-19.366
sUP	-2.783 (-5.847)	13.474 (1.673)	0.465 (14.569) (6.744)	0.0014	-19.759
fUP	-2.760 (-5.864)	13.462 (1.671)	0.465 (14.546) (6.733)	0.0014	-19.687
F100	3.505 (16.419)	0.049 (0.292)	0.366 (24.840) (7.857)	0.0074	-16.804
FAll	4.703 (11.350)	2.718 (0.965)	0.429 (5.158) (2.152)	0.0037	-17.111
S500	4.072 (12.938)	0.0497 (0.182)	0.401 (18.589) (7.003)	0.0096	-15.016

The data refers to the period 1^{st} Jan 1986 to 1^{st} Nov 1993 (3088 observations).

Each columns reports the estimate of the parameters ($\bar{\mu}$, $\bar{\rho}^2$, d)

with the standard Student-t in parentheses.

For d we report the Student-t (third figures in column 4) for the null hypothesis $H_0 : d = 1/4$.

All the calculations of this table are based on the fast Fourier transform.

Chapter 6

The 'one-shock' model: Lagrange Multipliers tests

We develop testing procedures for dynamic conditional heteroskedasticity with good power against long memory alternatives based on the Lagrange multiplier principle. We consider the behaviour of the test statistic under the null under two different set of assumptions, hence allowing for a great deal of heterogeneity in the underlying process. We propose frequency domain expressions for the test statistics. Finally the small sample properties (size and power) of the LM test statistics here introduced are assessed with some Montecarlo experiments. An empirical application concludes.

6.1 Introduction

In Chapter 3 we proposed a Gaussian PMLE for our model and in Chapter 4 we showed that the PMLE is asymptotically normal. This allows to perform standard inference, after the model has been estimated, based on the Wald test procedure. Alternatively, under a given null hypothesis, one can obtain the constrained PMLE and thus performing a Likelihood Ratio (LR) test.

From a practical point of view both procedures entail solving a highly nonlinear optimization problem by some numerical method. Hence before carrying full (and constrained in the LR test case) estimation of the model with empirical data it seems important to develop a preliminary test for conditional heteroskedasticity which would eventually justify the full estimation in a second stage.

6.2 The Lagrange Multiplier test

For this reason in this chapter we develop a test for dynamic conditional heteroskedasticity based on the Lagrange Multiplier (LM) principle with power against a weak-memory as well a long-memory alternative. ARCH tests (Engle 1982) (Weiss 1986) where the alternative is build an autoregression (moving average) over the past p (q) lags for some finite integer p (q), for the squared disturbances are likely to be consistent but inefficient against long memory heteroskedasticity (Robinson 1991a). Of course the gain in power with respect to standard ARCH tests comes at the cost of loosing the simple linear regression TR^2 formulation characterizing these tests.

We will develop two different versions of the tests, the second one being valid under weaker conditions than the first one, in particular allowing for some degree of heterogeneity of the unobservable ϵ_t in terms of the conditional sixth and higher moments. We also present an asymptotically equivalent test statistic expressed in the frequency domain. The methodology used follows Robinson (1991a).

6.3 The LM test: definitions and assumptions

The following sets of assumptions will completely replace Assumptions A', B'' both with respect to the coefficients $\alpha_i(\theta)$ and the unobservable process $\{\epsilon_t\}$.

Assumptions C

C_1 *The coefficients* $\{\alpha_i(\theta); i = 1, 2, \ldots, \}$ *are some invertible function of a* $p \times 1$ *vector* θ *such that* $\alpha_i(\theta) = 0$, $\forall i$ *if and only if* $\theta = 0$. *The partial derivatives of* $\alpha_i(\theta)$ *are square* *summable for $\theta = 0$ where we set $\tau_i(\theta) = \frac{\partial}{\partial\theta}\alpha_i(\theta)$ with $\tau_i = \tau_i(0)$.*

C_2 *The matrix $\Gamma = \sum_{i=1}^{\infty} \tau_i \tau_i'$ is invertible.*

Assumption D *The process $\{\epsilon_t\}$ satisfies*

$$E(\epsilon_t^a|\mathcal{F}_{t-1}) = \begin{cases} 0, & a = 1, \\ \sigma^2, & a = 2, \\ 3\sigma^4, & a = 4, \\ 15\sigma^6, & a = 6, \\ 105\sigma^8, & a = 8. \end{cases}$$

Remarks: (1) Note how nowhere have we assumed the ϵ_t to be an i.i.d. sequence, differently from Assumption B.
(2) On the other hand the process $\{\epsilon_t\}$ behaves as an independent identically Gaussian distributed sequence up to the eighth moment.

Assumption E

$$\rho^2 = 1. \tag{6.1}$$

We assume that the model is parameterized such that the null hypothesis is stated as

$$H_0 : \theta = 0.$$

Remarks: (1) under H_0 the observable process coincides with the unobservable process apart from the constant mean, viz. $x_t = \mu + \epsilon_t$, and so behaving as an independent and identically distributed Gaussian process up to the eighth order, in particular displaying conditional homoskedasticity.
(2) Assumption E is needed because under H_0 we are able to estimate the parameter $\phi = \sigma\rho$

only (beside μ). This will not affect the power of the test because the autocorrelation structure of the squared process, for θ different from zero, will depend entirely on whether ρ is zero or not.

(3) Under the alternative the squared process, certainly autocorrelated, may display long-memory depending on the parameterization chosen as in Assumption A_1 of Chapter 2 (cf. Robinson and Zaffaroni (1996a)).

Now by Assumption D, it follows

$$E_{t-1}(y_t) = \mu^2 + \sigma^2 h_{t-1}^2,$$
$$var_{t-1}(y_t) = 2\left(\sigma^4 h_{t-1}^4 + 2\sigma^2 \mu^2 h_{t-1}^2\right).$$

where $E_{t-1}(.)$ denotes the expectations conditional of the sigma-algebra $\mathcal{F}_{t-1}$ generated by $\{\epsilon_s, s \leq t-1\}$ and var_{t-1} the conditional variance.

Let us define the 'normalized' parameter

$$w = \mu^2/\sigma^2.$$

We will need to reparametrize the model with respect to the vector of parameters $(w, \theta')'$ so that replacing the α_i by the $\alpha_i(\theta)$ in h_t yields $h_t(\theta)$. We will need to evaluate the likelihood and scores under the null so that the issue of invertibility becomes irrelevant in this context and there is no need to introduce the $\bar{h}_t$.

The 'time domain' pseudo Gaussian likelihood is then

$$expL(\sigma^2, w, \theta) = \frac{1}{\left[(2\pi)^T \prod_{t=1}^T 2\left(\sigma^4 h_{t-1}^4(\theta) + 2\sigma^2 \mu^2 h_{t-1}^2(\theta)\right)\right]^{1/2}} \tag{6.2}$$

$$\times exp\left(-\frac{\sum_{t=1}^T (y_t - \mu^2 - \sigma^2 h_{t-1}^2(\theta))^2}{4(\sigma^4 h_{t-1}^4(\theta) + 2\sigma^2\mu^2 h_{t-1}^2(\theta))}\right), \tag{6.3}$$

so taking logs, concentrating σ^2 out and rearranging terms we get

$$L(\sigma^2, w, \theta) = -T/2\,ln(4\pi) - T\,ln(\sigma^2) - 1/2\,ln\left(\prod_{t=1}^T 2(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta))\right)$$
$$-1/4\sum_{t=1}^T (\frac{y_t}{\sigma^2} - w - h_{t-1}^2(\theta))^2/(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta)).$$

Let us define

$$\beta_t^*(\theta) = \sum_{i=1}^{\infty} \tau_i(\theta)\epsilon_{t+1-i}\,,$$
$$\beta_t(\theta) = \sum_{i=1}^{t} \tau_i(\theta)\epsilon_{t+1-i}\,.$$

so that $\beta_t(\theta)$ represents the part of $\beta_t^*(\theta)$ which belongs to $\mathcal{F}_1^T$, the sigma-field generated by $\{\epsilon_1, \ldots, \epsilon_T\}$, for every t such that $1 \le t \le T$. Again $\beta_t = \beta_t(0)$.

Differentiating with respect to the parameters (σ^2, w, θ) we obtain

$$\frac{\partial L^*(\sigma^2, w, \theta)}{\partial \sigma^2} = -T/\sigma^2 + 1/2\sigma^2 \sum_{t=1}^{T} \frac{(y_t/\sigma^2 - w - h_{t-1}^2(\theta))y_t/\sigma^2}{(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta))}\,, \tag{6.4}$$

$$\frac{\partial L^*(\sigma^2, w, \theta)}{\partial w} = \tag{6.5}$$

$$= 1/2 \sum_{t=1}^{T} \frac{1}{(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta))} \times \tag{6.6}$$

$$\left[(y_t/\sigma^2 - w - h_{t-1}^2(\theta))\left(\frac{(y_t/\sigma^2 - w - h_{t-1}^2(\theta))}{(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2)(\theta)} + 1\right) - 2\right], \tag{6.7}$$

$$\frac{\partial L^*(\sigma^2, w, \theta)}{\partial \theta} = \tag{6.8}$$

$$= \sum_{t=1}^{T} \frac{h_{t-1}(\theta)\,\beta_{t-1}^*(\theta)}{(h_{t-1}^4(\theta) + 2\,w\,h_{t-1}^2(\theta))}\Big((y_t/\sigma^2 - w - h_{t-1}^2(\theta)) + \tag{6.9}$$

$$(y_t/\sigma^2 - w - h_{t-1}^2(\theta))^2 \frac{(h_{t-1}^2(\theta) + w)}{(h_{t-1}^4(\theta) + 2wh_{t-1}^2(\theta))} - 2\,(h_{t-1}^2(\theta) + w)\Big)\,,$$

Let us consider the following definitions where m defines an integer smaller than T .

$$X_t = \left((\frac{y_t}{\sigma^2} - w - 1)((\frac{y_t}{\sigma^2} - w - 1)\frac{(1+w)}{(1+2w)} + 1) \;-\; 2(1+w)\right),$$

$$\eta_i = \sum_{t=1+i}^{T} \epsilon_{t-i}\,X_t\,,$$

$$\Gamma_m = \sum_{i=1}^{m} \tau_i \tau_i'\,,$$

$$
\begin{aligned}
e &= E(X_t^2) = 9 + \frac{(65 + 150w^3 + 338w^2 + 280w)}{(1+2w)^2}, \\
\lambda_m &= \frac{1}{(T\, e\, \sigma^2)^{1/2}} \Gamma_m^{1/2} \sum_{i=1}^{m} \tau_i \eta_i, \\
\hat{\sigma}^2 &= 1/T \sum_{t=1}^{T} (x_t - \bar{x})^2, \\
\hat{w} &= \bar{x}^2 / \hat{\sigma}^2.
\end{aligned}
$$

Whenever any of the above expressions becomes a function of $(\hat{\sigma}^2, \hat{w})$ we will write this explicitly as

$$X_t(\hat{\sigma}^2, \hat{w}) = \left((\frac{y_t}{\hat{\sigma}^2} - \hat{w} - 1)((\frac{y_t}{\hat{\sigma}^2} - \hat{w} - 1)\frac{(1+\hat{w})}{(1+2\hat{w})} + 1) \; - \; 2(1+\hat{w}) \right).$$

Then our LM test statistic is given by

$$LM = \lambda' \lambda, \tag{6.10}$$

where we set

$$\lambda = \lambda_{T-1}(\hat{\sigma}^2, \hat{w}).$$

Remark: we can use the matrix Γ directly in the expression for LM if the former has a closed form expression, depending on the parameterization chosen.

In the next section we will justify the expression of the LM statistic deriving its asymptotic distribution under H_0.

6.4 The LM test: asymptotic distribution under the null

We consider a set of preliminary lemmas with the main theorem reported at the end of this section. This theorem establishes the asymptotic distribution of the LM test under H_0.

The next lemma allows us to take the summation in the index t as ranging over $1 \le t \le T$.

Lemma 23 *Under Assumptions C, D, E and H_0*

$$\lambda_m - \bar{\lambda}_m = o_p(1)$$

with

$$\bar{\lambda}_m = \frac{1}{(T\,e\,\sigma^2)^{1/2}}\Gamma^{-1/2}\sum_{i=1}^{m}\tau_i\bar{\eta}_i\,, \tag{6.11}$$

$$\bar{\eta}_i = \sum_{t=1}^{T}\epsilon_{t-i}\,X_t\,. \tag{6.12}$$

Proof: We need to consider that

$$E(X_t\epsilon_{t-i}|\mathcal{F}_{t-1}) = 0, \text{ for any } i \geq 1\,, \tag{6.13}$$

and thus

$$E \parallel \sum_{i=1}^{m}\tau_i\tilde{\eta}_i \parallel^2 = O(m^2)\,, \tag{6.14}$$

with

$$\tilde{\eta}_i = \bar{\eta}_i - \eta_i = \sum_{t=1}^{i}\epsilon_{t-i}\,X_t\,,$$

□

Lemma 24 *Under H_0 and Assumptions C, D, E for $i = 1, 2, \ldots, p$,*

$$E(\frac{\partial L(\sigma^2, w, 0)}{\partial\sigma^2}\,\frac{\partial L(\sigma^2, w, 0)}{\partial\theta_i}) = 0\,,$$

$$E(\frac{\partial L(\sigma^2, w, 0)}{\partial w}\,\frac{\partial L(\sigma^2, w, 0)}{\partial\theta_i}) = 0\,.$$

Proof: We can express the score under H_0 as

$$\frac{\partial L(\sigma^2, w, 0)}{\partial\theta} = \frac{1}{(1+2w)}\sum_{t=2}^{T}\beta_{t-1}X_t\,, \tag{6.15}$$

$$\frac{\partial L(\sigma^2, w, 0)}{\partial\sigma^2} = -T/\sigma^2 + \frac{1}{2(1+2w)\,\sigma^2}\sum_{t=1}^{T}W_{1,t}\,, \tag{6.16}$$

$$\frac{\partial L(\sigma^2, w, 0)}{\partial w} = \frac{1}{2(1+2w)}\sum_{t=1}^{T}W_{2,t}\,, \tag{6.17}$$

where

$$W_{1,t} = (\frac{y_t}{\sigma^2} - w - 1)\,\frac{y_t}{\sigma^2}\;,,$$

$$W_{2,t} = (\frac{y_t}{\sigma^2} - w - 1)\left(1 + \frac{(y_t/\sigma^2 - w - 1)}{(1+2w)}\right) - 2.$$

From Assumption D, denoting by K_1, K_2 two real constants, it follows that for any $1 \leq t \leq T$

$$\begin{aligned}
&E(X_t \mid \mathcal{F}_{t-1}) = 0\,, \\
&E(X_t W_{1,t} \mid \mathcal{F}_{t-1}) = K_1 \, < \infty\,, \\
&E(X_t W_{2,t} \mid \mathcal{F}_{t-1}) = K_2 \, < \infty\,, \\
&E(\beta_t) = 0\,.
\end{aligned}$$

Thus the result follows by straightforward evaluation of the expectations of the cross-products of (6.15) with (6.16) and (6.15) with (6.17) .

□

Lemma 24 allows us to consider the following quantity as our LM test statistic:

$$\frac{\partial L(\hat{\sigma^2}, \hat{w}, 0)}{\partial \theta'} E\left(\frac{\partial L(\sigma^2, w, 0)}{\partial \theta}\frac{\partial L(\sigma^2, w, 0)}{\partial \theta'}\right)^{-1} |_{\sigma^2=\hat{\sigma^2}, w=\hat{w}} \frac{\partial L(\hat{\sigma^2}, \hat{w}, 0)}{\partial \theta}$$

In fact the covariance matrix of the score is block diagonal with respect to the two subsets of parameters (σ^2, w) on one hand and θ on the other.

In particular, concerning the variances-covariances matrix for the score with respect to θ, we obtain the following:

Lemma 25 *Under H_0 and Assumptions C, D, E,*

$$E\left(\frac{\partial L(\sigma^2, w, 0)}{\partial \theta}\frac{\partial L(\sigma^2, w, 0)}{\partial \theta'}\right) = \frac{e(w)}{(1+2w)^2}\sigma^2 \Gamma_{T-1}\,.$$

Proof: From Assumption D we have that $E(X_t \mid \mathcal{F}_{t-1}) = 0$ so that in the inner product of the score we need to evaluate the term $E(X_t^2 \mid \mathcal{F}_{t-1}) = E(X_t^2) = e$ only where the first equality holds under H_0. In particular, under H_0 we have that

$$\begin{aligned}
y_t &= \mu^2 + \epsilon_t^2 + 2\mu\epsilon_t\,, \\
y_t^2 &= \mu^4 + \epsilon_t^4 + 6\mu^2\epsilon_t^2 + 4\mu^3\epsilon_t + 4\mu\epsilon_t^3\,, \\
y_t^3 &= \mu^6 + \epsilon_t^6 + 20\mu^3\epsilon_t^3 + 15\mu^4\epsilon_t^2 + 15\mu^2\epsilon_t^4 + 6\mu^5\epsilon_t + 6\mu\epsilon_t^5\,,
\end{aligned}$$

$$
\begin{aligned}
y_t^4 \quad &= \mu^8 + \epsilon_t^8 + 70\mu^4\epsilon_t^4 + 28\mu^6\epsilon_t^2 + 28\mu^2\epsilon_t^6 + \\
&+ 8\mu^7\epsilon_t 8\mu\epsilon_t^7 + 56\mu^5\epsilon_t^3 + 56\mu^3\epsilon_t^5 .
\end{aligned}
$$

Thus after taking the expectations we obtain

$$
\begin{aligned}
&E(y_t) = \sigma^2(w+1)\,, \\
&E(y_t^2) = \sigma^4(w^2+3+6w) = \sigma^4\left((w+1)^2 + 2(1+2w)\right)\,, \\
&E(y_t^3) = \sigma^6(w^3+15+15w^2+60w) = \sigma^6\left((w+1)^3 + 14 + 12w^2 + 57w\right)\,, \\
&E(y_t^4) = \sigma^8(w^4+105+210w^2+28w^3+420w) = \\
&= \sigma^8\left((w+1)^4 + 104 + 204w^2 + 24w^3 + 416w\right)\,.
\end{aligned}
$$

Then let us consider the following expression

$$
\begin{aligned}
X_t^2 &= (\frac{y_t}{\sigma^2} - w - 1)^2 + \frac{(1+w)^2}{(1+2w)^2}(\frac{y_t}{\sigma^2} - w - 1)^4 + 4(1+w)^2 \\
&+ 2(\frac{y_t}{\sigma^2} - w - 1)\left(\frac{(1+w)}{(1+2w)}(\frac{y_t}{\sigma^2} - w - 1)^2 - 2(1+w)\right) - \frac{4(1+w)^2}{(1+2w)}(\frac{y_t}{\sigma^2} - w - 1)\,,
\end{aligned}
$$

expanding the squares, the cubes and the fourth power and taking the expectations, after simple but tedious algebra, yields

$$
E(X_t^2 \mid \mathcal{F}_{t-1}) = 9 + \frac{(65 + 150w^3 + 338w^2 + 280w)}{(1+2w)^2} = e\,.
$$

□

Lemma 26 *Under H_0 and Assumptions C, D, E,*

$$
\bar{\lambda}_m \rightarrow_d \mathcal{N}_p(0, I_p).
$$

Proof: We will make use of a martingale CLT (Brown 1971). Let us consider

$$
\sum_{i=1}^{m} \tau_i \bar{\eta}_i. \tag{6.18}
$$

Then, for a $p \times 1$ vector of real constants ν we shall define

$$
S_T = \nu' \sum_{i=1}^{m} \tau_i \sum_{t=1}^{T} X_t \epsilon_{t-i} = \sum_{t=1}^{T} U_t, \tag{6.19}
$$

where we set

$$U_t = (\sum_{i=1}^{m} \nu' \tau_i \epsilon_{t-i}) X_t.$$

Then

$$E(U_t | \mathcal{F}_{t-1}) = 0, \tag{6.20}$$

and

$$\sigma_t^2 = E(U_t^2 | \mathcal{F}_{t-1}) = (\nu' \sum_{i=1}^{m} \tau_i \epsilon_{t-i})^2 \, e, \tag{6.21}$$

$$V_T^2 = \sum_{t=1}^{T} \sigma_t^2 = e \sum_{t=1}^{T} (\nu' \sum_{i=1}^{m} \tau_i \epsilon_{t-i})^2, \tag{6.22}$$

$$s_T^2 = E(V_T^2) = E(S_T^2) = eT\sigma^2 (\sum_{i=1}^{m} (\nu' \tau_i)^2). \tag{6.23}$$

Expanding V_T^2 as

$$\begin{aligned} V_T^2 &= e \left(\sum_{i=1}^{m} (\nu' \tau_i)^2 \sum_{t=1}^{T} (\epsilon_{t-i}^2 - \sigma^2) + \sigma^2 T \sum_{i=1}^{m} (\nu' \tau_i)^2 + \sum_{\substack{i \neq j \\ =1}}^{m} (\nu' \tau_i)(\nu' \tau_j) \sum_{t=1}^{T} \epsilon_{t-i} \epsilon_{t-j} \right), \\ &= (i) + (ii) + (iii), \end{aligned}$$

by a martingale's LLN (Hall and Heyde 1980) we get for any $i, j = 1, \ldots, m$ and $i \neq j$

$$\frac{1}{T} \sum_{t=1}^{T} (\epsilon_{t-i}^2 - \sigma^2) \to_p 0,$$

$$\frac{1}{T} \sum_{t=1}^{T} \epsilon_{t-i} \epsilon_{t-j} \to_p 0,$$

yielding

$$\frac{(i)}{s_T^2} \to_p 0, \; \frac{(ii)}{s_T^2} = 1, \; \frac{(iii)}{s_T^2} \to_p 0.$$

so that

$$\frac{V_T^2}{s_T^2} \to_p 1. \tag{6.24}$$

For any $\epsilon > 0$ and some positive constant K, from

$$E(U_1^2) = e\sigma^2 \sum_{i=1}^{m} (\nu' \tau_i)^2 < \infty,$$

and from $s_T \sim K T^{1/2}$ as $T \to \infty$, it follows that

$$\lim_{T \to \infty} E(U_1^2 I(| U_1 | \geq \epsilon s_T) = 0.$$

Then the conditions for applying (Brown 1971, Theorem 2) hold, viz.

$$\frac{V_T^2}{s_T^2} \to_p 1 \text{ and } \lim_{T\to\infty} \frac{\sum_{t=1}^T E(U_1^2 I(\mid U_1 \mid \geq \epsilon\, s_T))}{s_T^2} = 0\,,$$

so that

$$\frac{S_T}{s_T} \to_d \mathcal{N}(0,1)\,, \tag{6.25}$$

where

$$\frac{S_T}{s_T} = \frac{\nu' \sum_{i=1}^m \tau_i \sum_{t=1}^T \epsilon_{t-i}\bar{\eta}_t}{(T\, e\, \sigma^2)^{1/2} (\sum_{i=1}^m (\nu'\tau_i)^2)^{1/2}}\,. \tag{6.26}$$

By choosing the vector ν appropriately we then obtain for a given m

$$\frac{1}{T^{1/2}} \sum_{i=1}^m \tau_i \sum_{t=1}^T \epsilon_{t-i} X_t \;\to_d \mathcal{N}_p(0, e\,\sigma^2\, \Gamma_m) \tag{6.27}$$

$\square$.

Theorem 15 *Under H_0 and Assumptions C, D, E we obtain*

$$LM \to_d \chi_p\,.$$

Proof: We need to make use of Bernstein's Lemma (Hannan 1970, p.242) so to extend the result of Lemma 26 for $m \to \infty$. At first we obtain that

$$\hat{\sigma^2} \to_p \sigma^2 \text{ and } \hat{w} \to_p w\,,$$

for $y_t^2 - E(y_t^2)$ and $y_t - E(y_t)$ being stationary square integrable martingale difference and a direct use of Slutzky lemma. Thus we have that

$$\begin{aligned}
&X_t(\hat{\sigma^2}, \hat{w}) - X_t = \\
&= (\frac{y_t}{\hat{\sigma}^2} - \hat{w} - 1)^2 \frac{2(w-\hat{w})}{(1+2w)(1+2\hat{w})} + \frac{1}{(1+2w)}\Big(y_t^2\,(\frac{1}{\hat{\sigma}^4} - \frac{1}{\sigma^4}) + \\
&+ (\hat{w}^2 - w^2) + 2(\hat{w} - w) - 2(\hat{w} - w)\,\frac{y_t}{\hat{\sigma}^2} - 2(w+1) y_t (\frac{1}{\hat{\sigma}^2} - \frac{1}{\sigma^2})\Big)\,,
\end{aligned}$$

so that it follows

$$\eta_i(\hat{\sigma}^2, \hat{w}) - \eta_i = \sum_{t=i+1}^T \Big((x_{t-i} - \hat{\mu}) X_t(\hat{\sigma}^2, \hat{w}) - (x_{t-i} - \mu) X_t\Big) = o_p(T^{1/2})\,.$$

Also from C_2 we have that

$$\lim_{m\to\infty}\sum_{i=m}^{\infty}\tau_i\tau_i' = 0. \tag{6.28}$$

Then from

$$E(\bar{\eta_i}\,\eta_{\bar{i+j}}) = e\,T\,\sigma^2\delta(0,j),\ i\neq 0, \tag{6.29}$$

we get as $m\to\infty$

$$E\parallel\sum_{i=m}^{T-1}\tau_i\bar{\eta}_i\parallel^2\leq Te\sigma^2\sum_{i=m}^{\infty}\parallel\tau_i\parallel^2 = o(T). \tag{6.30}$$

so that Bernstein's Lemma holds.

□

Remarks: (1) we can also consider a frequency domain expression, asymptotically equivalent, for the test statistic which, by using the FFT, would be much faster to compute for long series. Defining

$$\tilde{X}(\omega) = \sum_{a=0}^{T-1}X_{a+1}(\hat{\sigma^2},\hat{w})e^{(ia\omega)}, \tag{6.31}$$

$$\tilde{x}(\omega) = \sum_{a=0}^{T-1}(x_{a+1}-\bar{x})e^{(ia\omega)}, \tag{6.32}$$

$$\tilde{\tau}(\omega) = \sum_{a=0}^{\infty}\tau_{a+1}e^{(ia\omega)}, \tag{6.33}$$

then by setting $\omega_j = \frac{2\pi j}{T}$ and denoting with $\sum^*$ a summation made skipping the indexes for which $\tilde{\tau}(\omega_j)$ is unbounded, we have

$$\tilde{LM} = \tilde{\lambda}'\tilde{\lambda}, \tag{6.34}$$

with (6.35)

$$\tilde{\lambda} = (T\,e(\hat{w})\hat{\sigma}^2\Gamma)^{-1/2}\frac{1}{T}\sum_{a=0}^{T-1,*}\tilde{\tau}(\omega_a)\tilde{x}(-\omega_a)\tilde{X}(-\omega_a), \tag{6.36}$$

where now the matrix Γ can be expressed as

$$\Gamma = \frac{1}{4\pi}\int_{-\pi}^{\pi}\tilde{\tau}(\omega)\tilde{\tau}'(-\omega)d\omega, \tag{6.37}$$

approximated by

$$\frac{1}{2T}\sum_{a=0}^{T-1,*}\tilde{\tau}(\omega_a)\tilde{\tau}'(-\omega_a). \tag{6.38}$$

(2) The obvious mild condition for consistency is that when H_0 is false

$$\sum_{i=1}^{\infty} \tau_i E(\eta_i) \neq 0 .$$

6.5 The LM test: a robustified version

Obviously Assumption D puts a lot of structure in the moments of the unobservable ϵ_t, in particular imposing the ϵ_t to behave as an i.i.d. Gaussian process up to the eighth moment. We can weaken Assumption D as follows and still obtain a test statistic which allows us to perform standard inference, as shown in Theorem 16 below.

Assumptions D'

D'_1 $E(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$.

D'_2 $E[(\epsilon_t^2 - \sigma^2)\epsilon_t^2|\mathcal{F}_{t-1}] = 2\sigma^4$.

D'_3 $E(\epsilon_1^{16}) < \infty$.

$$D'_4\ E(X_1 X_{1+i}\epsilon_{1-k}\epsilon_{1-r}^a) = \begin{cases} 0, & a = 0, \\ \begin{cases} 0, & i \geq 1,\ r+i \geq 1, \\ = f_k, & i = 0,\ k = r, \end{cases} & a = 1 \end{cases}$$

where $k, r \geq 1$ *always.*

D'_5 $\inf_{i \geq 1}\ E(X_1^2 \epsilon_{1-i}^2) > 0$.

D'_6 $\frac{1}{T}\sum_{t=1}^{T} \epsilon_{t-i}\epsilon_{t-j} E(X_t^2|\mathcal{F}_{t-1}) \rightarrow_{a.s.} f_i \delta(i,j)$.

Remark: Assumptions D' allow a great deal of heterogeneity in the ϵ_t. In fact we only need the ϵ_t to have constant conditional moment up to the fourth order, the trade off being the strict unconditional moment condition D'_3. Assumption D'_5 guarantees that the distribution of the score is non-singular asymptotically and assumption D_4' expresses the minimal degree of stationarity required in the ϵ_t. Assumption D_6' is a mild ergodicity condition needed in order to apply Bernstein's lemma.

Then we can obtain the same distributional result as in Theorem 15, obviously with a different asymptotic covariance matrix in Lemma 26. Let us define the robustified test statistic as

$$LM_R = \lambda_R' \lambda_R,$$

$$\lambda_R = (T\hat{\Gamma}_R)^{-1/2} \sum_{i=1}^{T-1} \tau_i \eta_i(\hat{\sigma^2}, \hat{w}),$$

and

$$\Gamma_R = \sum_{i=1}^{\infty} \tau_i \tau_i' f_i \,,$$

where

$$\hat{\Gamma_R} = \sum_{i=1}^{T-1} \tau_i \tau_i' \bar{f}_i(\hat{\sigma^2}, \hat{w}) \,,$$

with

$$\bar{f}_i = \frac{1}{T} \sum_{t=1}^{T} X_t^2 \epsilon_{t-i}^2$$

Theorem 16 *Under H_0 and Assumptions C, E, D'*

$$LM_R \rightarrow_d \chi_p \,.$$

Proof: Lemma 23 and Lemma 24 can be easily seen to be satisfied by using Assumptions D_2', D_3' .

Then by Assumptions D_1', D_4', D_6' we obtain that as $T \to \infty$

$$\frac{1}{T^{1/2}} \sum_{i=1}^{m} \tau_i \bar{\eta}_i \rightarrow_d \mathcal{N}_p(0, \sum_{i=1}^{m} \tau_i \tau_i f_i) \,. \tag{6.39}$$

Assumption D_5' guarantees that the matrix Γ_R is invertible. Now, by assumptions D_1', D_2' we have that

$$X_t - X_t(\hat{\sigma^2}, \hat{w}) = o_p(1) \,,$$

and so

$$\sum_{i=1}^{m} \tau_i \tau_i' (\bar{f}_i - \bar{f}_i(\hat{\sigma^2}, \hat{w})) \to_p 0,$$

for fixed m. Directly by Assumption D_6' we obtain

$$\sum_{i=1}^{m} \tau_i \tau_i' (\bar{f}_i - f_i) \to_p 0,$$

as well for fixed m and by F_3, $E|\bar{f}_i| = O(1)$ and $E|\bar{f}_i(\hat{\sigma^2}, \hat{w})| = O(1)$ so that

$$\sum_{i=m}^{T-1} \tau_i \tau_i' (E|\bar{f}_i| + E|\bar{f}_i(\hat{\sigma^2})| + f_i) \to_p 0.$$

as m (and then T) goes to infinity thus allowing to use Bernstein's Lemma.

□

6.6 The LM test: small sample properties and applications

In this section we will apply the LM statistic proposed in 6.3 using simulated observations of the one-shock model, as described in section 5.1.1 and both the LM and LM_R statistic using the empirical financial data set introduced in section 1.1.1.

6.6.1 Montecarlo results

We consider the following parameterization:

$$\sigma^2 = 1, \;\; \alpha_i(d) = \prod_{j=1}^{i} \frac{j+d-1}{j}, \; 0 < d < 1/2, \; i = 1, 2, \ldots,$$

implying

$$\tau_i = \frac{1}{i}, \; i = 1, 2, \ldots.$$

We calculate the LM statistic using its frequency domain expression (6.34) both on the simulated series x_t as well as on the simulated disturbance e_t drawn from a standard normal using the GAUSS random generator. In particular, the same e_t are used to simulate the x_t and thus to evaluate the LM statistic under the alternative as well as to evaluate the LM statistic itself under the null.

Note that this parameterization does not yield a closed form expression for Γ which will be calculated as in (6.38) where:

$$\tilde{\tau}(\omega) = log(2\,sin(\frac{\omega}{2})e^{i\,(\omega-\pi)/2}).$$

Note that we derived the asymptotic distribution for the LM test statistic under the null only. Therefore we cannot confront the significance values taken under the alternative here considered, viz. using the x_t, with some distribution, but we might obtain some evidence regarding the consistency of the test.

Each Montecarlo experiments is made of 1000 replications. We consider the following sample sizes :

$$T = \{128, 256, 512, 1024, 2048\}$$

and the following data generating values for the parameter d:

$$d = \{0.1, 0.2, 0.375, 0.45, 0.49\}.$$

Thus we can express the null (H_0) and the alternative hypothesis (H_1) as:

$$\begin{array}{rcll} H_0 & : & \alpha_j = 0, & j = 1, \ldots, \\ H_1 & : & \alpha_j(d) = \prod_{k=1}^{j} \frac{k+d-1}{k},\ 0 < d < 1/2, & j = 1, \ldots. \end{array}$$

In Table 6.1 we report jointly the results for the power (first three columns) and the size (last three columns) of the LM statistic introduced in section 6.3. Obviously the latter results are unaffected by the different values taken by the parameter d but we report them for sake of comparison.

Concerning the power of the test, the results clearly indicate how the probability of rejecting the null when this is not true is increasing with the sample size, as we can expect. Furthermore, for any given sample size, we observe that the power is increasing with the parameter d so that it shows the highest power for values of d equal and bigger than 0.375. Finally note that this last relation is not monotonic, given that the power always decreases passing from 0.45 to 0.49.

The not optimal performance of the LM test statistic near the nonstationary region (when $\rho \neq 0$) has an analogy in terms of estimation (cf. section 5.1.2) which we conjecture as being due to the difficulties in simulating a series which is nearly nonstationary. Nevertheless the results show that the LM test statistic is still consistent in this case also.

In terms of size, we note that the empirical sizes are close to the nominal ones, the closer the bigger the sample size, in agreement with Theorem 15, but with no striking difference between different sample sizes differently from the power cases.

6.6.2 Empirical applications

We consider the seven time series introduced in section 1.1.1. Table 6.2 displays the LM test statistic from the data both in its non robust (LM) and robust version (LM_R). In fact, differently from the simulated data, the figures for the empirical kurtosis of the data (cf. Table 1.1) suggest a violation of Assumption D which the non robust LM statistic is based on. For each case we have calculated the test statistics for the entire sample (last two rows) as well as for the first $1,000$ (from 1st January 1993 to 31th October 1988) and first $2,000$ (from 1st January 1993 to 31 August 1992) in the first two rows (marked $1,000$) and in the second two rows (marked $2,000$) respectively. In this way, we can in principle take into account the effect of extraordinary events, as the crash of Monday 17th October 1987 for stock returns, on the data in terms of conditional heteroskedasticity.

The non robust LM statistic is always highly significant for all but the $Dollar/Pound$ series, especially for the stock return indexes where we find extremely strong evidence of long-range dependent volatility. When considering the LM_R statistic , as we can expect, the results are much less clear. For the stock return indexes, the test statistic is significant for the FT indexes over the entire sample. For the foreign exchange returns, the tests statistic significantly rejects the hypothesis of no conditional heteroskedasticity in the first subsample for the $Yen/Pound$ and in the second subsample for the $Dollar/Pound$.

Table 6.1: **LM test : size and power**

	H_1 (power)			H_0 (size)		
	significance value			significance value		
d	10 %	5 %	1 %	10 %	5 %	1 %
sample size : 128						
0.1	0.085	0.580	0.029	0.050	0.032	0.014
0.2	0.214	0.168	0.101	0.056	0.031	0.015
0.375	0.374	0.315	0.226	0.064	0.046	0.016
0.45	0.384	0.320	0.230	0.061	0.036	0.013
0.49	0.288	0.227	0.159	0.059	0.028	0.013
sample size : 256						
0.1	0.111	0.076	0.022	0.061	0.025	0.006
0.2	0.280	0.219	0.134	0.070	0.038	0.013
0.375	0.506	0.441	0.353	0.076	0.037	0.017
0.45	0.510	0.461	0.367	0.080	0.047	0.014
0.49	0.404	0.340	0.239	0.080	0.054	0.021
sample size : 512						
0.1	0.143	0.093	0.046	0.079	0.047	0.016
0.2	0.370	0.298	0.184	0.096	0.054	0.012
0.375	0.602	0.541	0.455	0.073	0.036	0.018
0.45	0.613	0.556	0.471	0.080	0.044	0.018
0.49	0.461	0.402	0.313	0.078	0.037	0.015
sample size : 1024						
0.1	0.184	0.124	0.066	0.082	0.052	0.014
0.2	0.424	0.343	0.214	0.084	0.041	0.009
0.375	0.692	0.647	0.541	0.094	0.045	0.010
0.45	0.673	0.625	0.534	0.092	0.048	0.016
0.49	0.529	0.475	0.380	0.095	0.056	0.017
sample size : 2048						
0.1	0.184	0.134	0.055	0.093	0.050	0.010
0.2	0.457	0.371	0.249	0.082	0.045	0.006
0.375	0.744	0.708	0.621	0.089	0.049	0.010
0.45	0.750	0.703	0.631	0.088	0.052	0.008
0.49	0.589	0.515	0.422	0.112	0.052	0.020

For each sample size and parameter d value
1000 Montecarlo replications are performed
with the other parameter values equal to $\sigma^2 = 1$, $\rho = 1$.
Columns $2-4$ reports the empirical power and
columns $5-7$ the empirical size of the (normal) LM test.

Table 6.2: **LM test : empirical applications**

size	test type	sYP	fYP	sUP	fUP	F100	FAll	S500
1000	LM	64.861	67.637	1.206	0.909	158.325	153.894	11803
1000	LM_R	2.678	2.831	0.247	0.185	0.109	0.098	0.573
2000	LM	12.767	12.363	29.142	31.388	451.836	1651	558.801
2000	LM_R	1.203	1.157	4.629	4.895	0.120	0.582	0.0173
3088	LM	37.961	11.407	9.650	13.297	15497	18355	70242
3088	LM_R	2.787	0.638	1.489	1.927	9.554	17.519	0.758

The data refers to the period 1^{st} Jan 1986 to 1^{st} Nov 1993 (3088 observations).
We report the values obtained by the (normal) LM test statistic (rows indexed LM) and the robust LM test statistic (rows indexed LM_R) for two subsamples of size 1000 and 2000 as well as for the entire sample to check for the effect of turbulent periods (as Monday $17^{th} October 1987$).
The critical values for a χ^2 with 1 d.o.f. at 10% and 1% are respectively 2.70 and 3.84.
All the calculations of this table are based on the discrete Fourier transform.

Chapter 7

Long memory moving average models: the 'two-shock' case

Here we introduce a nonlinear model of stochastic volatility within the class of 'product type' models. It allows different degrees of dependence for the 'raw' series and for the 'squared' series, for instance implying weak dependence in the former and long memory in the latter. Its main statistical properties and an estimation procedure based on a Gaussian PMLE are discussed.

7.1 Introduction

In this chapter we will introduce a second class of nonlinear MA models. In order better to understand its motivation, we must remember that an alternative approach to ARCH-type modelling is suggested by SV models (cf. section 1.1.5). Following Robinson and Zaffaroni (1996b), we replace (2.2) by

$$x_t = \eta_t h_{t-1}, \tag{7.1}$$

with h_t given in (2.5), but $\{\eta_t\}$ is an i.i.d. sequence independent of $\{\epsilon_t\}$. There is thus a decoupling of the two factors, and (7.1) can be called a 'two-shock' model as distinct from the 'one-shock' case (2.2). In the SV literature h_t and y_t have short memory, the weights α_j in (2.5) being assumed at least summable (so that they could be the coefficients in the moving average representation of a stationary autoregressive moving average sequence) but as in Chapter 2, we can choose the α_j to impart long memory to h_t and y_t. As has already been observed (cf. section 1.1.5), an advantage of (7.1) over (2.2) is that the independence of the two factors leads to simplification in moment formulae and thus should also simplify asymptotic theory for Gaussian estimation relative to (2.2). On the other hand, the latent structure of the volatility process makes 'in-sample' (smoothing) and 'out-of-sample' (filtering) forecasting a very difficult task. From an economic point of view, it seems difficult to give a rationale to the latent volatility process. Furthermore, in order to make the model apt to display the 'leverage' effect, one needs to relax the independence assumption, loosing one of the main qualities of the model. For these reasons, the analysis of the 'two-shock' model will not be as deep as for the 'one-shock' case. The analysis will be dealt within this chapter only with no formal proofs of the asymptotic properties of the Gaussian PMLE here proposed. An alternative two-shock model is in Harvey (1993).

In the following section we describe a model that is actually rather more general than (7.1) in two respects, and we derive its memory properties and kurtosis. The greater generality is due partly to allowing arbitrary memory in the raw x_t, and partly because we do not impose linearity anywhere, so far as consideration of autocorrelation properties is concerned. We also replace h_{t-1} by h_t, a function of $\mathcal{F}_t$. We make use of some general results on the second order properties results on the second order properties of certain nonlinear functions which are stated and proved in Appendix A. Section 7.3 specializes to the case of linear processes because these are the likely vehicles for parametric modelling. Section 7.4 discusses Gaussian estimation of a parametric model and section 7.5 adapts standard results to obtain a filtering and a forecasting algorithm. Section 7.6 estimates a simple version of this model from empirical data.

7.2 Definitions, assumptions and statistical properties

We extend (7.1) to

$$x_t = g_t + \eta_t h_t, \tag{7.2}$$

where the right hand side variates obey the following condition.

Assumption G

The process $\{\eta_t\}$ is serially uncorrelated, with $E(\eta_t) = E(\eta_t^3) = 0$, $var(\eta_t) = \sigma_{\eta\eta}$. The bivariate process $\{g_t, h_t\}$ is independent of $\{\eta_t\}$ and fourth order stationary with zero joint third cumulants, and for $a_t, b_t \in \{g_t, h_t\}$ we define

$$
\begin{aligned}
&E(a_t) = \mu_a,\\
&\gamma_{ab}(j) = cov(a_0, b_j),\ j = 0, \pm 1, \ldots,\\
&\ \textit{where } \gamma_{ab}(j) = \gamma_a(j) \textit{ when } a = b,\\
&\kappa_{ab}(j) = cum_4(a_0, a_0, b_j, b_j),\ j = 0, 1, \ldots.
\end{aligned}
$$

We also introduce a stronger condition, which holds under Gaussianity.

Assumption H

Assumption G holds and for $a_t, b_t \in \{g_t, h_t\}$ $\kappa_{ab}(j) = 0$, any $j = 0, 1, \ldots$.

Theorem 17 *Under Assumption G, for all $u = 0, 1, \ldots,$*

$$
\begin{aligned}
&(i)\gamma_x(u) \ = \gamma_g(u),\\
&(ii)\gamma_y(u) = 2\left[\gamma_g^2(u) + \kappa_{gg}(u)/2 + \sigma_{\eta\eta}\left(\gamma_{gh}^2(u) + \gamma_{hg}^2(u) + \kappa_{gh}(u)/2 + \kappa_{hg}(u)/2\right)\right.\\
&\left. + \sigma_{\eta\eta}^2\left(\gamma_h^2(u) + \kappa_{hh}(u)/2 + 2\mu_h^2\gamma_h\right) + 2\mu_g^2\gamma_g(u)\right] + \nu_y\delta(u, 0),
\end{aligned}
$$

where

$$
\begin{aligned}
\nu_y \ = \ & (\kappa_{\eta\eta} + 2\sigma_{\eta\eta}^2)\left(\kappa_{hh} + 2\gamma_h^2(0) + 4\mu_h^2\gamma_h(0) + (\mu_h^2 + \gamma_h(0))^2\right) + 8\sigma_{\eta\eta}\mu_g\mu_h\gamma_{gh}(0)\\
& + 4\sigma_{\eta\eta}\left(\kappa_{gh} + \gamma_g(0)\gamma_h(0) + \gamma_{gh}^2(0) + \mu_g^2\gamma_h(0) + \mu_h^2\gamma_g(0) + 2\mu_g\mu_h\gamma_{gh}(0)\right),
\end{aligned}
$$

$\kappa_{ab} = \kappa_{ab}(0)$.

Proof:
(i) Immediate given independence of $\{g_t, h_t\}$ and $\{\eta_t\}$, and observing that $\mu_x = \mu_g$.

(ii) From Lemma 31 in Appendix A,

$$\gamma_y(u) = cov\left((x_0 - \mu_x)^2, (x_u - \mu_x)^2\right) + 4\mu_x^2\gamma_x(u).$$

By Lemma 32 in Appendix A the first term on the right hand side becomes

$$\begin{aligned}
&cov\left((g_0 - \mu_g + \eta_0 h_0)^2, (g_u - \mu_g + \eta_u h_u)^2\right) \\
&= cov\left((g_0 - \mu_g)^2 + 2(g_0 - \mu_g)\eta_0 h_0 + \eta_0^2 h_0^2, (g_u - \mu_g)^2 + 2(g_u - \mu_g)\eta_u h_u + \eta_u^2 h_u^2\right) \\
&= a + \sigma_{\eta\eta}(b + c) + \sigma_{\eta\eta}^2 d
\end{aligned}$$

where $a = cov((g_0 - \mu_g)^2, (g_u - \mu_g)^2), b = cov((g_0 - \mu_g)^2, h_u^2), c = cov(h_0^2, (g_u - \mu_g)^2), d = cov(h_0^2, h_u^2)$.

From Lemmas 31 and 33 in Appendix A, $a = 2\gamma_g^2(u) + \kappa_{gg}(u), b = 2\gamma_{gh}^2(u) + \kappa_{gh}(u), c = 2\gamma_{hg}^2(u) + \kappa_{hg}(u), d = \kappa_{hh}(u) + 2\gamma_h^2(u) + 4\mu_h^2\gamma_h(u)$. The result follows from (i) and routine computation.

□

Under Assumption G the autocovariance properties of x_t are inherited from those of g_t. Moreover if η_t and g_t are also martingale differences, so is x_t.

The different possibilities allowed by the model in terms of degree of dependence for the squared process y_t are indicated as follows. Assuming that Assumption H and the mild ergodicity condition

$$\gamma_g(u) \to 0, \quad \gamma_h(u) \to 0, \quad \text{as } u \to \infty,$$

hold, we deduce that, as $u \to \infty$:

(*i*) When $\mu_g = 0$
$$\gamma_y(u) \sim 2\left[\gamma_g^2(u) + \sigma_{\eta\eta}\left(\gamma_{gh}^2(u) + \gamma_{hg}^2(u)\right) + 2\sigma_{\eta\eta}^2\mu_h^2\gamma_h(u)\right].$$
(*ii*) When $\mu_h = 0$
$$\gamma_y(u) \sim 2\left[\mu_g^2\gamma_g(u) + \sigma_{\eta\eta}\left(\gamma_{gh}^2(u) + \gamma_{hg}^2(u)\right) + \sigma_{\eta\eta}^2\gamma_h^2(u)\right].$$

(iii) When $\mu_g = \mu_h = 0$

$\gamma_y(u) \sim 2\left[\gamma_g^2(u) + \sigma_{\eta\eta}\left(\gamma_{gh}^2(u) + \gamma_{hg}^2(u)\right) + \sigma_{\eta\eta}^2\gamma_h^2(u)\right].$

(iv) Otherwise

$\gamma_y(u) \sim 2\left[\mu_g^2\gamma_g(u) + \sigma_{\eta\eta}\left(\gamma_{gh}^2(u) + \gamma_{hg}^2(u)\right) + \sigma_{\eta\eta}^2 2\mu_h^2\gamma_h(u)\right].$

Returning to the general setting, we can apply these results to the circumstances envisaged in section 1.1.1, to achieve a white noise or short memory x_t and a long memory y_t. Let us assume that for $j \to \infty$

$$\gamma_h(j) \sim K_h \mid j \mid^{2d-1},$$

$$\gamma_g(j) = o(\mid j \mid^{4d-2}),$$

with $0 < d < 1/2$ and $0 < K_h < \infty$. Here h_t has long memory (with memory parameter d) because the $\gamma_h(j)$ are not summable. Clearly g_t has shorter memory than y_t for all $d \in (0, 1/2)$, and for $d \in (0, 1/4)$ it does not have long memory because the $\gamma_g(j)$ are summable, while long memory in g_t is a possibility when $1/4 < d < 1/2$. Let us assume also that g_t and h_t are uncorrelated, so that $\gamma_{gh}(j) = 0$, or more generally that $\gamma_{gh}(j) = o(\mid j \mid^{2d-1})$, as $\mid j \mid \to \infty$. Then from $(i) - (iv)$ we deduce that asymptotically it does not matter whether or not μ_g is zero, and

$$\gamma_y(j) \sim \begin{cases} 4\sigma_{\eta\eta}^2\mu_h^2\gamma_h(j) \sim 4K_h\mu_h^2\sigma_{\eta\eta}^2 \mid j \mid^{2d-1}, & \mu_h \neq 0, \\ 2\sigma_{\eta\eta}^2\gamma_h^2(j) \sim 2K_h^2\sigma_{\eta\eta}^2 \mid j \mid^{4d-2}, & \mu_h = 0. \end{cases} \tag{7.3}$$

Thus when $\mu_h \neq 0$, y_t exhibits long memory for all $d \in (0, 1/2)$, while when $\mu_h = 0$, it does so when $d \in (1/4, 1/2)$.

We can also give an expression for the coefficient of kurtosis of the process x_t.

Theorem 18 *Under Assumption H*

$$kurt(x_t) = 3 + \frac{12\sigma_{\eta\eta}\gamma_{gh}^2(0) + 6\sigma_{\eta\eta}^2\gamma_h(0)(\gamma_h(0) + 2\mu_h^2)}{(\gamma_g(0) + \sigma_{\eta\eta}(\mu_h^2 + \gamma_h(0)))^2}.$$

Proof: Writing

$$kurt(x_t) = \frac{var\left((x_t - \mu_x)^2\right)}{(var(x_t))^2} + 1,$$

we will evaluate the numerator and the denominator separately. By direct calculation using Lemmas 32 and 33 in Appendix A we get

$$
\begin{aligned}
&var(x_t) = \gamma_g(0) + \sigma_{\eta\eta}(\gamma_h(0) + \mu_h^2),\\
&var\left((x_t - \mu_x)^2\right)\\
&= cov\left((g_t - \mu_g)^2 + \eta_t^2 h_t^2 + 2\eta_t h_t(g_t - \mu_g), (g_t - \mu_g)^2 + \eta_t^2 h_t^2 + 2\eta_t h_t(g_t - \mu_g)\right)\\
&= var((g_t - \mu_g)^2) + 2cov((g_t - \mu_g)^2, \eta_t^2 h_t^2) + var(\eta_t^2 h_t^2) + 4var(\eta_t h_t(g_t - \mu_g))\\
&= a + 2b + c + 4d,\\
&\text{ where}\\
&a = 2\gamma_g^2(0),\\
&b = \sigma_{\eta\eta} 2\gamma_{gh}^2(0),\\
&c = 3\sigma_{\eta\eta}^2(4\mu_h^2\gamma_h(0) + 2\gamma_h^2(0)) + 2\sigma_{\eta\eta}^2(\gamma_h(0) + \mu_h^2)^2,\\
&d = \sigma_{\eta\eta}\left(\gamma_h(0)\gamma_g(0) + \mu_h^2\gamma_g(0) + 2\gamma_{gh}^2(0)\right).
\end{aligned}
$$

The result follows by straightforward manipulation.

□

Because the second term on the right hand side in Theorem 18 is positive, x_t has fatter tails than a Gaussian process, but as $\mu_h^2 \to \infty$, Gaussian kurtosis is approached.

We shall now derive the power spectra for the x_t and the y_t processes, assuming $\{g_t, h_t\}$ have jointly absolutely continuous spectral distribution function. We denote the cross spectral density of processes a_t, b_t by $f_{ab}(\lambda)$, satisfying

$$\gamma_{ab}(u) = \int_{-\pi}^{\pi} f_{ab}(\omega)e^{iu\omega}d\omega, \quad u = 0, \pm1, \ldots. \tag{7.4}$$

where $f_{ab}(\lambda) = f_a(\lambda)$ when $a = b$.

Theorem 19 *For any $-\pi \leq \lambda < \pi$:*

(i) under Assumption G

$$f_x(\lambda) = f_g(\lambda);$$

(ii) under Assumption H

$$
\begin{aligned}
f_y(\lambda) \quad &= 2\left[\int_{-\pi}^{\pi} f_g(\mu) f_g(\lambda-\mu)d\mu + 2\sigma_{\eta\eta}\int_{-\pi}^{\pi} \mathcal{R}e(f_{gh}(\mu)f_{gh}(\lambda-\mu))d\mu \right. \\
&\left. +2\mu_g^2 f_g(\lambda) + \sigma_{\eta\eta}^2\left(\int_{-\pi}^{\pi} f_h(\mu)f_h(\lambda-\mu)d\mu + 2\mu_h^2 f_h(\lambda)\right)\right] + \frac{v_y}{2\pi}.
\end{aligned}
$$

Proof: (*i*) The proof follows directly from Theorem 17.

(*ii*) In the expression for $\gamma_y(u)$ from (ii) Theorem 17 with all fourth order cumulants terms set to zero, substitute using (7.4) to obtain

$$
\begin{aligned}
\gamma_y(u) = 2&\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_g(\lambda)f_g(\omega)e^{iu(\lambda+\omega)}d\lambda d\omega + 2\mu_g^2\int_{-\pi}^{\pi} f_g(\lambda)e^{iu\lambda}\right. \\
&+\sigma_{\eta\eta}\left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(f_{gh}(\lambda)f_{gh}(\omega)+f_{hg}(\lambda)f_{hg}(\omega))e^{iu(\lambda+\omega)}d\omega d\lambda\right) \\
&+\sigma_{\eta\eta}^2\left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_h(\lambda)f_h(\omega)e^{iu(\lambda+\omega)}d\lambda d\omega + 2\mu_h^2\int_{-\pi}^{\pi} f_h(\lambda)e^{iu\lambda}d\lambda\right).
\end{aligned}
$$

Now make the change of variables from ω to $\mu = \omega + \lambda$ and equate the integrand with respect to λ to $f_y(\lambda)e^{iu\lambda}$, in view of (7.4).

□

7.3 Linear g_t and h_t

For the purpose of finite-parameter modelling it is likely that we will specify g_t and h_t to be linear processes, as in:

Assumption J

$$
g_t = \mu_g + \sum_{i=0}^{\infty}\beta_i\epsilon_{t-i}, \quad h_t = \mu_h + \sum_{i=0}^{\infty}\alpha_i\epsilon_{t-i},
$$

where the coefficients $\{\alpha_i\}$ and $\{\beta_i\}$ at minimum satisfy

$$\sum_{i=0}^{\infty}\alpha_i^2 < \infty, \ \sum_{i=0}^{\infty}\beta_i^2 < \infty,$$

and

$$E(\epsilon_t) = 0\,, t = 0, \pm 1, \ldots$$

$$E(\epsilon_t\epsilon_s) = \begin{cases} \sigma_{\epsilon\epsilon}\,, \ s = t \\ 0\,, \ s \neq t \end{cases} \tag{7.5}$$

$$E(\epsilon_s\epsilon_t\epsilon_u) = 0\,, \ \forall s, t, u$$

$$E(\epsilon_s\epsilon_t\epsilon_u\epsilon_v) = \begin{cases} 3\sigma_{\epsilon\epsilon}^2 + \kappa_{\epsilon\epsilon} & s = t = v = u \\ \sigma_{\epsilon\epsilon}^2 & \begin{cases} s = t \neq v = u \\ s = u \neq t = v \\ s = v \neq t = u \end{cases} \\ 0 & \textit{otherwise.} \end{cases}$$

Thus ϵ_t behaves as an *i.i.d.* sequence up to fourth moments.

Under Assumption J, h_t and g_t satisfy Assumption G.

Corollary 5 *Under Assumption J, for $u = 0, 1, \ldots,$*

$$(i)\gamma_x(u) = \sigma_{\epsilon\epsilon}\sum_{i=0}^{\infty}\beta_i\beta_{i+u}.$$

$$(ii)\ \gamma_y(u) = \kappa_{\epsilon\epsilon}\sum_{i=0}^{\infty}\beta_i^2\beta_{i+u}^2 + 2\sigma_{\epsilon\epsilon}^2(\sum_{i=0}^{\infty}\beta_i\beta_{i+u})^2 + 4\mu_g^2\sigma_{\epsilon\epsilon}\sum_{i=0}^{\infty}\beta_i\beta_{i+u}$$

$$+\sigma_{\eta\eta}\left(\kappa_{\epsilon\epsilon}\sum_{i=0}^{\infty}\beta_i^2\alpha_{i+u}^2 + 2\sigma_{\epsilon\epsilon}^2(\sum_{i=0}^{\infty}\beta_i\alpha_{i+u})^2 + \kappa_{\epsilon\epsilon}\sum_{i=0}^{\infty}\alpha_i^2\beta_{i+u}^2 + 2\sigma_{\epsilon\epsilon}^2(\sum_{i=0}^{\infty}\alpha_i\beta_{i+u})^2\right)$$

$$+\sigma_{\eta\eta}^2\left(\kappa_{\epsilon\epsilon}\sum_{i=0}^{\infty}\alpha_i^2\alpha_{i+u}^2 + 2\sigma_{\epsilon\epsilon}^2(\sum_{i=0}^{\infty}\alpha_i\alpha_{i+u})^2 + 4\mu_h^2\sigma_{\epsilon\epsilon}\sum_{i=0}^{\infty}\alpha_i\alpha_{i+u}\right) + \nu_y\delta(u,0),$$

with

$$\begin{aligned}\nu_y \ = \ & (\kappa_{\eta\eta}+2\sigma_{\eta\eta}^2)\left(\kappa_{\epsilon\epsilon}\Sigma_{\alpha^4} + 2\sigma_{\epsilon\epsilon}^2(\Sigma_{\alpha^2})^2 + 4\mu_h^2\sigma_{\epsilon\epsilon}\Sigma_{\alpha^2} + (\mu_h^2 + \sigma_{\epsilon\epsilon}\Sigma_{\alpha^2})^2\right) + 8\sigma_{\eta\eta}\mu_g\mu_h\sigma_{\epsilon\epsilon}\Sigma_{\alpha\beta} \\ & +4\sigma_{\eta\eta}\left(\kappa_{\epsilon\epsilon}\Sigma_{\alpha^2\beta^2} + \sigma_{\epsilon\epsilon}^2\Sigma_{\alpha^2}\Sigma_{\beta^2} + \sigma_{\epsilon\epsilon}^2(\Sigma_{\alpha\beta})^2 + \mu_g^2\sigma_{\epsilon\epsilon}\Sigma_{\alpha^2} + \mu_h^2\sigma_{\epsilon\epsilon}\Sigma_{\beta^2} + 2\mu_g\mu_h\sigma_{\epsilon\epsilon}\Sigma_{\alpha\beta}\right),\end{aligned}$$

where $\sum_c = \sum_{i=0}^{\infty} c_i$ for any sequence $\{c_i\}$.

It follows that when $\beta_i = 0$, $i \geq 1$ the raw process x_t is a white noise, but not a martingale difference sequence. To achieve the latter property we would require ϵ_t to be a martingale difference sequence, a stronger condition than (7.5).

Corollary 6 *Under Assumption J*

$$kurt(x_t) = 3 + \frac{12\sigma_{\epsilon\epsilon}^2 \sigma_{\eta\eta} (\sum_{\alpha\beta})^2 + 6\sigma_{\epsilon\epsilon}\sigma_{\eta\eta}^2 \left(\sigma_{\epsilon\epsilon}(\sum_{\alpha^2})^2 + 2\mu_h^2 \sum_{\alpha^4}\right)}{\left(\sigma_{\epsilon\epsilon} \sum_{\beta^2} + \sigma_{\eta\eta}(\mu_h^2 + \sigma_{\epsilon\epsilon} \sum_{\alpha^2})\right)^2}$$
$$+ \frac{1}{(\sigma_{\epsilon\epsilon} \sum_{\beta^2} + \sigma_{\eta\eta}(\mu_h^2 + \sigma_{\epsilon\epsilon} \sum_{\alpha^2}))^2} \times$$
$$\left[\kappa_{\epsilon\epsilon} \sum_{\beta^4} + 3\sigma_{\eta\eta}^2 \kappa_{\epsilon\epsilon} \sum_{\alpha^4} + 6\sigma_{\eta\eta}\kappa_{\epsilon\epsilon} \sum_{\alpha^2\beta^2} + \kappa_{\eta\eta} \left(\kappa_{\epsilon\epsilon} \sum_{\alpha^4} + 2\sigma_{\epsilon}^4 (\sum_{\alpha^2})^2 + 4\mu_h^2 \sigma_{\epsilon\epsilon} \sum_{\alpha^4} + (\mu_h^2 + \sigma_{\epsilon\epsilon} \sum_{\alpha^2})^2\right)\right].$$

Again note that with $\kappa_{\eta\eta} = \kappa_{\epsilon\epsilon} = 0$ the coefficient of kurtosis decreases to 3 as $\mu_h^2 \to \infty$.

By denoting the transfer functions of the α_i and β_i coefficients respectively by

$$\alpha(\lambda) = \sum_{j=0}^{\infty} \alpha_j e^{ij\lambda}, \ \beta(\lambda) = \sum_{j=0}^{\infty} \beta_j e^{ij\lambda},$$

we obtain the power spectra for the raw process x_t and the squared process y_t.

Corollary 7 *Under Assumption J*

$$(i) f_x(\lambda) = \frac{\sigma_{\epsilon\epsilon}}{2\pi} \mid \beta(\lambda) \mid^2,$$

and if also $\kappa_{\epsilon\epsilon} = 0$

$$(ii) f_y(\lambda) = \frac{\sigma^2_{\epsilon\epsilon}}{2\pi}\left(\frac{2}{2\pi}\int_{-\pi}^{\pi} \mid \beta(\mu)\beta(\lambda-\mu)\mid^2 d\mu\right.$$

$$+\frac{4\sigma_{\eta\eta}}{2\pi}\int_{-\pi}^{\pi} \mathcal{R}e(\alpha(\mu)\beta(-\mu)\alpha(\lambda-\mu)\beta(-\lambda+\mu))d\mu$$

$$\left.+4\frac{\mu_g^2}{\sigma_{\epsilon\epsilon}} \mid \beta(\lambda)\mid^2 + \frac{2\sigma^2_{\eta\eta}}{2\pi}\int_{-\pi}^{\pi} \mid \alpha(\mu)\alpha(\lambda-\mu)\mid^2 d\mu + 4\sigma^2_{\eta\eta}\frac{\mu_h^2}{\sigma_{\epsilon\epsilon}} \mid \alpha(\lambda)\mid^2\right) + \frac{v_y}{2\pi}.$$

7.4 Estimation

Following chapter 3, we propose as a simple expedient a PMLE based on a Gaussian likelihood as if y_t were Gaussian. Of course, from the assumptions made on the distribution of the unobservable processes g_t, η_t, h_t, the squared process y_t cannot be Gaussian, being always non-negative. Indeed, x_t also cannot be Gaussian even assuming g_t, h_t, η_t to be so. However, given the latent structure of the model there is no simple way to invert the model and to write down the true likelihood on the basis of, say, Gaussian g_t, h_t, η_t.

In the linear set-up of the previous section introduce functions $\alpha(\lambda;\theta), \beta(\lambda;\theta)$ of λ and a $p \times 1$ vector θ and define $f_y(\lambda;\psi)$ by the formula for $f_y(\lambda)$ in Corollary 7 with $\alpha(\lambda), \beta(\lambda)$ replaced by $\alpha(\lambda;\theta), \beta(\lambda;\theta)$, for $\psi = (\mu_g, \mu_h, \sigma_{\eta\eta}, \theta', \sigma_{\epsilon\epsilon})'$. Denote by ψ_0 the true value of ψ, so that

$$f_y(\lambda) = f_y(\lambda;\psi_0).$$

Introducing the periodogram based on T observations

$$I(\lambda) = \frac{1}{2\pi T} \mid \sum_{t=1}^{T} y_t e^{it\lambda} \mid^2,$$

and denoting $\lambda_j = 2\pi j/T$, we consider the discrete version of Whittle's Gaussian pseudo log likelihood introduced in Chapter 3 :

$$Q_T(\psi) = 1/T \sum_{j=1}^{T-1}\left(\log f_y(\lambda_j;\psi) + \frac{I(\lambda_j)}{f_y(\lambda_j;\psi)}\right). \tag{7.6}$$

The PMLE is

$$\hat{\psi} = argmin_{\psi \in \Psi} Q_T(\psi),$$

for a compact Ψ.

It should not be difficult to establish $T^{1/2}$ consistency and asymptotic normality of the PMLE given the stationarity of the processes involved and the relatively simple moment structure. With respect to consistency we should be able to adapt the approach of Hannan (1973) under ergodicity assumptions, following the proof for the PMLE of the 'one-shock' case (cf. Chapter 4, section 4.1). A critical aspect is checking identifiability, which depends on the parameterization chosen, e.g. in the linear case above we might need to set $\sigma_{\epsilon\epsilon} = 1$ depending on whether we set $\mu_g = 0$ or not. So far as asymptotic normality is concerned (cf. Chapter 4 section 4.2), we cannot use central limit theorems for weakly dependent processes on the one hand, or the available results on linear long memory processes on the other (e.g. see Giraitis and Surgailis (1990b), Heyde and Gay (1993)), but it seems that the method of moments can be applied, especially under the simplifying assumption that η_t and ϵ_t are Gaussian. Indeed, the independence assumption of the two shocks will yield a much simpler analysis than the 'one-shock' case in section 4.2.

We conjecture that, as $T \to \infty$,

$$T^{1/2}(\hat{\psi} - \psi) \to_d \mathcal{N}(0, A^{-1}BA^{-1}),$$

where

$$\begin{aligned} A(\psi) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} c(\lambda; \psi) c'(\lambda; \psi) d\lambda, \\ B(\psi) &= 4\pi \int_{-\pi}^{\pi} c(\lambda; \psi) c'(\lambda; \psi) d\lambda \\ &+ 2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{c(\lambda; \psi)}{f_y(\lambda; \psi)} \frac{c'(\omega; \psi)}{f_y(\omega; \psi)} Q_{yyy}(-\lambda, \omega, -\omega) d\lambda d\omega, \end{aligned}$$

and

$$c(\lambda; \psi) = \frac{\partial log\, f_y(\lambda; \psi)}{\partial \psi},$$

with $A = A(\psi_0), B = B(\psi_0)$ and where $Q_{yyy}(\lambda_1, \lambda_2, \lambda_3)$ denotes the trispectrum of y_t. To perform approximate statistical inference we can plug $\hat{\psi}$ in the expression for $A(\psi)$ and $B(\psi)$ or discrete approximations of these, for example we can replace $A(\psi)$ by

$$\bar{A}_T(\psi) = \frac{1}{T} \sum_{j=1}^{T-1} c(\lambda_j; \psi) c'(\lambda_j; \psi),$$

and estimate the double integral on the right hand side of $B(\psi)$, involving $Q_{yyy}(.,.,.)$, as in Taniguchi (1982) or Keenan (1987).

7.5 Signal extraction and prediction.

Given a long memory parameterization, we cannot use the State Space/Kalman Filter devices because the state-vector would have infinite dimension due to the long memory parameterization, whereas the state space approach requires the state vector to follow a Markovian law.

Thus we propose to extract the g_t signal and the $\eta_t^2 h_t^2$ signal by a direct approach of signal extraction based on the classic minimum mean square linear estimator (MMSLE) (e.g. in Naylor and Sell (1982)). The same methodology is employed to obtain forecasts that are optimal in the MMSLE sense.

Assume that the set of $(m+n)$ random variables $z = (z_1, \ldots, z_m)', u = (u_1, \ldots, u_n)'$ all with finite second moment, are such that z is unobservable and u is observable. Then let us assume that we want to estimate the random variables z_i, $i = 1, \ldots, m$ by a linear function of u (including a constant) called $\hat{z}_i$:

$$\hat{z}_i = a_0^i + a_1^i u_1 + a_2^i u_2 + \ldots + a_n^i u_n, \; i = 1, \ldots, m,$$

so that $E(z_i - \hat{z}_i)^2$ is minimum for the unknown constant weights $a_0^i, a_1^i, \ldots, a_n^i$, $i = 1, \ldots, m$. Classical results yield the weights as the solution of the following linear system:

$$B = CA,$$

where $B = E(v\, z'), C = E(v\, v')$ is assumed to have full rank and the $n \times m$ matrix A has $(i,j)th$ element (a_{i-1}^j) where $v = (1, u')'$. Hence the MMSLE of z is

$$\hat{z} = B'C^{-1}v.$$

This MMSLE-based procedure is very practical albeit not fully efficient, unless the vector (z', u') is multivariate normal.

In order to apply this methodology to our model, we propose a two-stage approach. In the first stage we extract the g_t component. We rewrite (7.2) as

$$x_t = \mu_x + (g_t - \mu_g) + \eta_t h_t = \mu_x + m_t + \xi_t, \tag{7.7}$$

where $\xi_t = \eta_t h_t$ is white noise, with $\sigma_{\xi\xi} = var(\xi)$, uncorrelated with $m_t = g_t - \mu_g$ given the assumptions made (Assumption G). In vector notation the last expression becomes

$$x = 1_T \mu_x + m + \xi, \tag{7.8}$$

where $x = (x_1, \ldots, x_T)', m = (m_1, \ldots, m_T)', \xi = (\xi_1, \ldots, \xi_T)', 1_a = (1, \ldots, 1)'$, with T again denoting the sample size and 1_a being $a \times 1$.

Then if M_m and M_ξ denote the covariance matrices of the vectors f and ξ and given that $M_x = M_m + M_\xi$ with M_x denoting the covariance matrix of x, the MMLSE of m is

$$\hat{m} = M_m M_x^{-1}(x - 1_T \mu_x) = (I_T - \sigma_{\xi\xi} M_x^{-1})(x - 1_T \mu_x). \tag{7.9}$$

Then, in the second stage we extract the volatility component. In fact, setting $\tilde{y}_t = (x_t - (g_t - \mu_g))^2$ we get, developing the square,

$$\tilde{y}_t = \mu_x^2 + (\eta_t h_t)^2 + 2\mu_x \eta_t h_t = \mu_x^2 + k_t + \zeta_t, \tag{7.10}$$

where again the residual $\zeta_t = 2\mu_x \eta_t h_t$ is white noise with $\sigma_{\zeta\zeta} = var(\zeta_t)$ uncorrelated with $k_t = (\eta_t h_t)^2$. Then, using the same notation, the MMSLE of k will be

$$\hat{k} = M_k M_{\tilde{y}}^{-1}(\tilde{y} - 1_T \mu_x^2) = (I_T - \sigma_{\zeta\zeta} M_{\tilde{y}}^{-1})(\tilde{y} - 1_T \mu_x^2). \tag{7.11}$$

Of course in this second stage we will subtract the estimated signal $\hat{m}$ from the raw series in first instance and thus obtain the signal for the squared adjusted series $\tilde{y}$.

In order to forecast, consider the case of the $\tilde{y}_t$ series, and for some integer s denote by N_s the $s \times T$ matrix of covariances between $\tilde{y}_s = (\tilde{y}_{T+1}, \ldots, \tilde{y}_{T+s})'$ and $\tilde{y}$. We obtain the forecast vector

$$\hat{\tilde{y}}_s = N_s M_{\tilde{y}}^{-1}(\tilde{y} - 1_T \mu_x^2) + 1_s \mu_x^2, \tag{7.12}$$

We forecast $x_s = (x_{T+1}, \ldots, x_{T+s})'$ by

$$\hat{x}_s = P_s M_x^{-1}(x - 1_T \mu_x) + 1_s \mu_x, \tag{7.13}$$

where P_s denotes the $s \times T$ matrix of covariances between x_s and x.

As noted in section 7.4 both the raw process x_t and, especially, the squared process y_t cannot be Gaussian even assuming g_t, η_t, h_t to be so, hence both the filtering and the forecasting procedures here considered will not be fully optimal by the MMS criteria.

Table 7.1: **Summary Statistics:**

Data	$Q(24)\,(p-value)$	$Q^2(24)\,(p-value)$	$Q^2(70)\,(p-value)$	kurtosis	skewness
sYP	26.34 (0.33)	50.83 (0.001)	65.23 (0.62)	8.43	1.27
fYP	26.01 (0.35)	50.85 (0.001)	65.93 (0.61)	8.34	1.25
sUP	32.07 (0.12)	102.17 (0.00)	118.94 (0.00)	5.8	0.27
fUP	31.99 (0.13)	101.16 (0.00)	117.79 (0.00)	5.69	0.25
F100	19.35 (0.73)	49.73 (0.001)	94.29 (0.02)	3.41	0.02
FAll	16.38 (0.87)	51.91 (0.001)	101.45 (0.008)	3.10	0.02
S500	33.001 (0.10)	37.33 (0.04)	83.63 (0.13)	4.98	0.07

"sYP" = spot Yen/Dollar, "fYP" = forward Yen/Pound,
"sUP" = spot Dollar/Pound, "fUP" = forward Dollar/Pound,
"F100" = FTSE 100, "FAll" = FTSE All, "S500" = S&P500 .
The forex data are weekly and run from 8^{th} Jan 1985 through 7^{th} June 1995
while the stock index data are daily and run from 1^{st} Jan 1993 through 6^{th} Feb 1995.
Columns $2-3$ report the Ljung and Box (1984) test statistic with 24 lags
for the data in the level (columns 2) and in the squares (columns 3) with p-value in parentheses.
Column 4 reports the test statistic with 70 lags for the data in the squares .
The test for the level with 70 lags is not reported because highly nonsignificant.
All the calculations of this table are based on the fast Fourier transform.

7.6 An empirical application.

We consider the same series as in section 1.1.1 but with different sampling frequency and period. The foreign exchange rate (forex) data are weekly and run from 8 Jan 1985 through 7 June 1995 while the stock index data are daily and run from 1 Jan 1993 through 6 Feb 1995. In Table 7.1 we report the Ljung and Box (1978) statistic based on the first 24 sample autocorrelations for the raw data x_t ($Q(24)$) and for the squares y_t ($Q^2(24)$) in the first and second column respectively. In the third column we report the Ljung-Box statistic based on the first 70 sample autocorrelations for the squared returns ($Q^2(70)$). In parentheses we report the p-value based on the usual χ^2 approximation. Finally in the last two columns we report the sample coefficients of kurtosis and skewness for the raw returns.

The results clearly indicate little serial correlation in the levels, but significant serial correlation in the squares. In particular, for the forex Dollar/Pound and the stock indixes, the

Table 7.2: **PMLE ('two-shock' model) : empirical applications**

Data	$\hat{\mu}_g^2$ $(t_{\mu_g^2})$	$\hat{\mu}_h^2$ $(t_{\mu_h^2})$	$\hat{\beta}$ (t_β)	$\hat{d}$ (t_d)	$\hat{\sigma}_{\epsilon\epsilon}$
sYP	exp(-13.00) (0.00)	1.22 (2.28)	0.03 (0.07)	0.313 (3.22)	0.000132
fYP	exp(-13.09) (0.00)	1.21 (2.27)	0.03 (0.06)	0.312 (3.11)	0.000133
sUP	exp(-14.89) (0.00)	1.33 (2.46)	0.11 (0.19)	0.336 (5.43)	0.000143
fUP	exp(-14.93) (0.00)	1.34 (2.50)	0.10 (0.17)	0.337 (5.38)	0.000144
F100	exp(-15.97) (0.00)	19.69 (3.83)	0.09 (0.04)	0.475 (35.03)	0.000020
FAll	exp(-16.26) (0.00)	19.13 (3.92)	0.04 (0.002)	0.474 (25.49)	0.000027
S500	exp(-17.24) (0.00)	1.00 (1.56)	0.02 (0.004)	0.476 (38.39)	0.000016

The data description is at the bottom of Table 7.1.
Each columns reports the estimate of the parameters μ_g^2, μ_h^2, β, d
with the standard Student-t in parentheses. We set $\sigma_{\eta\eta} = 1$ for identification.
All the calculations of this table are based on the fast Fourier transform.

degree of dependence in the squares appears particularly strong given the high significance of the portmanteau statistic up to the 70th lag. For all but the FTSE series the kurtosis is much greater than that for Gaussian data, and the forex Yen/Pound series show the greatest skewness.

We consider the linear parameterization reported in section 7.3 with

$$\beta_i = \beta^i, \mid \beta \mid < 1, \; i = 0, 1, \ldots,$$

$$\alpha_i = \alpha_i(d) = \begin{cases} 1, & i = 0, \\ \prod_{j=1}^{i} \frac{j+d-1}{j} & ,0 < d < 1/2, \; i = 1, 2, \ldots, \end{cases}$$

so that

$$\psi = (\mu_g^2, \mu_h^2, \beta, d, \sigma_{\epsilon\epsilon})'.$$

Hence we specify $g_t - \mu_g$ as a stationary AR(1) and $h_t - \mu_h$ as a stationary ARFIMA(0,d,0) thus obtaining a very parsimonious model. Finally we take $\sigma_{\eta\eta} = 1$ so that the mean parameters and $\sigma_{\epsilon\epsilon}$ will assume the meaning of 'variance-ratios' .

To optimize the pseudo likelihood for y_t (see section 7.4) we used the Gauss subroutine OPTMUM with the Polak-Ribiere-type option, with 50 iterations from estimates obtained by a grid search. Standard errors and thus Student-t statistics use the estimates of the trispectrum for the squared data of Taniguchi (1982) and Keenan (1987) with a Fejer window.

The results are displayed in Table 7.2, the hatted quantities indicating parameter estimates with t-ratios in parentheses.

For each raw series the estimates of the mean $\mu_x = \mu_g$ are insignificantly different from zero as are those of the AR(1) coefficient β. Things are much more interesting once we consider the estimates of the parameters of the nonlinear part of the model. In fact all the data display a strong degree of dependence in the squares, some of the d values being close to the boundary of the stationary region. For all but the $S\&P500$ index $\hat{\mu}_h^2$ is significantly different from zero, so in view of (7.3) the d estimates are directly interpretable as expressing the memory property of the squared process. For the $S\&P500$ index, taking e as the memory parameter of the squares we have $\hat{e}_{S500} = 2\hat{d}_{S500} - 1/2 = 0.452$ from the relation $2e - 1 = 4d - 2$. Then in agreement with the preliminary analysis of Table 7.1 we find that the weakest (yet significant) degree of persistence of volatility characterizes the forex yen/Pound data. Finally, we observe how the biggest estimates of the μ_h^2 parameter characterize the series with the smallest coefficient of kurtosis in agreement with the theoretical result of Corollary 6.

Appendix A

Lemmata

Lemma 27 *Under Assumption A_1, for some integer m, n with $m, n \geq 0, m+n \geq 2$, as $l \to \infty$*

$$\sum_{j=1}^{\infty} \alpha_j^m \, \alpha_{j+l}^m \sim K(l^{n(d-1)} + l^{(n+m)(d-1)+1}) \,. \tag{A.1}$$

Proof:

$$\sum_{j=1}^{\infty} \alpha_j^m \, \alpha_{j+l}^n = \sum_{j=1}^{l} \alpha_j^m \, \alpha_{j+l}^n + \sum_{j=l+1}^{\infty} \alpha_j^m \, \alpha_{j+l}^n$$

$$\sim K\,(l^{n(d-1)} \sum_{j=1}^{l} \alpha_j^m + \sum_{j=l+1}^{\infty} \alpha_j^{(m+n)}) \sim K(l^{n(d-1)} + l^{(m+n)(d-1)+1}),$$

as $l \to \infty$. $\square$

Lemma 28 *Let $\{p_j\,, j = 0, \pm 1 \ldots\}$ and $\{q_j\,, j = 0, \pm 1 \ldots\}$ be two sequences satisfying* [1], *for $r_j \in \{p_j, q_j\}$*

$$| \, r_j - r_{j+1} \, | \leq K \frac{| \, r_j \, |}{j} \tag{A.2}$$

for all $j > J$, some $0 < J < \infty$ and some $0 < K < \infty$ and

$$r_j \to 0 \;\; as \; j \to \infty \,. \tag{A.3}$$

[1] More formally, condition (A.2) is stronger than QMC, implying bounded variation as well (Yong 1974) .

Then the product sequence and the convolution sequence given by

$$m_j = p_j q_j \,,$$

and

$$n_u = \sum_{i=-\infty}^{\infty} p_i q_{i+u} \,,$$

satisfy (A.2) and (A.3).

Proof: Let us start from the product sequence m_j. Then

$$m_{j+1} - m_j = p_{j+1}(q_{j+1} - q_j) + q_j(p_{j+1} - p_j) = (i) + (ii).$$

By the assumptions made we have, for some $0 < K < \infty$, $J < \infty$ and for all $j > J$

$$| (i) | \leq K \mid p_{j+1} \mid \frac{| q_j |}{j} ,$$

$$| (ii) | \leq K \mid q_{j+1} \mid \frac{| p_j |}{j} .$$

so that we obtain, as $k \to \infty$

$$| m_k - m_{k+1} | = O(\frac{| m_k |}{k}) .$$

Considering now the convolution, as $k \to \infty$ we directly have

$$n_{k+1} - n_k = \sum_{i=-\infty}^{\infty} p_i(q_{i+k+1} - q_{i+k}) = \sum_{|i| \leq k} p_i(q_{i+k+1} - q_{i+k}) + \sum_{|i| > k} p_i(q_{i+k+1} - q_{i+k}) = (i) + (ii) .$$

Then we have as $k \to \infty$

$$| (i) | = O(| \sum_{|i| \leq k} \frac{p_i q_{i+k}}{i+k} |) = O(\frac{| q_k |}{k} \sum_{|i| \leq k} | p_i |) ,$$

and

$$| (ii) | = O(| \sum_{|i| > k} p_i(q_{i+1} - q_i) |) = O(\frac{1}{k} \sum_{|i| > k} | p_i q_i |) .$$

□

Lemma 29 *By Assumptions A_1, A_2 the following sequences are QMC:*
$\delta_{\alpha\alpha}(u)$, $\delta^2_{\alpha\alpha}(u)$, $\delta_{\alpha\alpha}(u)\alpha_{|l|}$, $\alpha^2_{|l|}$.

Proof: The result follows by the direct use of Lemma 28 in each of the sequences obtained as products or convolutions (and product of) of the α_j.

□

Lemma 30 *By Assumptions A'_1, A'_2 the following sequences are QMC :*
$\delta_{\alpha\alpha}(u;\theta)$, $\delta^2_{\alpha\alpha}(u;\theta)$, $\alpha^2_{|l|}(\theta)$, $\alpha_{|l|}\delta_{\alpha\alpha}(u;\theta)$,.
By considering Assumptions A'_3 and A'_4 as well, the result holds for the following sequence:

$$\frac{\partial \delta^2_{\alpha\alpha}(u;\theta)}{\partial \theta_i} \quad i = 1, 2, .., p.$$

Finally, by considering Assumptions $A_5(s)', A_6(s)', s = 2, \ldots, S$ as well, the result follows for

$$\frac{\partial^S \delta^2_{\alpha\alpha}(u;\theta)}{\partial \theta_{i_1} \ldots \partial \theta_{i_S}}, \quad i_j = 1, 2, .., p, \; j = 1, 2, \ldots, S.$$

Proof: The result follows by the direct use of Lemma 28 in each of the sequences obtained as products or convolutions (and product) of the $\alpha_j(\theta)$ and their derivatives with respect of θ.

□

Lemma 31 *If X and Y have zero mean, finite fourth moments and zero third cumulants,*

$$(i) cov\left((X+a)^2, (Y+b)^2\right) = cov(X^2, Y^2) + 4ab\, cov(X, Y),$$
$$(ii) cov(X^2, Y^2) = cum_4(X, X, Y, Y) + 2(cov(X, Y))^2.$$

Proof:

(i) Developing the squares on the left hand side we obtain

$$\begin{aligned}&cov(X^2+2aX+a^2,Y^2+2bY+b^2)=cov(X^2+2aX,Y^2+2bY)=\\&=cov(X^2,Y^2)+4ab\,cov(X,Y),\end{aligned}$$

the other terms being zero given the assumptions made.

(ii) We just need to take into account all the indecomposible partitions (Proposition 1) of the 2×2 array

$$\begin{array}{cc} X & X \\ Y & Y, \end{array}$$

and the result follows.

□

Lemma 32 *For W, Y, X, Z with finite variance and such that W, Y and (X, Z) are independent*

$$cov(WX, YZ) = E(W)E(Y)cov(X, Z).$$

Proof: By direct calculation.

□

Lemma 33 *For any X, Y with finite fourth moments and zero third cumulants*

$$\begin{aligned}var(XY) \quad &= cum_4(X,X,Y,Y)+var(X)var(Y)+(cov(X,Y))^2\\&+E(X)^2var(Y)+2E(X)E(Y)cov(X,Y)+E(Y)^2var(X).\end{aligned}$$

Proof: The result follows by simple use of Proposition 1, taking into consideration all the indecomposible partitions of the 2×2 array

$$\begin{array}{cc} X & Y \\ X & Y \end{array}$$

□

Appendix B

The trispectrum: basic cumulants evaluation

Now we evaluate the cumulants needed in Chapter 2 Theorem 11.

B.1 Cumulant: $cum_1(h_{t-1}^2)$

For any t by direct evaluation we get

$$cum_1(h_{t-1}^2) = \sigma^2 \delta_{(0)} + \rho^2. \tag{B.1}$$

B.2 Cumulant: $cum_2(\bar{\epsilon}_{t+s}^2, h_{t+k-1}^2)$

For any $0 \le s \le k$,

$$cum_2(\bar{\epsilon}_{t+s}^2, h_{t+k-1}^2) = \alpha_{k-s}^2 2\sigma^4 . \tag{B.2}$$

B.3 Cumulant: $cum_2(h_{t-1+a}^2, h_{t-1+b}^2)$

Given that for any $b \geq a \geq 0$

$$h_{t-1+b} = \rho + w_{b,(b-1,a)} + w_{b,(a-1,-\infty)},$$

so that

$$\begin{aligned} h_{t-1+b}^2 &= \rho^2 + w_{b,(b-1,a)}^2 + w_{b,(a-1,-\infty)}^2 \\ &+2w_{b,(b-1,a)}w_{b,(a-1,-\infty)} + 2\rho\left(w_{b,(b-1,a)} + w_{b,(a-1,-\infty)}\right), \end{aligned}$$

and

$$h_{t-1+a}^2 = \rho^2 + w_{a,(a-1,-\infty)}^2 + 2\rho w_{a,(a-1,-\infty)}.$$

Then

$$\begin{aligned} & cum_2(h_{t-1+a}^2, h_{t-1+b}^2) \\ &= (i) + (ii) \text{ where} \\ & (i) = cum_2\left(w_{b,(a-1,-\infty)}^2, w_{a,(a-1,-\infty)}^2\right), \\ & (ii) = 4\rho^2 cum_2(w_{b,(a-1,-\infty)}, w_{a,(a-1,-\infty)}), \end{aligned}$$

for the other terms being zero, so that by simple algebra and taking expectations

$$cum_2(h_{t-1+a}^2, h_{t-1+b}^2) = 2\sigma^4\delta_{(b-a)}^2 + 4\rho^2\sigma^2\delta_{(b-a)}. \tag{B.3}$$

B.4 Cumulant: $cum_3(\bar{\epsilon}_{t+a}^2, h_{t-1+b}^2, h_{t-1+c}^2)$

For any $0 \leq a \leq b \leq c$, from

$$h_{t+c-1} = \rho + w_{c,(c-1,b)} + w_{c,(b-1,a+1)} + w_{c,(a,a)} + w_{c,(a-1,-\infty)},$$

and

$$h_{t+b-1} = \rho + w_{b,(b-1,a+1)} + w_{b,(a,a)} + w_{b,(a-1,-\infty)},$$

considering only the relevant terms we obtain

$$cum_3(\bar{\epsilon}_{t+a}^2, h_{t-1+b}^2, h_{t-1+c}^2) =$$

$$
\begin{aligned}
= \; & cum_3 \Big(\bar{\epsilon}^2_{t+a}, \Big[w^2_{b,(a,a)} + w^2_{b,(a-1,-\infty)} + 2\rho(w_{b,(a,a)} + w_{b,(a-1,-\infty)}) \\
& + 2w_{b,(b-1,a+1)}(w_{b,(a,a)} + w_{b,(a-1,-\infty)})) + 2w_{b,(a,a)}w_{b,(a-1,-\infty)}\Big], \\
& , \Big[w^2_{c,(a,a)} + w^2_{c,(a-1,-\infty)} + 2\rho(w_{c,(a,a)} + w_{c,(a-1,-\infty)}) + \\
& + 2w_{c,(c-1,b)}(w_{c,(a,a)} + w_{c,(a-1,-\infty)}) + 2w_{c,(b-1,a+1)}(w_{c,(a,a)} + w_{c,(a-1,-\infty)}) \\
& + 2w_{c,(a,a)}w_{c,(a-1,-\infty)}\Big]\Big) .
\end{aligned}
$$

Then

$$
\begin{aligned}
& cum_3(\bar{\epsilon}^2_{t+a}, w^2_{b,(a,a)}, w^2_{c,(a,a)}) = \alpha^2_{c-a}\alpha^2_{b-a}8\sigma^6 , \\
& cum_3(\bar{\epsilon}^2_{t+a}, 2\rho w_{c,(a,a)}, 2\rho w_{b,(a,a)})) = 8\rho^2\alpha_{c-a}\alpha_{b-a}\sigma^4 , \\
& cum_3(\bar{\epsilon}^2_{t+a}, 2w_{c,(b-1,a+1)}w_{c,(a,a)}, 2w_{b,(b-1,a+1)}w_{b,(a,a)}) \\
& = 8\sigma^6\alpha_{b-a}\alpha_{c-a}\sum_{i=1}^{b-a-1}\alpha_i\alpha_{i+c-b} , \\
& cum_3(\bar{\epsilon}^2_{t+a}, 2w_{c,(a,a)}w_{c,(a-1,-\infty)}, 2w_{b,(a,a)}w_{b,(a-1,-\infty)}) \\
& = 8\sigma^6\alpha_{b-a}\alpha_{c-a}\sum_{i=1}^{\infty}\alpha_{i+b-a}\alpha_{i+c-a} .
\end{aligned}
$$

Finally

$$
cum_3(\bar{\epsilon}^2_{t+a}, h^2_{t-1+b}, h^2_{t-1+c}) = 8\sigma^6\alpha_{c-a}\alpha_{b-a}\delta_{(c-b)} + 8\rho^2\sigma^4\alpha_{c-a}\alpha_{b-a} . \tag{B.4}
$$

B.5 Cumulant: $cum_3(h^2_{t-1+a}, h^2_{t-1+b}, h^2_{t-1+c})$

Again taking only the relevant terms yields

$$
\begin{aligned}
& cum_3 \Big(w^2_{a,(a-1,-\infty)} + 2\rho w_{a,(a-1,-\infty)}, \\
& \Big[w^2_{a,(a-1,-\infty)} + 2w_{b,(b-1,a+1)}w_{b,(a-1,-\infty)} + 2w_{b,(a,a)}w_{b,(a-1,-\infty)} + 2\rho w_{b,(a-1,-\infty)}\Big], \\
& \Big[w^2_{c,(a-1,-\infty)} + 2\rho w_{c,(a-1,-\infty)} + 2w_{c,(b-1,a+1)}w_{c,(a-1,-\infty)} + 2w_{c,(a,a)}w_{c,(a-1,-\infty)}\Big]\Big) .
\end{aligned}
$$

Then

$$
\begin{aligned}
& cum_3(w^2_{a,(a-1,-\infty)}, w^2_{b,(a-1,-\infty)}, w^2_{c,(a-1,-\infty)}) \\
& = 8\sigma^6\delta_{(b-a)}\delta_{(c-a)}\sum_{i=1}^{\infty}\alpha_{i+c-a}\alpha_{i+b-a} .
\end{aligned}
$$

$$cum_3(w^2_{a,(a-1,-\infty)}, 2w_{b,(b-1,a+1)}w_{b,(a-1,-\infty)}, 2w_{c,(b-1,a+1)}w_{c,(a-1,-\infty)})$$
$$= 8\sigma^6(\sum_{i=1}^{b-a} \alpha_i\alpha_{c-b+i})\delta_{(b-a)}\delta_{(c-a)}.$$

$$cum_3(2\rho w_{a,(a-1,-\infty)}, w^2_{b,(a-1,-\infty)}, 2\rho w_{c,(a-1,-\infty)})$$
$$= 8\rho^2\sigma^4\delta_{(b-a)}(\sum_{i=1}^{\infty} \alpha_{i+c-a}\alpha_{i+b-a}).$$

$$cum_3(2\rho w_{a,(a-1,-\infty)}, 2\rho w_{b,(a-1,-\infty)}, w^2_{c,(a-1,-\infty)})$$
$$= \quad 4\rho^2\sigma^4\delta_{(c-a)}(\sum_{i=1}^{\infty} \alpha_{i+b-a}\alpha_{i+b-a}).$$

$$cum_3(w^2_{a,(a-1,-\infty)}, 2\rho w_{b,(a-1,-\infty)}, 2\rho w_{c,(a-1,-\infty)}) =$$
$$8\rho^2\sigma^4\delta_{(b-a)}\delta_{(c-a)}.$$

$$cum_3(2\rho w_{a,(a-1,-\infty)}, 2\rho w_{b,(b-1,a)}, 2w_{c,(a-1,-\infty)}w_{c,(b-1,a)}) =$$
$$8\rho^2\sigma^4\delta_{(c-a)}(\sum_{i=1}^{b-a} \alpha_i\alpha_{i+c-b}).$$

$$cum_3(2\rho w_{a,(a-1,-\infty)}, 2w_{c,(a-1,-\infty)}w_{c,(b-1,a)}, 2\rho w_{b,(b-1,a)})$$
$$= 8\rho^2\sigma^4\delta_{(b-a)}(\sum_{i=1}^{b-a} \alpha_i\alpha_{i+c-b}).$$

Summing up we obtain

$$cum_3(h^2_{t-1+a}, h^2_{t-1+b}, h^2_{t-1+c}) = 8\sigma^6\delta_{(b-a)}\delta_{(c-a)}\delta_{(c-b)}$$
$$+8\sigma^4\rho^2\left[\delta_{(b-a)}\delta_{(c-a)} + \delta_{(b-a)}\delta_{(c-b)} + \delta_{(c-a)}\delta_{(c-b)}\right]. \tag{B.5}$$

B.6 Cumulant: $cum_4(\bar{\epsilon}^2_t, h^2_{t-1+a}, h^2_{t-1+b}, h^2_{t-1+c})$

The relevant terms are given by

$$cum_4(\bar{\epsilon}^2_t, w^2_{a,(a-1,1)}, 2w_{b,(a,1)}w_{b,(0,0)}, 2w_{c,(a,1)}w_{c,(0,0)})$$
$$= 16\sigma^8\alpha_{b-a}\alpha_{c-a}(\sum_{i=1}^{a-1} \alpha_i\alpha_{b-a+i}\sum_{i=1}^{a-1} \alpha_i\alpha_{c-a+i}).$$
$$cum_4(\bar{\epsilon}^2_t, w^2_{a,(0,0)}, w^2_{b,(0,0)}, w^2_{c,(0,0)}) = 48\sigma^8\alpha_a^2\alpha_b^2\alpha_c^2.$$
$$cum_4(\bar{\epsilon}^2_t, w^2_{a,(0,0)}, 2\rho w_{b,(0,0)}, 2\rho w_{c,(0,0)}) = 32\rho^2\sigma^6\alpha_a^2\alpha_b\alpha_c.$$

$$cum_4(\bar{\epsilon}_t^2, w_{a,(0,0)}^2, 2w_{b,(b-1,a)}w_{b,(0,0)}, 2w_{c,(b-1,a)}w_{c,(0,0)})$$
$$= 32\sigma^8\alpha_a^2\alpha_b\alpha_c \sum_{i=1}^{b-a} \alpha_i\alpha_{c-b+i}.$$

$$cum_4(\bar{\epsilon}_t^2, w_{a,(0,0)}^2, 2w_{b,(a+1,1)}w_{b,(0,0)}, 2w_{c,(a+1,1)}w_{c,(0,0)})$$
$$= 32\sigma^8\alpha_a^2\alpha_b\alpha_c \sum_{i=1}^{a-1} \alpha_{i+b-a}\alpha_{i+c-a}$$

$$cum_4(\bar{\epsilon}_t^2, w_{a,(0,0)}^2, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, 2w_{c,(0,0)}w_{c,(0-1,-\infty)})$$
$$= 32\sigma^8\alpha_a^2\alpha_b\alpha_c \sum_{i=1}^{\infty} \alpha_{i+b}\alpha_{c+i}.$$

$$cum_4(\bar{\epsilon}_t^2, w_{a,(-1,-\infty)}^2, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, 2w_{c,(0,0)}w_{c,(0-1,-\infty)})$$
$$= 16\sigma^8\alpha_b\alpha_c \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{c+i} \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{b+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, w_{b,(0,0)}^2, 2\rho w_{c,(0,0)}) = 32\rho^2\alpha_a\alpha_b^2\alpha_c\sigma^6.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2\rho w_{b,(0,0)}, w_{c,(0,0)}^2) = 32\rho^2\alpha_a\alpha_b\alpha_c^2\sigma^6.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(a-1,1)}, 2w_{b,(0,0)}w_{b,(a+1,1)}, 2\rho w_{c,(0,0)})$$
$$= 16\rho^2\sigma^6\alpha_b\alpha_c \sum_{i=1}^{a-1} \alpha_i\alpha_{i+b-a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(a-1,1)}, 2\rho w_{b,(0,0)}, 2w_{c,(0,0)}w_{c,(a+1,1)})$$
$$= 16\rho^2\sigma^6\alpha_b\alpha_c \sum_{i=1}^{a-1} \alpha_i\alpha_{i+c-a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2w_{b,(0,0)}w_{b,(b-1,a)}, 2\rho w_{c,(b-1,a)})$$
$$= 16\rho^2\alpha_a\alpha_b\sigma^6 \sum_{i=1}^{b-a} \alpha_i\alpha_{c-b+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2\rho w_{b,(b-1,a)}, 2w_{c,(b-1,a)}w_{c,(0,0)})$$
$$= 16\rho^2\alpha_a\alpha_c\sigma^6 \sum_{i=1}^{b-a} \alpha_i\alpha_{c-b+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2\rho w_{b,(a+1,1)}, 2w_{c,(0,0)}w_{c,(a+1,1)})$$
$$= 16\rho^2\alpha_a\alpha_c\sigma^6\sum_{i=1}^{a-1}\alpha_{i+b-a}\alpha_{c-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2w_{b,(a+1,1)}w_{b,(0,0)}, 2\rho w_{c,(a+1,1)})$$
$$= 16\rho^2\alpha_a\alpha_b\sigma^6\sum_{i=1}^{a-1}\alpha_{i+b-a}\alpha_{c-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, 2\rho w_{c,(-1,-\infty)})$$
$$= 16\rho^2\alpha_a\alpha_b\sigma^6\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{c+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(0,0)}, 2\rho w_{b,(-1,-\infty)}, 2w_{c,(0,0)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\alpha_a\alpha_c\sigma^6\sum_{i=1}^{\infty}\alpha_{b+i}\alpha_{c+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(-1,-\infty)}, 2\rho w_{b,(0,0)}, 2w_{c,(0,0)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\alpha_b\alpha_c\sigma^6\sum_{i=1}^{\infty}\alpha_{a+i}\alpha_{c+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2\rho w_{a,(-1,-\infty)}, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, 2\rho w_{c,(0,0)})$$
$$= 16\rho^2\alpha_b\alpha_c\sigma^6\sum_{i=1}^{\infty}\alpha_{a+i}\alpha_{b+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(0,0)}, 2\rho w_{b,(0,0)}, 2\rho w_{c,(a-1,1)})$$
$$= 16\rho^2\alpha_a\alpha_b\sigma^6\sum_{i=1}^{a-1}\alpha_i\alpha_{c-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(0,0)}, 2\rho w_{b,(a-1,1)}, 2\rho w_{c,(0,0)})$$
$$= 16\rho^2\alpha_a\alpha_c\sigma^6\sum_{i=1}^{a-1}\alpha_i\alpha_{b-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(0,0)}, w_{b,(0,0)}^2, 2w_{c,(a-1,1)}w_{c,(0,0)})$$
$$= 32\sigma^8\alpha_a\alpha_b^2\alpha_c\sum_{i=1}^{a-1}\alpha_i\alpha_{c-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(0,0)}, 2w_{b,(a-1,1)}w_{b,(0,0)}, w_{c,(0,0)}^2)$$
$$= 32\sigma^8\alpha_a\alpha_b\alpha_c^2\sum_{i=1}^{a-1}\alpha_i\alpha_{c-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(0,0)}, 2w_{b,(0,0)}w_{b,(a-1,1)}, w_{c,(a-1,1)}^2) =$$
$$= 32\sigma^8\alpha_a\alpha_b\sum_{i=1}^{a-1}\alpha_i\alpha_{c-a+i}\sum_{i=1}^{a-1}\alpha_i\alpha_{b-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(0,0)}, w_{b,(a-1,1)}^2, 2w_{c,(a-1,1)}w_{c,(0,0)})$$
$$= 32\sigma^8\alpha_a\alpha_c\sum_{i=1}^{a-1}\alpha_{i+b-a}\alpha_{c-a+i}\sum_{i=1}^{a-1}\alpha_i\alpha_{b-a+i}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(-1,-\infty)}, 2w_{b,(0,0)}w_{b,(a-1,1)}, 2w_{c,(0,0)}w_{c,(-1,-\infty)})$$
$$= 16\sigma^8\alpha_b\alpha_c\sum_{i=1}^{a-1}\alpha_i\alpha_{b-a+i}\sum_{i=1}^{\infty}\alpha_{i+c}\alpha_{i+a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(a-1,1)}w_{a,(-1,-\infty)}, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, 2w_{c,(0,0)}w_{c,(a-1,1)})$$
$$= 16\sigma^8\alpha_b\alpha_c\sum_{i=1}^{a-1}\alpha_i\alpha_{c-a+i}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(-1,-\infty)}, 2\rho w_{b,(-1,-\infty)}, 2\rho w_{c,(0,0)})$$
$$= 16\rho^2\sigma^6\alpha_a\alpha_c\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(-1,-\infty)}, 2\rho w_{b,(0,0)}, 2\rho w_{c,(-1,-\infty)})$$
$$= \quad 16\rho^2\sigma^6\alpha_a\alpha_b\sum_{i=1}^{\infty}\alpha_{i+c}\alpha_{i+a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(-1,-\infty)}w_{a,(0,0)}, 2w_{b,(-1,-\infty)}w_{b,(0,0)}, w_{c,(0,0)}^2)$$
$$= 32\alpha_a\alpha_b\alpha_c^2\sigma^8 \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+b}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(-1,-\infty)}w_{a,(0,0)}, w_{b,(0,0)}^2, 2w_{c,(-1,-\infty)}w_{c,(0,0)})$$
$$= 32\alpha_a\alpha_b^2\alpha_c\sigma^8 \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+c}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(-1,-\infty)}w_{a,(0,0)}, w_{b,(-1,-\infty)}^2, 2w_{c,(-1,-\infty)}w_{c,(0,0)})$$
$$= 32\alpha_a\alpha_c\sigma^8 \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+b} \sum_{i=1}^{\infty} \alpha_{i+b}\alpha_{i+c}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(-1,-\infty)}w_{a,(0,0)}, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, w_{c,(-1,-\infty)}^2)$$
$$= 32\alpha_a\alpha_b\sigma^8 \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+c} \sum_{i=1}^{\infty} \alpha_{i+b}\alpha_{i+c}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(-1,-\infty)}, 2w_{b,(-1,-\infty)}w_{b,(0,0)}, 2w_{c,(a,-\infty)}w_{c,(-1,-\infty)})$$
$$= 16\sigma^8\alpha_a\alpha_b \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+c} \sum_{i=1}^{a-1} \alpha_{i+b-a}\alpha_{i+c-a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(-1,-\infty)}, 2w_{b,(b,a+1)}w_{b,(0,0)}, w_{c,(b,a+1)}w_{c(-1,-\infty)})$$
$$= 16\sigma^8\alpha_a\alpha_b \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+c} \sum_{i=1}^{b-a} \alpha_i\alpha_{i+c-b}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(a-1,1)}, 2w_{b,(0,0)}w_{b,(-1,-\infty)}, 2w_{c,(a-1,1)}w_{c,(-1,-\infty)}$$
$$= 16\sigma^8\alpha_a\alpha_b \sum_{i=1}^{\infty} \alpha_{i+b}\alpha_{i+c} \sum_{i=1}^{a-1} \alpha_i\alpha_{i+c-a}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(a-1,1)}, 2w_{b,(a-1,1)}w_{b,(-1,-\infty)}, 2w_{c,(0,0)}w_{c,(-1,-\infty)})$$
$$= 16\sigma^8\alpha_a\alpha_c \sum_{i=1}^{a-1} \alpha_i\alpha_{i+b-a} \sum_{i=1}^{\infty} \alpha_{i+b}\alpha_{i+c}.$$

$$cum_4(\bar{\epsilon}_t^2, 2w_{a,(0,0)}w_{a,(-1,-\infty)}, 2w_{b,(a-1,1)}w_{b,(-1,-\infty)}, 2w_{c,(0,0)}w_{c,(a-1,1)})$$
$$= 16\sigma^8\alpha_a\alpha_c \sum_{i=1}^{a-1} \alpha_{i+b-a}\alpha_{i+c-a} \sum_{i=1}^{\infty} \alpha_{i+a}\alpha_{i+b}.$$

Thus by summing up and using simple algebraic steps one obtains

$$
\begin{aligned}
& cum_4(\bar{\epsilon}_t^2, h_{t+a-1}^2, h_{t+b-1}^2, h_{t+c-1}^2) = 16\sigma^8\alpha_b\alpha_c\delta_{(b-a)}\delta_{(c-a)} \\
& +16\sigma^8\alpha_a\alpha_b\delta_{(c-a)}\delta_{(c-b)} + 16\sigma^8\alpha_a\alpha_c\delta_{(b-a)}\delta_{(c-b)} \\
& +16\rho^2\sigma^4\alpha_b\alpha_c(\delta_{(b-a)} + \delta_{(c-a)}) + 16\rho^2\sigma^4\alpha_a\alpha_b(\delta_{(c-b)} + \delta_{(c-a)}) \\
& +16\rho^2\sigma^4\alpha_a\alpha_c(\delta_{(c-b)} + \delta_{(b-a)}).
\end{aligned}
$$

B.7 Cumulant: $cum_4(h_{t-1}^2, h_{t+a-1}^2, h_{t+b-1}^2, h_{t+c-1}^2)$

With respect of the latter case we have approximately the double of the terms due to the fact that the first argument in the cumulant expression is given by

$$
h_{t-1}^2 = \rho^2 + w_{0,(-1,-\infty)}^2 + 2\rho w_{0,(-1,-\infty)}. \tag{B.6}
$$

We can then write for notational convenience

$$
cum_4(h_{t-1}^2, h_{t+a-1}^2, h_{t+b-1}^2, h_{t+c-1}^2) = (i) + (ii), \tag{B.7}
$$

where

$$
\begin{aligned}
(i) &= cum_4(w_{0,(-1,-\infty)}^2, h_{t+a-1}^2, h_{t+b-1}^2, h_{t+c-1}^2), \\
(ii) &= cum_4(2\rho w_{0,(-1,-\infty)}, h_{t+a-1}^2, h_{t+b-1}^2, h_{t+c-1}^2).
\end{aligned}
$$

Then, with respect to (i), where $w_{0,(-1,-\infty)}^2$ appears as the first argument in the cumulant expression, one obtains

$$
\begin{aligned}
& cum_4(w_{0,(-1,-\infty)}^2, w_{a,(a-1,0)}^2, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)}, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)}) \\
= \; & 16\sigma^8\delta_{(b)}\delta_{(c)}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a}.
\end{aligned}
$$

$$
\begin{aligned}
& cum_4(w_{0,(-1,-\infty)}^2, w_{a,(a-1,0)}^2, 2\rho w_{b,(-1,-\infty)}, 2\rho w_{c,(-1,-\infty)}) \\
= \; & 16\sigma^6\rho^2(\delta_{(a)}\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c} + \delta_{(a)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, w^2_{b,(-1,-\infty)}, w^2_{c,(-1,-\infty)}) \\
&= 16\sigma^8\delta_{(a)}\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c} \\
&+16\sigma^8\delta_{(a)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c} \\
&+16\sigma^8\delta_{(b)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}.
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2w_{b,(-1,-\infty)}w_{b,(a-1,0)}, 2w_{c,(-1,-\infty)}w_{c,(a-1,0)}) \\
&= 16\sigma^8\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}(\delta_{(a)}\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}+\delta_{(a)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2w_{b,(-1,-\infty)}w_{b,(b-1,a)}, 2w_{c,(-1,-\infty)}w_{c,(b-1,a)}) \\
&= 16\sigma^8(\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b})(\delta_{(a)}\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}+\delta_{(a)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, 2\rho w_{a,(a-1,0)}, 2w_{b,(-1,-\infty)}w_{b,(a-1,0)}, 2\rho w_{c,(-1,-\infty)}) \\
&= 16\rho^2\sigma^6\delta_{(b-a)}(\delta_{(b)}\delta_{(c)}).
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, 2\rho w_{a,(a-1,0)}, 2\rho w_{b,(-1,-\infty)}, 2w_{c,(a-1,0)}w_{+c,(-1,-\infty)}) \\
&= 16\rho^2\sigma^6\delta_{(b)}\delta_{(c)}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}.
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, w^2_{b,(-1,-\infty)}, 2\rho w_{c,(-1,-\infty)} \\
&= 16\rho^2\sigma^6(\delta_{(a)}\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c}+\delta_{(b)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(0-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2\rho w_{b,(-1,-\infty)}, w^2_{c,(-1,-\infty)}) \\
&= 16\rho^2\sigma^6(\delta_{(a)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{b+i}\alpha_{c+i}+\delta_{(b)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{a+i}\alpha_{c+i}).
\end{aligned}
$$

$$
\begin{aligned}
&cum_4(w^2_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2w_{b,(a+1,0)}w_{b,(-1,-\infty)}, 2\rho w_{c,(a+1,0)}) \\
&= 16\rho^2\sigma^6\delta_{(a)}\delta_{(b)}\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}.
\end{aligned}
$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2\rho w_{b,(a+1,0)}, 2w_{c,(a+1,0)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\delta_{(a)}\delta_{(c)}\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}.$$

$$cum_4(w^2_{0,(0-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2w_{b,(b-1,a)}w_{b,(-1,-\infty)}, 2\rho w_{c,(b-1,a)})$$
$$= 16\rho^2\sigma^6\delta_{(a)}\delta_{(b)}\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b}.$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2\rho w_{b,(b-1,a)}, 2w_{c,(b-1,a)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\delta_{(a)}\delta_{(c)}\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b}.$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a+1,0)}w_{a,(-1,-\infty)}, 2\rho w_{b,(-1,-\infty)}, 2\rho w_{c,(a+1,0)})$$
$$= 16\rho^2\sigma^6\delta_{(a)}\delta_{(b)}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}.$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a-1,0)}w_{a,(-1,-\infty)}, w^2_{b,(a-1,0)}, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)})$$
$$= 16\sigma^8\delta_{(a)}\delta_{(c)}\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a}\sum_{i=1}^{a}\alpha_{i+c-a}\alpha_{i+b-a}.$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a+1,0)}w_{a,(-1,-\infty)}, 2\rho w_{b,(a+1,0)}, 2\rho w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\delta_{(c)}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}\sum_{i=1}^{\infty}\alpha_i\alpha_{i+a}.$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a-1,0)}w_{a,(-1,-\infty)}, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)}, w^2_{c,(a-1,0)})$$
$$= 16\sigma^8\delta_{(a)}\delta_{(b)}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}.$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a-1,1)}w_{a,(-1,-\infty)}, 2w_{b,(b-1,a)}w_{b,(a-1,1)}, 2w_{c,(b-1,a)}w_{c,(-1,-\infty)})$$
$$= 16\sigma^8\delta_{(a)}\delta_{(c)}(\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a})(\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b}).$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a-1,1)}w_{a,(-1,-\infty)}, 2w_{b,(b-1,a)}w_{b,(-1,-\infty)}, 2w_{c,(b-1,a)}w_{c,(a-1,1)})$$
$$= 16\sigma^8\delta_{(a)}\delta_{(b)}(\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a})(\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b}).$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a-1,1)}w_{a,(-1,-\infty)}, w^2_{b,(-1,-\infty)}, 2w_{c,(a-1,1)}w_{c,(-1,-\infty)})$$
$$= 16\sigma^8(\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a})(\delta_{(a)}\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c} + \delta_{(b)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

$$cum_4(w^2_{0,(-1,-\infty)}, 2w_{a,(a-1,1)}w_{a,(-1,-\infty)}, 2w_{b,(a-1,1)}w_{b,(-1,-\infty)}, w^2_{c,(-1,-\infty)})$$
$$= 16\sigma^8(\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a})(\delta_{(a)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c} + \delta_{(b)}\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}).$$

Thus, after simple but tedious algebra, one obtains

$$(i) = 16\sigma^8\left(\delta_{(a)}\delta_{(b)}\delta_{(c-b)}\delta_{(c-a)} + \delta_{(a)}\delta_{(c)}\delta_{(b-a)}\delta_{(c-b)} + \delta_{(b)}\delta_{(c)}\delta_{(b-a)}\delta_{(c-a)}\right)$$
$$+16\rho^2\sigma^6\left(\delta_{(a)}\delta_{(b)}(\delta_{(c-b)} + \delta_{(b-a)}) + \sigma^6\delta_{(a)}\delta_{(c)}(\delta_{(c-b)} + \delta_{(c-a)}) + \sigma^6\delta_{(b)}\delta_{(c)}(\delta_{(b-a)} + \delta_{(c-a)})\right).$$

With respect to (ii), where $2\rho w_{0,(-1,-\infty)}$ appears as the first argument in the cumulant expression, one obtains

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(a-1,0)}, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)}, 2\rho w_{c,(a-1,0)})$$
$$= 16\rho^2\sigma^6\delta_{(b)}\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}.$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(a-1,0)}, 2\rho w_{b,(a-1,0)}, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\delta_{(c)}\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a}\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a}.$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2\rho w_{b,(-1,-\infty)}, w^2_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\left[\delta_{(a)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c})(\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c}) + \delta_{(c)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b})(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c})\right]$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, w^2_{b,(-1,-\infty)}, 2\rho w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\left[\delta_{(a)}(\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c})(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}) + \delta_{(b)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b})(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c})\right]$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2\rho w_{b,(a-1,1)}, 2w_{c,(-1,-\infty)}w_{c,(a-1,1)})$$
$$= 16\rho^2\sigma^4(\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a})\delta_{(a)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2w_{b,(-1,-\infty)}w_{b,(a-1,1)}, 2\rho w_{c,(a-1,1)})$$
$$= 16\rho^2\sigma^4(\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a})\delta_{(a)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2w_{b,(-1,-\infty)}w_{b,(b-1,a)}, 2\rho w_{c,(b-1,a)})$$
$$= 16\rho^2\sigma^4(\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b})\delta_{(a)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, w^2_{a,(-1,-\infty)}, 2\rho w_{b,(b-1,a)}, 2w_{c,(-1,-\infty)}w_{c,(b-1,a)})$$
$$= 6\rho^2\sigma^4(\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b})\delta_{(a)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2\rho w_{a,(a-1,0)}, w^2_{b,(a-1,0)}, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6\delta_{(c)}(\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a})(\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2\rho w_{a,(a-1,0)}, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)}, w^2_{c,(a-1,0)})$$
$$= 16\rho^2\sigma^6\delta_{(b)}(\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a})(\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}).$$

$$cum_4(2\rho w_{0,(0-1,-\infty)}, 2\rho w_{a,(a-1,0)}, w^2_{b,(-1,-\infty)}, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)}) =$$
$$= \quad 16\rho^2\sigma^6\delta_{(b)}(\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a})(\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2\rho w_{a,(a-1,0)}, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)}, w^2_{c,(-1,-\infty)}) =$$
$$= \quad 16\rho^2\sigma^6\delta_{(c)}(\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a})(\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)},$$
$$, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)}) =$$
$$= \quad 16\rho^2\sigma^6\sum_{i=1}^{a}\alpha_{i+b-a}\alpha_{i+c-a}(\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c} + \delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, w^2_{b,(-1,-\infty)}, w^2_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6(\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c})(\delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b} + \delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2w_{a,(a+1,0)}w_{a,(-1,-\infty)}, 2\rho w_{b,(-1,-\infty)}, 2w_{c,(a-1,0)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6(\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a})(\delta_{(a)}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c} + \delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2w_{a,(a+1,0)}w_{a,(-1,-\infty)}, 2w_{b,(a-1,0)}w_{b,(-1,-\infty)}, 2\rho w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6(\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a})(\delta_{(a)}\sum_{i=1}^{\infty}\alpha_{i+b}\alpha_{i+c} + \delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2w_{a,(a-1,0)}w_{a,(-1,-\infty)}, 2\rho w_{b,(a-1,0)}, w^2_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^6(\sum_{i=1}^{a}\alpha_i\alpha_{i+b-a})\delta_{(c)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2w_{a,(-1,0)}w_{a,(-1,-\infty)}, w^2_{b,(-1,-\infty)}, 2\rho w_{c,(a-1,0)})$$
$$= 16\rho^2\sigma^6(\sum_{i=1}^{a}\alpha_i\alpha_{i+c-a})(\delta_{(b)}(\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

$$cum_4(2\rho w_{0,(-1,-\infty)}, 2\rho w_{a,(-1,-\infty)}, 2w_{b,(b-1,a)}w_{b,(-1,-\infty)}, , 2w_{c,(b-1,a)}w_{c,(-1,-\infty)})$$
$$= 16\rho^2\sigma^4(\sum_{i=1}^{b-a}\alpha_i\alpha_{i+c-b})(\delta_{(b)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+c} + \delta_{(c)}\sum_{i=1}^{\infty}\alpha_{i+a}\alpha_{i+b}).$$

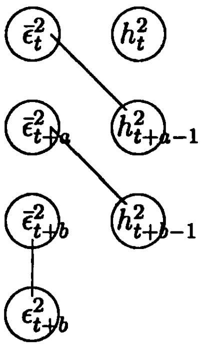
Thus

$$(ii) = 16\rho^2\sigma^4\delta_{(b)}\left[\delta_{(c-a)}\delta_{(c-b)} + \delta_{(c-a)}\delta_{(b-a)}\right] + 16\rho^2\sigma^4\delta_{(c)}\left[\delta_{(c-a)}\delta_{(b-a)} + \delta_{(c-b)}\delta_{(b-a)}\right]$$
$$+16\rho^2\sigma^4\delta_{(a)}\left[\delta_{(c-b)}\delta_{(b-a)} + \delta_{(c-b)}\delta_{(c-a)}\right].$$

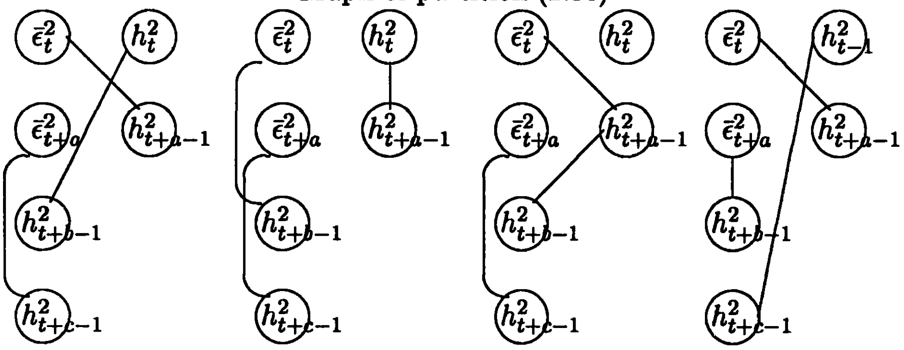
Appendix C

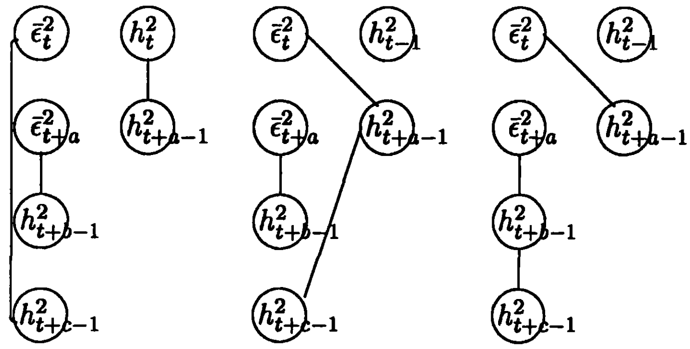
The trispectrum: graphs of the partitions

Graph of partition (2.34)

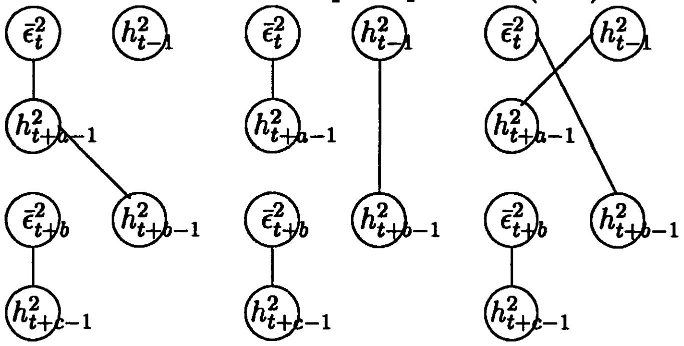


Graph of partition (2.36)

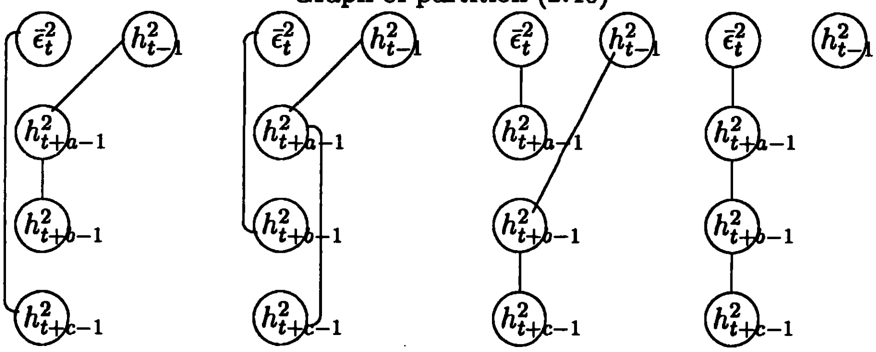


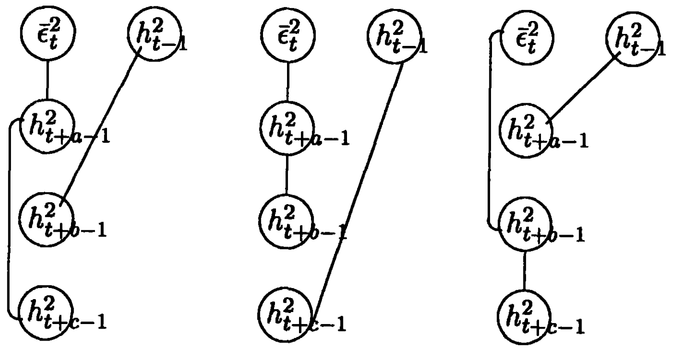


Graph of partition (2.38)

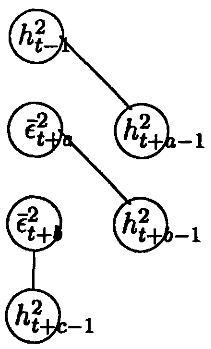


Graph of partition (2.40)

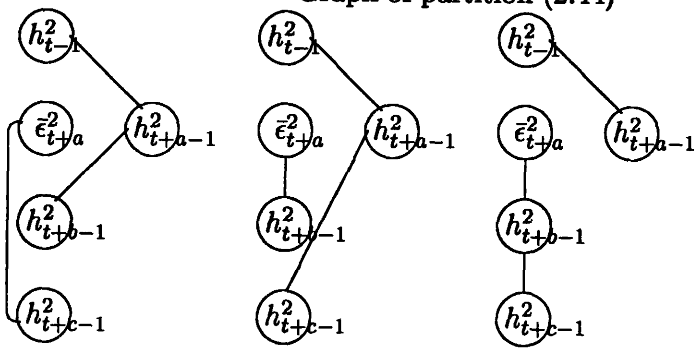




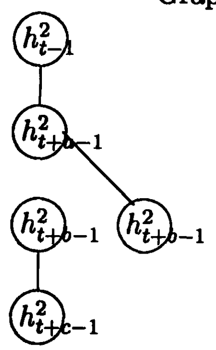
Graph of partition (2.42)



Graph of partition (2.44)



Graph of partition (2.46)



Bibliography

ADENSTEDT, R. (1974): "On large sample estimation for the mean of a stationary random sequence," *The Annals of Statistics*, 6/2, 1095–1107.

ANDERSEN, T. (1996): "Return volatility and trading volume: an information flow interpretation of stochastic volatility," *Journal of Finance*, 51/1, 169–204.

AROIAN, L. (1947): "The probability function of the product of two normally distributed variables," *Annals of Mathematical Statistics*, 18, 265–271.

BAILLIE, R. T., T. BOLLERSLEV, AND H. O. A. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74(1), 3–530.

BARNDORFF-NIELSEN, O., AND D. COX (1989): *Asymptotic techniques for use in statistics.* London: Chapman and Hall.

BEKAERT, G. (1996): "The time variation of risk and return in foreign exchange markets: a general equilibrium perspective," *Review of Financial Studies*, 9(2), 427–470.

BENTKUS, R. (1976): "Cumulants of estimates of the spectrum of a stationary time series," *Lietuvos Matematikos Rinkinys*, 16, 37–61.

BERAN, J. (1994): *Statistics for Long Memory Processes.* New York : Chapman and Hall.

BLACK, F. (1976): "Studies of stock market volatility changes," in *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

BOLLERLSEV, T., R. CHOU, AND K. KRONER (1992): "ARCH modelling in finance: a review of the theory and empirical evidence," *Journal of Econometrics*, 52, 5–59.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 302–327.

BOLLERSLEV, T. (1987): "A conditional heteroscedastic time series model for speculative prices and rates of return," *Review of Economics and Statistics*, 69, 542–547.

BOLLERSLEV, T., R. ENGLE, AND D. NELSON (1995): "ARCH models," in *Handbook of Econometrics vol IV*, ed. by R. Engle,and D. McFadden. Amsterdam : North Holland.

BREIDT, F., N. CRATO, AND DELIMA P. (1993): "Modelling long-memory stochastic volatility," Working paper, Iowa State University.

BRILLINGER, D. (1975): *Time series, data analysis and theory.* New York : Holt Rinehart and Winston.

BRILLINGER, D. R., AND M. ROSENBLATT (1966): "Asymptotic theory of higher-order spectra," in *Spectral Analysis of Time Series*, ed. by B. Harris, pp. 189–232. New York: Wiley.

BROCKWELL, P., AND R. DAVIS (1987): *Time series: theory and methods.* New York: Springer Verlag.

BROWN, B. (1971): "Martingale central limit theorem," *The Annals of Mathematical Statistics*, 35, 1296–1303.

CAMPBELL, J., S. GROSSMAN, AND J. WANG (1993): "Trading volume and serial correlation in stock returns," *Quarterly Journal of Economics*, 1, 905–939.

CAMPBELL, J., AND L. HENTSCHEL (1990): "No news is good news: an asymmetric model of changing volatility in stock returns," Unpublished paper, Princeton University.

CHENG, B., AND P. M. ROBINSON (1994): "Semiparametric estimation from time series with long-range dependence," *Journal of Econometrics*, 64, 335–353.

CHESNEY, M., AND L. SCOTT (1989): "Pricing European options: a comparison of the modified Black-Scholes model and a random variance model," *Journal of Financial and Quantitative Analysis*, 24, 267–284.

CHOW, Y., AND H. TEICHER (1978): *Probability theory.* New York: Springer Verlag.

CLARK, P. (1973): "A subordinate stochastic process model with fixed variance for speculative prices," *Econometrica*, 41, 135–156.

CORRADI, V. (1995): "Nonlinear transformations of integrated time series : a reconsideration," *Journal of Time Series Analysis*, 16/6, 539–550.

DAHLHAUS, R. (1983): "Spectral analysis with tapered data," *Journal of Time Series Analysis*, 4/3, 163–175.

DAHLHAUS, R. (1989): "Efficient parameter estimation for self-similar processes," *The Annals of Statistics*, 17, 1749–1782.

DAVIES, R., AND D. HARTE (1987): "Tests for Hurst effect," *Biometrika*, 74, 95–102.

DUFFIE, D., AND K. SINGLETON (1993): "Simulated moments estimation of Markov models of asset prices," *Econometrica*, 61, 929–952.

DUNSMUIR, W. (1979): "A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed white noise," *The Annals of Statistics*, 7, 490–506.

ENGLE, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom," *Econometrica*, 50, 987–1007.

FOX, R., AND M. TAQQU (1985): "Non-central limit theorems for quadratic forms in random variables having long-range dependence," *The Annals of Statistics*, 13, 428–446.

FOX, R., AND M. TAQQU (1986): "Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *The Annals of Statistics*, 14/2, 517–532.

FOX, R., AND M. TAQQU (1987): "Central limit theorems for quadratic forms in random variables having long-range dependence," *Probability Theory and Related Fields*, 74, 213–240.

GALLANT, A., D. HSIEH, AND G. TAUCHEN (1991): "On fitting recalcitrant series: the pound/dollar exchange rate, 1973-1983," in *Nonparametric and semiparametric methods in economics and statistics*, ed. by W. Barnett, J. Powell,and G. Tauchen. Cambridge: Cambridge University Press.

GHYSELS, E., A. HARVEY, AND E. RENAULT (1995): "Stochastic Volatility," in *Handbook of Statistics vol 14*. Amsterdam: North Holland.

GIRAITIS, L., AND D. SURGAILIS (1990a): "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotically normality of Whittle's estimate," *Probability Theory & Related Fields*, 86, 105–129.

GIRAITIS, L., AND D. SURGAILIS (1990b): "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotically normality of Whittle's estimate," *Probability Theory & Related Fields*, 86, 105–129.

GRANGER, C. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

GRANGER, C. (1991): "Nonlinear transformations of integrated time series," *Journal of Time Series Analysis*, 12, 207–224.

GRANGER, C., AND A. ANDERSEN (1978): "On the invertibility of time series models," *Stochastic Processes and their Applications*, 8, 87–92.

GRAYBILL, F. (1983): *Matrices with applications in statistics.* Belmont, Calif. (USA) : Wadsworth International Group, second edn.

GRENANDER, U., AND G. SZEGO (1958): *Toeplitz forms and their applications.* Berkeley ; Los Angeles : University of California Press.

HALL, P., AND C. HEYDE (1980): *Martingale limit theory and its applications*, Probability and mathematical statistics. London ; New York: Academic Press.

HANNAN, E. (1970): *Multiple time series.* New York: Wiley.

HANNAN, E. (1973): "The asymptotic theory of linear time series models," *Journal of Applied Probability*, 10, 130–145.

HANNAN, E. (1976): "The asymptotic distribution of serial covariances," *Annals of Statistics*, 4, 396–399.

HARDY, G., J. LITTLEWOOD, AND G. POLYA (1964): *Inequalities.* Cambridge: Cambridge University Press .

HARVEY, A. (1993): "Long memory in stochastic volatility," Discussion Paper 10, London School of Economics, SPS.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

HEYDE, C., AND R. GAY (1993): "Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence," *Stochastic Processes and their Applications.*, 45, 169–182.

HOSKING, J. (1984): "Modelling persistence in hydrological time series using fractional differencing," *Water Resources Research*, 20/12, 1898–1908.

HOSOYA, Y. (1993): "A limit theory on stationary processes with long-range dependence and related statistical models," Discussion Paper 106, Faculty of Economics, Tohoku University.

HOSOYA, Y., AND M. TANIGUCHI (1982): "A central limit theorem for stationary processes and the parameter estimation of linear processes," *The Annals of Statistics*, 10-1, 132–153.

IBRAGIMOV, I. (1959): "Some limit theorems for stochastic processes stationary in the strict sense," *Doklady Akademia Nauk USSR*, 125, 711–714.

KEENAN, D. M. (1987): "Limiting behaviour of functionals of high-order sample cumulants spectra," *Annals of Statistics*, 15-1, 134–151.

KNOPP, K. (1961): *Theory and applications of infinite series.* London: Blackie and son.

LEE, S., AND B. HANSEN (1994): "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory*, 11, 17–27.

LEONOV, V., AND A. SHIRYAEV (1959): "On a method of calculation of semi-invariants," *Theory of Probability Applications*, 4, 319–329.

LJUNG, G. M., AND G. E. P. BOX (1978): "On a measure of lack of fit in time series models," *Biometrika*, 65, 297–303.

LUMSDAINE, R. (1996): "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64/3, 575–596.

MANN, H., AND A. WALD (1943): "On the statistical treatment of linear stochastic difference equations," *Econometrica*, 11, 173–220.

McLeish, D. (1975): "Invariance principles for dependent variables," *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32, 165–178.

Mitrinovic, D. (1970): *Analytic inequalities.* New York: Springer Verlag.

Naylor, A., and G. Sell (1982): *Linear operator theory in engineering and science.* New York : Springer Verlag, second edn.

Nelson, D. (1991): "Conditional heteroscedasticity in asset pricing: a new approach," *Econometrica*, 59, 347–370.

Robinson, P. (1991a): "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," *Journal of Econometrics*, 47, 67–84.

Robinson, P. M. (1977): "The estimation of a nonlinear moving average model," *Stochastic Processes and their Applications*, 5, 81–90.

Robinson, P. M. (1978a): "Alternative models for stationary stochastic processes," *Stochastic Processes and their Applications*, 8, 141–152.

Robinson, P. M. (1978b): "Statistical inference for a random coefficient autoregressive model," *Scandinavian Journal of Statistics*, 5, 163–168.

Robinson, P. M. (1983): "Nonparametric estimators for time series," *Journal of Time Series Analysis*, 4, 186–207.

Robinson, P. M. (1991b): "Nonparametric function estimation for long-memory time series," in *Semiparametric and Nonparametric Methods in Statistics and Econometrics*, ed. by W.Barnett, et al., pp. 189–232. Cambridge: Cambridge University Press.

Robinson, P. M. (1991c): "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," *Journal of Econometrics*, 47, 67–84.

Robinson, P. M. (1994a): "Rates of convergence and optimal bandwidth for long-range dependence," *Probability Theory and Related Fields*, 99, 443–473.

Robinson, P. M. (1994b): "Semiparametric analysis of long-memory time series," *The Annals of Statistics*, 22-1, 515–539.

ROBINSON, P. M. (1994c): "Time series with strong dependence," in *Advances in Econometrics Sixth World Congress vol.1*, ed. by C. Sims, pp. 47–96. Cambridge: Cambridge University Press.

ROBINSON, P. M. (1995a): "Gaussian semiparametric estimation of long range dependence," *Annals of Statistics*, 23/5, 1630–1661.

ROBINSON, P. M. (1995b): "Log-periodogram regression of time series with long-range dependence," *Annals of Statistics*, 23, 1048–1072.

ROBINSON, P. M., AND P. ZAFFARONI (1996a): "Modelling nonlinearity and long memory in time series," in *Nonlinear Dynamics and Time Series: Building a Bridge between the Natural and Statistical Sciences*, ed. by C. Cutler, and D. Kaplan. Providence (USA) : American Mathematical Society.

ROBINSON, P. M., AND P. ZAFFARONI (1996b): "Nonlinear time series with long memory: a model for stochastic volatility," Econometrics Discussion Paper 97/320, STICERD, London School of Econmics.

ROSENBLATT, M. (1956): "A central limit theorem and a strong mixing condition," *Proceedings of the National Academy of Sciences, USA*, 42, 43–47.

ROSENBLATT, M. (1987): "Scale renormalization and random solutions of the Burgers equation," *Journal of Applied Probability*, 24, 328–338.

ROYDEN, H. (1968): *Real analysis.* New York : Mac-Millan Publishing Co. In., second edn.

ROZANOV, Y. (1963): *Stationary random processes.* San Francisco : Holden-Day.

RUIZ, E. (1994): "Quasi-maximum likelihood estimation of stochastic volatility models," *Journal of Econometrics*, 63/1, 289–306.

SENTANA, E. (1995): "Quadratic ARCH models," *Review of Economic Studies*, 62, 639–661.

SHEPHARD, N. (1996): "Statistical aspect of ARCH and stochastic volatility," in *Time Series models in Econometrics, Finance and other Fields*, ed. by O. Barndorff-Nielsen, D. Cox, and D. Hinkley. London : Chapman Hall.

STOUT, W. (1974): *Almost sure convergence.* New York ; London : Academic Press Inc.

SUBBA RAO, T., AND M. GABR (1984): *Introduction to Bispectral Analysis and Bilinear Time Series Models*, vol. 24 of *Lecture Notes in Statistics*. New York : Springer Verlag.

TANIGUCHI, M. (1982): "On estimation of the integrals of the fourth order cumulant spectral density," *Biometrika*, 69, 117–122.

TAQQU, M. (1975): "Weak convergence to fractional Brownian motion and the Rosenblatt process," *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 31, 287–302.

TAQQU, M. (1987): "Random Processes with long-range dependence and high variability," *Journal of Geophysical Resources*, 92, 9683–9986.

TAUCHEN, G., AND M. PITTS (1983): "The price variability volume relationship on speculative markets," *Econometrica*, 51, 485–505.

TAYLOR, S. (1986): *Modelling Financial Time Series*. Chichester (UK) : Wiley.

TIERNEY, L. (1994): "Markov Chains for exploring posterior distributions," *Annals of Statistics*, 22/4, 1701–1762.

TONG, H. (1991): *Non-Linear Time Series : a Dynamical System Approach*. Oxford : Clarendon Press ; New York : Oxford University Press.

VOLTERRA, V. (1931): *Theory of functionals and of integro-differential equations*. London.

WEISS, A. (1986): "Asymptotic theory for ARCH models: estimation and testing," *Econometric Theory*, 2, 101–131.

WHITTLE, P. (1962): "Gaussian estimation in stationary time series," *Bulletin of the International Statistical Institute*, 39, 105–129.

WIENER, N. (1958): *Non-linear problems in random theory*. Cambridge (USA) : MIT press.

YONG, C. H. (1974): *Asymptotic Behaviour of Trigonometric Series*. Honk Kong: Chinese University of Hong Kong.

ZYGMUND, A. (1977): *Trigonometric series*. Cambridge: Cambridge University Press .