Chapter 3

Theory and Methodology

3.1 Unit Root Test Model

Consider a simple AR(1) process

$$y_t = \rho y_{t-1} + \chi_t^1 \delta + \varepsilon_t \tag{1}$$

where χ_t are optional exogenous which may consist of constant, or a constant and trend ρ , δ and are parameters to be estimated, and the ε_t are assumed to be white noise. If $|\rho| \geq 1$, y is a non-stationary series and the variance of increases with time and approaches infinity. If, $|\rho| < 1$, y is a trend-stationary series. Thus, the hypothesis of trend-stationary can be evaluated by testing whether the absolute value of ρ is strictly less than one.

The null hypothesis $H_0 : \rho = 1$ against the one-sided alternative. In some cases, the null is tested against a point alternative $H_1 : \rho < 1$ .In some cases, the null is tested against a point alternative. In contrast, the KPSS Lagrange Multiplier test evaluates the null of $H_0 : \rho < 1$ against the alternative $H_1 : \rho = 1$.

3.1.1 The Augmented Dickey-Fuller (ADF) Test

The standard DF test is carried out by estimating Equation (1) after y_{t-1} subtracting from both sides of the equation:

$$\Delta y_t = \alpha y_{t-1} + \chi_t' \delta + \varepsilon_t \tag{2}$$

where $\alpha = \rho - 1$. The null and alternative hypotheses can be written as,

$$H_0 : \alpha = 0$$

$$H_1 : \alpha < 0 \tag{3}$$

and evaluated using the conventional t -ratio for α :

$$t_\alpha = \hat{\alpha} / (se(\hat{\alpha})) \tag{4}$$

where $\hat{\alpha}$ is the estimate of α, and $se(\hat{\alpha})$ is the coefficient standard error. Dickey and Fuller show that under the null hypothesis of a unit root, this statistic does not follow the conventional Student's t-distribution, and they derive asymptotic results and simulate critical values for various test and sample sizes.

The simple Dickey-Fuller unit root test described above is valid only if the series is an AR(1) process. If the series is correlated at higher order lags, the assumption of white noise disturbances is ε_t violated. The Augmented Dickey-Fuller (ADF) test constructs a parametric correction for higher-order correlation by assuming that the y series follows an AR(ρ) process and adding ρ lagged difference terms of the dependent variable to the right hand side of the test regression:

$$\Delta y_t = \alpha y_{t-1} + \chi_t' \delta + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \nu_t \tag{5}$$

This augmented specification is then used to test (3) using the t -ratio (10). An important result obtained by Fuller is that the asymptotic distribution of the t -ratio for α is independent of the number of lagged first differences included in the ADF regression. Moreover, the assumption demonstrate that the ADF test is asymptotically valid in the presence of a moving average (MA) component, provided that sufficient lagged difference terms are included in the test regression.

In practically, first, the data must be chosen whether to include exogenous variables in the test regression. There is a choice of including a constant, a constant and a linear time trend, or neither in the test regression.

One approach would be to run the test with both a constant and a linear trend since the other two cases are just special cases of this more general specification. However, including irrelevant regressors in the regression will reduce the power of the test to reject the null of a unit root. The standard recommendation is to choose a specification that is a plausible description of the data under both the null and alternative hypotheses.

Second, this paper has to specify the number of lagged difference terms to be added to the test regression (0 yields the standard DF test; integers greater than 0 correspond to ADF tests). The usual (though not particularly useful) advice is to include a number of lags sufficient to remove serial correlation in the residuals.

3.1.2 The Phillips-Perron (PP) Test

The alternative (nonparametric) method of controlling for serial correlation when testing for a unit root. The PP method estimates the non-augmented-DF test equation (2), and modifies the t -ratio of the α coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. The PP test is based on the statistic:

$$\tilde{t}_{\alpha} = t_{\alpha}\left(\frac{\gamma_0}{f_0}\right)^{1/2} - \frac{T(f_0 - \gamma_0)(se(\hat{\alpha}))}{2 f_0^{1/2} s} \qquad (6)$$

where $\hat{\alpha}$ is the estimate, and t_{α} the t-ratio of α, $se(\hat{\alpha})$ is coefficient standard error, and s is the standard error of the test regression. In addition γ_0, is a consistent estimate of the error variance in (2) (calculated as $(T - k)s^2/T$, where k is the number of regressors). The remaining term, 0 f is an estimator of the residual spectrum at frequency zero.

There are two choices you will have make when performing the PP test. First, you must choose whether to include a constant, a constant and a linear time trend, or neither, in the test regression. Second, you will have to choose a method for estimating f_0.

3.1.3 The Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) Test

The testing differs from the other unit root tests described here in that the series y_t is assumed to be trend stationary under the null. The KPSS statistic is based on the residuals from the OLS regression of y_t on the exogenous variables χ_t :

$$y_t = \chi_t' \delta + u_t \tag{7}$$

The LM statistic is be defined as:

$$LM = \sum_t S(t)^2 /(T^2 f_0) \tag{8}$$

where f_0, is an estimator of the residual spectrum at frequency zero and where $S(t)$ is a cumulative residual function:

$$S(t) = \sum_{\gamma=1}^{t} \hat{u}_t \tag{9}$$

based on the residuals $\hat{u}_t = y_t - \chi_t' \delta(0)$. This paper points out that the estimator of δ used in this calculation differs from the estimator for δ used by GLS distending since it is based on a regression involving the original data and not on the quasi-differenced data. To specify the KPSS test, you must specify the set of exogenous regressors χ_t and a method for estimating f_0.

3.2 ARCH - GARCH Model

The autoregressive conditional heteroskedasticity (ARCH) model is the first model of conditional heteroskedasticity. The original idea was to find a model that could assess uncertainty changing over time. Let ε_t be a random variable that has a mean and a variance conditionally on the information set F_{t-1} (the σ - field generated by , ε_{t-n}, $n \geq 1$) :The ARCH model of ε_t has the following properties.

First, $E\{\varepsilon_t \mid F_{t-1}\}$and, second, the conditional variance $h_t = E\{\varepsilon_t \mid F_{t-1}\}$ is a nontrivial positive-valued parametric function of $F_{t-1}$the sequence $\{\varepsilon_t\}$ may be observed directly, or it may be an error or innovation sequence of an econometric model. In the latter case,

$$\varepsilon_t = y_t - \mu_t(y_t) \tag{10}$$

where y_t is an observable random variable and $\mu_t(y_y) = E\{\varepsilon_t \mid F_{t-1}\}$ the conditional mean of y_t given F_{t-1} the application was of this type. In what follows, the focus will be on parametric forms of h_t, and $\mu_t(y_t)$will be ignored. Engle assumed that ε_t can be decomposed as follows:

$$\varepsilon_t = z_t h_t^{1/2} \tag{11}$$

where $\{z_t\}$ is a sequence of independent, identically distributed (iid) random variables with zero mean and unit variance. This implies $\varepsilon_t \mid F_{t-1} \sim D(0, h_t)$where D stands for the distribution (typically assumed to be a normal or a leptokurtic one). The following conditional variance defines an ARCH model of order q:

$$h_t = \alpha_0 + \sum_{n=1}^{q} \alpha_n \varepsilon_{t-n}^2 \tag{12}$$

Where $\alpha_0 > 0, \alpha_n \geq 0, n = 1,..., q-1$ and $\alpha_q > 0$ the parameter restrictions in (12) form a necessary and sufficient condition for positively of the conditional variance. Suppose the unconditional variance $E\varepsilon_t^2 = \alpha^2$ the definition of ε_t through the decomposition (11) involving z_t then guarantees the white noise property of the sequence $\{\varepsilon_t\}$, since $\{z_t\}$ is a sequence of iid variables. Engle and others soon realized the potential of the ARCH model in financial applications that required forecasting volatility. The ARCH model and its generalizations are applied to modeling, among other things, exchange rates and export volumes. Forecasting volatility of these series is different from forecasting the conditional mean of a process because volatility, the object to be forecast, is not observed. The question then is how volatility should be measured. Using ε_t^2 is an obvious but not necessarily.

In applications, the ARCH model has been replaced by the so-called generalized ARCH (GARCH) model, the conditional variance is also a linear function of its own lags and has the form

$$h_t = \alpha_0 + \sum_{n=1}^{q} \alpha_n \varepsilon_{t-n}^2 + \sum_{n=1}^{p} \beta_n h_{t-n} \qquad (13)$$

The conditional variance defined by (13) has the property that the unconditional autocorrelation function of ε_t^2 if it exists, can decay slowly, albeit still exponentially. For the ARCH family, the decay rate is too rapid compared to what is typically observed in financial time series, unless the maximum lag q in (12) is long. As (12) is a more parsimonious model of the conditional variance than a high-order ARCH model, most users prefer it to the simpler ARCH alternative. The overwhelmingly most popular GARCH model in applications has been the GARCH(1,1) model, that is, p = q = 1 in (13). A sufficient condition for the conditional variance to be positive with probability one is $\alpha_0 > 0, \alpha_n \geq 0, n = 1,..., q; \beta \geq 0, n = 1,..., p$. The necessary and sufficient conditions for positivity of the conditional variance in higher-order GARCH models are more complicated than the sufficient conditions just mentioned and have been given in

Nelson and Cao (1992). Note that for the GARCH model to be identified if at least one $\beta_n > 0$ (the model is a genuine GARCH model) one has to require that also at least one $\alpha_n > 0$. If $\alpha_1 = ... = \alpha_q = 0$ the conditional and unconditional variances of ε_t are equal and $\beta_1,...,\beta_p$ are unidentified nuisance parameters. The GARCH(*p,q*) process is weakly stationary if and only if $\sum_{n-1}^{q} \alpha_n + \sum_{n-1}^{p} \beta_n < 1$

The stationary GARCH model has been slightly simplified by variance targeting. This implies replacing the intercept α_0 in (13) by $\left(1 - \sum_{n-1}^{q} \alpha_n - \sum_{n-1}^{p} \beta_n\right)\sigma^2$ where $\sigma^2 = E\varepsilon_t^2$.The estimate $\hat{\sigma}^2 = T^{-1}\sum_{t-1}^{T} \varepsilon_t^2$ is substituted for σ^2 before estimating the other parameters. As a result, the conditional variance converges towards the long-run unconditional variance, and the model contains one parameter less than the standard GARCH(*p,q*) model. It may be pointed out that the GARCH model is a special case of an infinite-order (ARCH(∞)) model (11) with

$$h_t = \alpha_0 + \sum_{n=1}^{\infty} \alpha_n \varepsilon_{t-n}^2 \tag{14}$$

The ARCH(∞) representation is useful in considering properties of ARCH and GARCH models such as the existence of moments and long memory.

3.3 Definition and properties of the FIGARCH process

Following Engle (1982), consider the discrete time real-valued stochastic ARCH process, $\{\varepsilon_t\}$,

$$\varepsilon_t \equiv z_t \sigma_t \tag{15}$$

where $E_{t-1}(z_t) = 0$ and $VAR_{t-1}(z_t) = 1$, and σ_t is a positive time-varying and measurable function with respect to the information set available at time $t-1$. Throughout, $E_{t-1}(\cdot)$ and $VAR_{t-1}(\cdot)$ refer to the conditional expectation and variance with respect to this same information set. Thus, by definition, the $\{\varepsilon_t\}$ process is serially uncorrelated with mean zero, but the conditional variance of the process, σ_t^2, is changing over time.

In the classic ARCH(q) model of Engle (1982), the conditional variance σ_t^2 is postulated to be a linear function of the lagged squared innovations implying Markovian dependence dating back only q periods; i.e., ε_t^2 for $i = 1,2,...,q$. The GARCH(p,q) specification of Bollerslev (1986) provides a more flexible lag structure. Formally, this model is defined by

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \tag{16}$$

where L denotes the lag or backshift operator, and $\alpha(L) \equiv \alpha_1 L + \alpha_2 L^2 + ... + \alpha_q L^q$ and $\beta(L) \equiv \beta_1 L + \beta_2 L^2 + ... + \beta_p L^p$. For stability and covariance stationarity of the $\{\varepsilon_t\}$ process, all the roots of $[1 - \alpha(L) - \beta(L)]$ and $[1 - \beta(L)]$ are constrained to lie outside the unit circle. The GARCH(p, q) process may be rewritten as the infinite-order ARCH process,

$$\begin{aligned} \sigma_t^2 &= \omega[1 - \beta(1)]^{-1} + \alpha(L)[1 - \beta(L)]^{-1}\varepsilon_t^2 \\ &\equiv \omega[1 - \beta(1)]^{-1} + \lambda(L)\varepsilon_t^2 \end{aligned} \tag{17}$$

The before mentioned stationarity condition implies that the effect of the past squared innovations on the current conditional variance decays exponentially with the lag length. Alternatively, the GARCH(p, q) process in Eq. (16) may also be expressed as an ARMA(m, p) process in ε_t^2,

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \tag{18}$$

where $m \equiv \max\{p, q\}$, and $\nu_t \equiv \varepsilon_t^2 - \sigma_t^2$ is mean zero serially uncorrelated. Thus, the $\{\nu_t\}$ process is readily interpreted as the 'innovations' for the conditional variance. When the autoregressive lag polynomial, $[1 - \alpha(L) - \beta(L)]$, contains a unit root, the GARCH(*p, q*) process is defined by Engle and Boilerslev (1986) to be integrated in variance. The corresponding Integrated GARCH(*p, q*), or IGARCH(*p, q*), class of models is given succinctly by

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + [1 - \beta(L)]\nu_t \tag{19}$$

where $\phi(L) \equiv [1 - \alpha(L) - \beta(L)](1 - L)^{-1}$ is of order $m - 1$. The Fractionally Integrated GARCH, or FIGARCH, class of models is simply obtained by replacing the first difference operator in Eq. (19) with the fractional differencing operator.

In order to motivate this development, it is worth briefly considering the fractionally integrated process for the mean. The concept of long-memory and fractional Brownian motion was originally developed by Hurst (1951) and Mandelbrot and Van Ness (1968). However, the ideas were essentially operationalized for applications with discrete time representations by Granger (1980, 1981), Granger and Joyeux (1980), and Hosking (1981). In particular, the ARFIMA(*k, d, l*) class of models for the discrete time real-valued process $\{y_t\}$ is defined by

$$a(L)(1 - L)^d y_t = b(L)\varepsilon_t \tag{20}$$

where $a(L)$ and $b(L)$ are polynomials in the lag operator of orders k and l respectively, and $\{\varepsilon_t\}$ is a mean-zero, serially uncorrelated process. The fractional differencing operator, (1 - L) d, has a binomial expansion which is most conveniently expressed in terms of the hypergeometric function,

$$(1-L)^d = F(-d,1,1;L)$$

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

$$= \sum_{k=0,\infty} \pi_k L^k \quad (21)$$

where $\Gamma(\cdot)$ denotes the Gamma function. Provided that $\text{var}(\varepsilon_t) < \infty$ and $-0.5 < d < 0.5$, the $\{y_t\}$ process in Eq. (20) is weakly stationary and invertible, and will possess unique infinite moving average and autoregressive representations. However, for d > 0 the process is long memory in the sense that $\lim_{k\to\infty} \sum_{j=-k,k} |\rho_j|$, where ρ_j denotes the autocorrelation of the process at lag *j*, does not converge to a finite limit. As argued forcefully by Sowell (1992b), the ARFIMA model essentially disentangles the short-run and the long-run dynamics, by modelling the short-run behavior through the conventional ARMA lag polynomials, *a(L)* and *b(L),* while the long-run characteristic is captured by the fractional differencing parameter, *d*.

Analogously to the ARFIMA(*k, d, l*) process for the mean, the FIGARCH(*p, d, q*) process for $\{\varepsilon_t\}$ is naturally defined by

$$\phi(L)(1-L)^d \varepsilon_t^2 = \omega + [1-\beta(L)]v_t \quad (22)$$

where $0 < d < 1$, and all the roots of $\phi(L)$ and $[1-\beta(L)]$ lie outside the unit circle. Rearranging the terms in Eq. (22), an alternative representation for the FIGARCH(*p, d, q*) model is

$$[1-\beta(L)]\sigma_t^2 = \omega + [1-\beta(L)-\phi(L)(1-L)^d]\varepsilon_t^2 \quad (23)$$

Thus, the conditional variance of *e*, is simply given by

$$\sigma_t^2 = \omega[1-\beta(1)]^{-1} + \{1 - [1-\beta(L)]^{-1}\phi(L)(1-L)^d\}\varepsilon_t^2$$

$$\equiv \omega[1-\beta(1)]^{-1} + \lambda(L)\varepsilon_t^2 \qquad (24)$$

where $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \ldots$. Of course, for the FIGARCH(*p, d, q*) process in Eq. (22) to be well-defined and the conditional variance to be positive almost surely for all *t*, all the coefficients in the infinite ARCH representation in Eq. (24) must be nonnegative; i.e., $\lambda_k \geq 0$ for *k* = 1, 2 . . . . . As for the GARCH(*p, q*) class of models analyzed by Nelson and Cao (1992), general conditions to ensure nonnegativity of all the lag coefficients in $\lambda(L)$ have proven elusive. Fortunately, as illustrated below, sufficient conditions are fairly easy to establish on a caseby- case basis.

For $0 < d \leq 1$ the hypergeometric function evaluated at *L* = l equals zero, F(-*d*, 1, 1;1) =0, so that $\lambda(1) = 1$. The $\omega > 0$ term therefore has the same interpretation as in the IGARCH model. Consequently, the second moment of the unconditional distribution of ε_t is infinite, and the FIGARCH process is clearly not weakly stationary; a feature it shares with the IGARCH class of processes. However, as shown by Nelson (1990a) for the IGARCH(1, 1) model and extended to the general IGARCH(p, q) model by Bougerol and Picard (1992), IGARCH models are strictly stationary and ergodic. Since the high-order lag coefficients in the infinite ARCH representation of any FIGARCH model may be dominated in an absolute value sense by the corresponding IGARCH coefficients from Eq. (19), it follows by a direct extension of the proofs for the IGARCH case that the FIGARCH(p, d, q) class of processes is strictly stationary and ergodic for $0 \leq d \leq 1$.

As highlighted by this discussion, considerable care should be exercised in interpreting persistence in nonlinear models. Formally, Bollerslev and Engle (1993) define a process to be persistent in variance if lim $\sup_{k\to\infty} |\mathrm{E}_{t+s}(\varepsilon_{t+k}^2) - \mathrm{E}_t(\varepsilon_{t+k}^2)| > 0$ for some *s* > 0. This same notion of infinite dependence on the initial conditions for the optimal forecasts of the future conditional variances also underlies the conditional moment profiles analyzed by Gallant, Rossi, and Tauchen (1993). However, in the present context in which the conditional variance is parameterized as a linear function of the past squared innovations, the persistence of the conditional variance is most simply characterized in terms of the impulse response coefficients for the optimal forecast of the future conditional variance as a function of the time *t* innovation, ν_t,

$$\gamma_k \equiv \partial \mathrm{E}_t\left(\varepsilon_{t+k}^2\right)/\partial \nu_t - \partial \mathrm{E}_t\left(\varepsilon_{t+k-1}^2\right)/\partial \nu_t \tag{25}$$

Of course, in more general conditional variance models the γ_i's will depend on the time *t* information set. However, for the FIGARCH class of models analyzed here, the impulse response coefficients are independent of t, and the persistence as measured by the γ_i coefficients corresponds directly to the generalization of the linear impulse response analysis to nonlinear models developed by Gallant, Rossi, and Tauchen (1993). Specifically, the impulse response coefficients may be found from the coefficients in the $\gamma(L)$ lag polynomial,

$$(1-L)\varepsilon_t^2 = (1-L)^{1-d}\phi(L)^{-1}\omega + (1-L)^{1-d}\phi(L)^{-1}\left[1-\beta(L)\right]\nu_t$$

$$\equiv \zeta + \gamma(L)\nu_t \tag{26}$$

where the first equality follows directly from the definition of the FIGARCH(*p, d, q*) process in Eq. (22). Analogously to conventional impulse response analysis for the mean, the long-run impact of past shocks for the volatility process may now be assessed in terms of the limit of the cumulative impulse response weights, i.e.,

$$\gamma(1) = \lim_{k\to\infty}\sum_{i=0,k}\gamma_i = \lim_{k\to\infty}\lambda_k$$

$$= F\left(d-1,1,1;1\right)\phi(1)^{-1}\left[1-\beta(1)\right] \tag{27}$$

As noted above, for $0 \le d < 1, F(d-1,1,1;1) = 0$, so that for the covariance stationary GARCH(*p, q*) model and the FIGARCH(*p, d, q*) model with $0 < d < 1$, shocks to the conditional variance will ultimately die out in a forecasting sense. There are important differences in the shock dissipation for $d = 0$ and $0 < d < 1$, however. Whereas shocks to the GARCH process die out at a fast exponential rate, for the FIGARCH model λ_k will eventually be dominated by a hyperbolic rate of decay; see, e.g., Diebold, Husted, and Rush (1991). Thus, even though the cumulative impulse response function converges to zero for $0 \le d < 1$, the fractional differencing parameter provides important information regarding the pattern and speed with which shocks to the volatility process are propagated. In contrast, for $d = 1$, $F(d-1, 1, 1; 1) = 1$, and the cumulative impulse response weights will converge to the nonzero constant $\gamma(1) = \phi(1)^{-1} \times [1 - \beta(1)]$. Thus, from a forecasting perspective shocks to the conditional variance of the IGARCH model persist indefinitely. For values of $d > 1$, $F(d-1, 1, 1; 1) = \infty$, resulting in an unrealistic explosive conditional variance process and $\gamma(1)$ being undefined.

In most practical applications relatively simple first-order models have been found to provide good representations of the conditional variance processes. To illustrate the ideas developed above consider therefore the simple GARCH(1,1) model,

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

rewritten in ARMA(1,1) form as

$$(1 - \phi_1 L)\varepsilon_t^2 = \omega + (1 - \beta_1 L)v_t$$

where $\phi_1 \equiv \alpha_1 + \beta_1$. The impulse response weights for this model are given by the coefficients in the polynomial, $\gamma(L) = (1-L)(1 - \phi_1 L)^{-1}(1 - \beta_1 L)$, so that $\gamma_0 = 1, \gamma_1 = \phi_1 - \beta_1 - 1$, and $\gamma_k = (\phi_1 - \beta_1)(\phi_1 - 1)\phi^{k-2}$ for $k > 2$; see also Engle and

Bollerslev (1986). The cumulative impulse response weights for the process equals $\lambda_k = (\phi_1 - \beta_1)\phi_1^{k-1}$ for $k > 1$, and in the limit $\gamma(1) = 0$ provided that $0 < \phi_1 < 1$. Hence, the effect of a shock for the forecast of the future conditional variance tend to zero at a fast exponential rate. The IGARCH(1,1) model occurs when $\phi_t = 1$,

$$(1 - L)\varepsilon_t^2 = \omega + (1 - \beta_1 L)v_t$$

In this situation, $\lambda_k = (1 - \beta_1)$ for all lags $k > 1$, and all the cumulative impulse response weights are equal to the nonzero constant $\gamma(1) = 1 - \beta_1$. The corresponding FIGARCH(1, *d*, 0) model is

$$(1 - L)^d \varepsilon_t^2 = \omega + (1 - \beta_1 L)v_t$$

By analogy to the properties for the ARFIMA(0,*d*, 1) model developed in Hosking (1981), it is possible to show that the cumulative impulse response coefficients in the infinite ARCH representation for the FIGARCH(1,*d*, 0) model, $\lambda(L) \equiv 1 - (1 - \beta_1 L)^{-1}(1 - L)^d$, equal

$$\lambda_k = \left[1 - \beta_1 - (1 - d)k^{-1}\right] \cdot \Gamma(k + d - 1)\Gamma(k)^{-1}\Gamma(d)^{-1}$$

for $k > 1$, and $\lambda_0 = 1$. Thus, provided that to $\omega > 0$, the condition $0 \le \beta_1 < d \le 1$ is both necessary and sufficient to ensure that the conditional variance in the FIGARCH(1,d, 0) model is positive almost surely for all *t*. Furthermore, it follows by a straightforward application of Sterling's formula, that for high lags, *k*,

$$\lambda_k \approx \left[(1 - \beta_1)\Gamma(d)^{-1}\right]k^{d-1}$$

In contrast to the covariance stationary GARCH(1,1) model or the IGARCH (l,1) model, where shocks to the conditional variance either dissipates exponentially or persist indefinitely; for the FIGARCH(I, d, 0) model the response of the conditional variance to past shocks decays at a slow hyperbolic rate.

3.4 The Normal Inverse Gaussian distribution

The normal inverse Gaussian distribution is a variance-mean mixture of a Gaussian distribution with an inverse Gaussian. A stochastic variable *X* is said to be normal inverse Gaussian if it has a probability density function of the form [l, 2, 5]

$$f_X(x) = \frac{\alpha\delta}{\pi}\frac{\exp[p(x)]}{q(x)}K_1[\alpha q(x)] \tag{28}$$

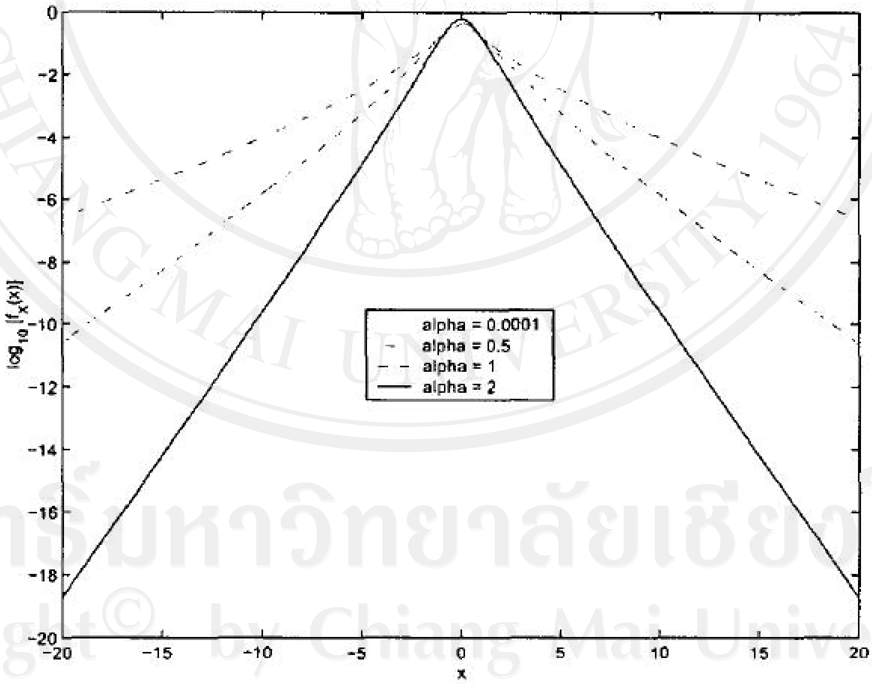

Figure 3.1 NIG-density (logarithmic scale) for different values of α. Here, $\beta = \mu = 0$, and $\delta = 1$.

where K_1 (x) is the modified Bessel function of the second kind with index 1, $p(x) = \delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu), q(x) = \left((x - \mu)^2 + \delta^2\right)^{1/2}$.Furthermore, $0 \le |\beta| < \alpha, \delta > 0$, and $-\infty < \mu < \infty$.

As seen from the definition in Eq. (28), the shape of the NIG-density is specified by a four dimensional parameter vector ($\alpha, \beta, \mu, \delta$). This parameterization is very flexible indeed, making it possible to model a large variety of shapes and with various decay rates of the tail.

The four parameters of the NIG-distribution have natural interpretations relating to the overall shape of the density as follows. The a-parameter controls the steepness of the density, in the sense that the steepness or pointiness of the density increases monotonically with increasing α. This has implications also for the tail behavior, by the fact that large values of α implies light tails, while smaller values of α implies heavier tails. Note the similarity between this parameter and the α-parameter in the α-stable distribution. Figure 3.1 shows the dependency on α for $\beta = \mu = 0$ and $\delta = 1$. Note that the tails become heavier as the value of α decreases.

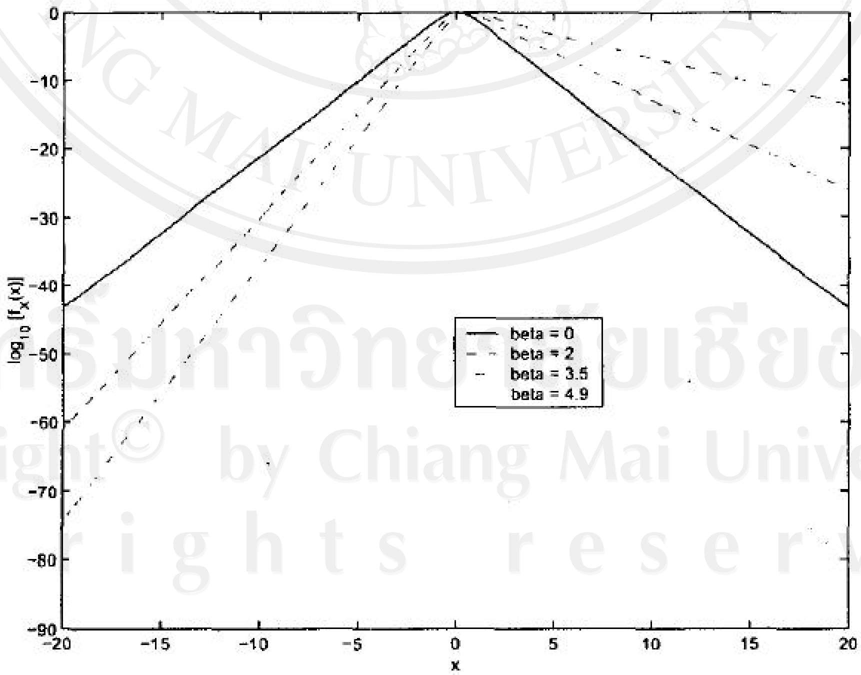

Figure 3.2 NIG-density (logarithmic scale) for different values of β. Here, $\alpha = 5, \mu = 0$, and $\delta = 1$.

The $\beta-$parameter is a skewness parameter, in the sense that $\beta<0$ implies a density skew to the left, $\beta>0$ implies a density skew to the right, and ,Ll = 0 implies a density that is symmetric around μ, which is obviously a centrality or translation parameter. Figure 3.2 shows the dependency on β. Note that the skewness increases as β increases.

Last, the $\delta-$parameter is a scale parameter in the sense that the rescaled parameters $\alpha \rightarrow \alpha\delta$ and $\beta \rightarrow \beta\delta$ are invariant under location-scale changes of *x*.

3.5 Properties of NIG Variables

NIG-variables obey several desirable properties that make them suitable for practical noise modeling. We will now demonstrate the attractiveness of the NIG-distribution in terms of some of its properties.

3.5.1 Cumulants

Bamdorff-Nielsen(1997) derived the moment generating function of the NIG-distribution. By generalizing his result, we readily derive the characteristic function of the NIG as

$$\Phi_X(\omega)=e^{\delta\sqrt{\alpha^2-\beta^2}}e^{-\delta\sqrt{\alpha^2-(\beta+j\omega)^2}}e^{j\mu\omega} \qquad (29)$$

where $j=\sqrt{-1}$ is the imaginary unit, and $-\infty<\omega<\infty$.

The cumulant generating function $\Psi_X(\omega)=\ln\Phi_X(\omega)$ therefore has the following simple form

$$\Psi_X(\omega)=\delta\left[\sqrt{\alpha^2-\beta^2}-\sqrt{\alpha^2-(\beta+j\omega)^2}\right]+j\mu\omega \qquad (30)$$

With the cumulant generating function at hand, it is now straightforward to calculate the cumulant of order *n* by means of

$$\kappa_X^{(n)} = (-j)^n \frac{d^n \Psi_X(\omega = 0)}{d\omega^n} \tag{31}$$

The first four cumulants are readily found to be

$$\kappa_X^{(1)} = \mu + \frac{\beta\delta}{\sqrt{\alpha^2 - \beta^2}} \qquad \kappa_X^{(2)} = \mu + \frac{\alpha^2\delta}{\left(\alpha^2 - \beta^2\right)^{3/2}} \tag{32}$$

$$\kappa_X^{(3)} = \mu + \frac{3\alpha^2\beta\delta}{\left(\alpha^2 - \beta^2\right)^{5/2}} \qquad \kappa_X^{(4)} = \mu + \frac{3\alpha^2(\alpha^2 + 4\beta^2)\delta}{\left(\alpha^2 - \beta^2\right)^{7/2}} \tag{33}$$

We thus see that all cumulants exist and that they are expressible as simple algebraic functions of the parameters.

It is particularly interesting to notice that the skewness and kurtosis has the following elegant closed form expressions

$$\gamma_3 = \frac{\kappa_X^{(4)}}{\left[\kappa_X^{(2)}\right]^2} = \frac{3\rho}{\sqrt{\xi}} \tag{34}$$

and

$$\gamma_4 = \frac{\kappa_X^{(4)}}{\left[\kappa_X^{(2)}\right]^2} = \frac{3(1 + 4\rho^2)}{\xi} \tag{35}$$

where $\rho = \beta / \alpha$ and $\xi = \delta\sqrt{\alpha^2 - \beta^2}$.

Since -1 < ρ < 1 and $0 < \xi < \infty$, the expressions for skewness and kurtosis, Eqs. (34) and (35) show that we can model data that covers a very large range of non-Gaussian shapes. By combining Eqs. (34) and (35) it is easy to show that we may model variables with any simultaneous skewness and kurtosis in the region $\gamma_4 \geq 4\gamma_3^2 / 3$.

3.5.2 Exact limits

If we assume $\beta = 0$ and μ arbitrary, one can readily show that the Gaussian~X density is a limit when either $\alpha \to \infty$ or $\delta \to \infty$, with the identification that $\sigma^2 = \delta / \alpha$.

Another important special case of the NIG is the Cauchy distribution which results when $\alpha = \beta = 0$, and β and δ arbitrary.

3.5.3 Tail behavior

Asymptotically, the Bessel function behaves as

$$K_1(x) \sim \sqrt{\frac{\pi}{2x}} \exp(-x); |x| \to \infty \qquad (36)$$

Hence, the tail of the NIG decays as

$$f_x(x) \sim |x|^{-3/2} \exp(\beta x - \alpha |x|) \qquad (37)$$

Note that Eq. (37) is invalid when a - << 1. In that special case, the tail of the NIG decays as

$$fx(x) \sim |x|^{-2} \qquad (38)$$

which is of course the tail behavior of the Cauchy.

3.5.4 Convolution property

A very attractive and useful property of the NIG that cannot be overrated, is that it is closed under convolution [1, 2]. This has far reaching consequences when considering sums of NIG variables.

Let $X_1,...,X_M$ be M independent NIG-variables with common parameters α and β, but having individual location parameters $\mu_1,...,\mu_M$ and individual scale parameters $\alpha_1,...,\alpha_M$. Then the sum variable $Y = X_1 + ... + X_M$ is also NIG distributed, with parameters $(\alpha, \beta, \mu_{tot,\delta_{tot}})$, where $\mu_{tot} = \sum_m^M \mu_M$ and $\alpha_{tot} = \sum_{m=1}^M \alpha_m$.

3.6 FIGARCH – NIG Model

In FIGARCH-NIG model, the time series return data from the exchange rate can be written as

$$r_t = \mu + \frac{b\sqrt{\gamma}}{a}\sigma_t + z_t\sigma_t, for\ t = 1,...,T$$

which z_t is zero-mean and unit variance process. From the study of Anderson and Jenson and Lunde defined the z_t to be NIG distributed as

$$z_t \sim NIG(a, b, -\frac{b\sqrt{\gamma}}{a}, \frac{\gamma^{\frac{3}{2}}}{a})$$

where $\gamma = \sqrt{a^2 - b^2}$ in NIG distributed defined that the conditional distribution of returns will be NIG as well

$$r_t \mid \Omega_{t-1} \sim NIG(a, b, \mu, \frac{\gamma^{\frac{3}{2}}}{a}\sigma_t)$$

where Ω_{t-1} is the information set from the previous day return and $\Omega_{t-1} = \sigma(r_{t-1}, r_{t-1}, r_{t-2}, ...)$ and conditional mean and conditional variance can be defined as

$$E[r_t \mid \Omega_{t-1}] = \mu + \sigma_t \frac{b\sqrt{\gamma}}{a}, for\ t = 1,...,T$$

$$Var(r_t \mid \Omega_{t-1}) = \sigma_t^2, for\ t = 1,...,T$$

It is given that $u_t = r_t - E(r_t \mid \Omega_{t-1}) = r_t - \sigma_t \frac{b\sqrt{a^2 - b^2}}{a} - \mu$ which is the innovation of return process.