The non-Gaussian exponential SEMIFAR
and its application as a semiparametric long-memory volatility model

Yuanhua Feng, Sebastian Letmathe

Faculty of Business Administration and Economics, Paderborn University

Abstract

The well-known ESEMIFAR model introduced by **beran2015modelling** simultaneously identifies a deterministic trend as well as short- and long-range dependence. The authors derived statistical properties under the assumption that the stationary part of the ESEMIFAR follows a zero mean FARIMA process with Gaussian innovations. Another finding was that the long-memory parameter could be estimated from the log-transformed process as the Hermite rank of the exponential transformation has rank one. Consequently, the long-memory parameter in the original and the log-transformed process is the same.This paper proposes an extension of the ESEMIFAR to the non-Gaussian case. Under suitable additional assumptions on the distribution of the innovations we show that the Appell rank of the exponential transformation is one as well. Hence, the long-memory parameter is expected to be the same for the original and log-transformed process.Theoretical asymptotic properties of our model are derived. We ensure that data-driven SEMIFAR algorithms are applicable to our proposal. The appropriateness of our approach is then illustrated by the application to financial data. Trend estimation and model selection is implemented with the statistical software R.

Key Words: Long-memory, Exponential SEMIFAR, SEMIFAR algorithms, non-Gaussian, Hermite rank, Appell rank

1 Introduction

2 Non-Gaussian EFARIMA and ESEMIFAR models

The following models are formulated to analyse non-negative equidistant financial time series data. Our proposal is particularly designed for cases where the normality assumption would not be suitable, i.e. financial data with heavy tails for example. We follow Beran et. al (2015) and extend the Gaussian EFARIMA and ESEMFIAR by assuming that the innovations follow some other well-known non-Gaussian distribution. To motivate our proposals we first summarize the results under the normality assumption derived by Beran et. al (2015) and illustrate the relationship of the EFARIMA to the EACD_1.

2.1 EFARIMA and ESEMIFAR models

We define X_t $(t = 1, ...N)$ as a nonnegative stationary financial time series. A well-established model for analysing financial time series data, particularly trade durations, is the multiplicative error model (MEM) which is given by

$$X_t = \nu \lambda_t \eta_t. \tag{1}$$

The scale parameter is denoted by $\nu > 0$, $\lambda_t > 0$ denotes the conditional mean of $X^* = X_t/\nu$, and η_t are i.i.d. random variables. Following Feng and Zhou (2015) we rewrite (1) as a semiparametric MEM given by

$$X_t = \nu(\tau_t)X^* = \nu(\tau_t)\lambda_t \eta_t, \tag{2}$$

where $\tau_t = t/n$ denotes the rescaled time and where we replaced the scale parameter in (1) with a nonparametric scale function denoted by $\nu(\tau_t)$. By taking the logs we can rewrite (2) as a semiparametric FARIMA model:

$$Y_t = \mu(\tau_t) + Z_t = \mu(\tau_t) + \xi_t + \epsilon_t, \tag{3}$$

where $Y_t = ln(X_t)$, $\mu(\tau_t) = ln[\nu(\tau_t)]$, $Z_t = ln(X^*)$, $\xi_t = ln(\lambda_t)$ and $\epsilon_t = ln(\eta_t)$. Following Beran et. al (2015) we assume that Z_t follows a FARIMA (p, d, q) process with zero mean:

$$(1 - B)^d \phi(B)(\xi_t + \epsilon_t) = (1 - B)^d \phi(B)Z_t = \psi(B)\epsilon_t, \tag{4}$$

where $d \in (0, 0.5)$ is the long-memory parameter, $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ and $\psi(z) = 1 - \sum_{i=1}^{q} \psi_i z^i$ are AR- and MA-polynomials with all roots outside the unit circle. Equation (4) defines a stationary and invertible FARIMA process with $E(\epsilon_t) = 0$ and $var(\epsilon_t) = \sigma_\epsilon^2$. Under the strong assumption that $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ in Z_t and hence $\eta \sim LN(0, \sigma_\epsilon^2)$ in $X_t^* = exp(Z_t) = exp(\xi_t + \epsilon_t)$, Z_t and X_t^* are explicitly called Gaussian FARIMA and Gaussian EFARIMA respectively.

2.1.1 Stationary solutions

Please note that we do not assume that ϵ_t is not normally distributed. We extend the results of **beran2015modelling** which were obtained under the log-normal assumption to more general distributions. We can write Z_t as a MA(∞) process:

$$Z_t = \sum_{i=0}^{\infty} \beta_i \epsilon_{t-i}, \tag{5}$$

where $\beta(B) = \frac{\psi(B)}{(1-B)^d \phi(B)}$, and $\beta_i / c_\beta i^{d-1} \longrightarrow 1$ for large i with

$$c_\beta = \frac{|\psi(1)|}{|\phi(1)|} \frac{1}{\Gamma(d)} > 0, \tag{6}$$

where $\Gamma(\cdot)$ denotes the Gamma-function. Under long range dependence the autocorrelations of Z_t are not summable, since $\rho_Z(k) \approx c_\rho^Z |k|^{2d-1}$ with $c_\rho^Z > 0$ being a constant. It was shown by Beran (1994), Bondon and Palma (2007), Kokoszka and Taqqu (1995) and Palma (2007) that an FARIMA process has a unique stationary solution given by an infinite moving average representation, for $\phi(B)$ and $\psi(B$ having no common zeros, $\phi(B)$ having roots lying outside the unit circle, and $d \in (-0.5, 0.5)$. They further demonstrate, that the MA coefficients in (5) can then be stated as $\beta_i = \psi(1) i^{d-1} (\phi(1)\Gamma(d))^{-1} + O(i^{-1})$ as $i \to \infty$. Let $\beta_{max} = sup(\beta_i)$ and $\beta_{min} = inf(\beta_i)$. Feng and Zhou (2015) derived the following conditions for the existence of a stationary solution of X^*:

A1: Z_t is a stationary and invertible FARIMA process given by (4)

A2: The second order moments $E(\eta_t^{2u\beta_{max}})$ and $E(\eta_t^{2u\beta_{min}})$ exist, where u is some positive number. Condition A2 is sufficient for the existence of all the terms in (7) and A1 indicates that in (5) $\sum_{i=0}^{\infty} \beta_i < \infty$ and $E(\epsilon_t) = 0$. If both conditions hold X_t^* converges and hence Lemma 1 in Feng and Zhou (2015) implies with regard to A1 and A2 that the stationary solution for X_t^* is given by:

$$X^* = \prod_{i=0}^{\infty} \eta_{t-1}^{\beta_i}. \tag{7}$$

Moreover, the authors indicated that condition A2 is not affected by d and jointly depends on the marginal distribution of X^*, the value of u, and β_i. For log-normal innovations A2 is always fulfilled.

Hier noch weiter (stationary solutions)

2.2 Long memory properties

In the case of Gaussian innovations **beran2015modelling** derived precise formulas for moments, autocovariances and autocorrelations as well as long-memory properties. The authors indicated that the long-memory parameter is the same in the original process X_t and the transformed process Y_t (Z_t). **menendez2013trend** indicated that every monotonous transformation has Hermite rank one. Consequently, it holds that $X_t = G(Z_t) = \nu(\tau_t)X_t^* = \exp(Y_t) = \exp(\mu(\tau_t) + Z_t)$ has Hermite rank one as well. The expansion of transformed RVs by means of orthogonal polynomials is possible only for a very limited amount of distribution classes, e.g., normal and exponential distribution (**beran2013limit**). Therefore, in the face of non-Gaussian innovations it is not possible to represent $G(Z_t)$ in terms of a Hermite expansion. However, if the underlying transformation function fulfils certain conditions and under suitable assumptions on the marginal distribution of Z_t, it is feasible to write the transformation function by means of an Appell polynomial expansion which can be considered to be a generalization of Hermite polynomials (**schutzner2009asymptotic**).

FIXME

2.3 The SEMI-LM-GARCH model

The autoregressive conditional heteroscedasticity (ARCH) model proposed by **engle1982autoregressive** and its generalisation, the generalized ARCH (GARCH)) model, introduced by **bollerslev1986generalize** , is a well-known volatility process approach for modelling non-constant conditional variances . This is of major interest especially for fitting financial time-series and their frequently observed characteristic of conditional heteroscedasticity and volatility clustering (**mandelbrot1963new**). Following **baillie1996fractionally** a common representation of a GARCH(p,q) model is given by

$$\epsilon_t^2 = e_t^2 h_t \iff \epsilon_t = e_t\sqrt{h_t}, \tag{8}$$

$$h_t = \alpha_0 + \sum_{i=0}^{q} \alpha_i \epsilon_{t-i} + \sum_{j=1}^{p} \beta_j h_j = \alpha_0 + \alpha(B)\epsilon_t^2 + \beta(B)h_t \tag{9}$$

where e_t being an i.i.d. standard normal random variable, the square root of h_t denotes the conditional standard deviation, $\alpha_0 > 0$ and $\alpha_1, \ldots, \alpha_q, \beta_1 \ldots \beta_p \geq 0$. $\alpha(B) = \sum_{i=1}^{q} \alpha_i B^i$ and $\beta(B) = \sum_{j=1}^{p} \beta_j B^j$. The GARCH process can also be given by an ARMA(max(p,q)q) type representation in ϵ_t^2 of the form

$$(1 - \alpha(B) - \beta(B))\epsilon_t^2 = \alpha_0 + (1 - \beta(B))u_t, \tag{10}$$

where $u_t = \epsilon_t^2 - h_t$ is a zero-mean process describing the innovations in the conditional variances. See also **feng2008modelling** for an extension of this model to high-frequency financial data. In the context of this paper, we are particularly interested in the long memory structures to arise in the ACF of absolute and squared market returns (see **ding1993long**). Following **baillie1996fractionally** a fractionally integrated GARCH (FIGARCH) may be written as

$$\tilde{\phi}(B)(1 - B)^{\tilde{d}}\epsilon_t^2 = \alpha_0 + [1 - \beta(B)]u_t, \tag{11}$$

where $\tilde{\phi}(B) \equiv (1 - \alpha(B) - \beta(B))(1 - B)^{-1}$ with all roots of $\tilde{\phi}(B)$ and $[1 - \beta(B)]$ outside the unit circle and $0 < \tilde{d} < 1$. By rearranign the terms in (11) The conditional volatility h_t is now given by

$$h_t = \alpha_0[1 - \beta(1)]^{-1} + \{1 - [1 - \beta(B)]^{-1}\tilde{\phi}(B)(1 - B)^{\tilde{d}}\}\epsilon_t^2. \tag{12}$$

Note that the FIGARCH is not second order stationary because it exhibits an infinite unconditional variance. A model closely related to the FIGARCH model is the LM-GARCH Model which belongs to the framework of conditionally heteroscedastic processes introduced by **robinson1991testing**. Following **karanasos2004autocorrelation** (cf. Robinson 2001)FIXME we define the LMGARCH(p, d, q) model as

$$\tilde{\phi}(B)(1 - B)^{\tilde{d}}(\epsilon_t^2 - \alpha_0) = [1 - \beta(B)]u_t. \tag{13}$$

Here the authors model the squared residuals accounting for the deviations from $\alpha_0 \in \mathbb{R}_+$. This centralisation enables us to analogously define the LMGARCH to the FARIMA model for the mean (**conrad2006garch**). According to (23) and (13) we have $E(\epsilon_t) = 0$, $\mathrm{Cov}(\epsilon_t, \epsilon_{t-i}) = 0$ for $i \geq 1$ and the unconditional variance with $E(\epsilon_t^2) = \alpha_0 < \infty$. Moreover, based on the exponential GARCH (EGARCH) proposed by Nelson (1991), Bailie and Mikkelsen (1996a)FIXME introduced the fractionally integrated exponential GARCH (FIEGARCH) where the logarithm of the conditional variance is modelled as a fractionally integrated process. The EGARCH and FIEGARCH accounts for the so-called leverage effect which usually has short-term effects on the dependence structure of the underlying process. Additionally, no constraints on the parameters are needed as the conditional variance is positive by definition. Following **nelson1991conditional** the EGARCH can be parameterized as an ARMA (p,q)

$$\log(h_t) = \alpha_0 + \tilde{\phi}(B)^{-1}(1 + \psi(B))g(e_{t-1}), FIXME \tag{14}$$

where $\psi(B) = 1 + \sum_{i=1}^{q} \psi_i B^i$ are the MA coefficients and $\phi(B)$ and $\psi(B)$ have no common roots. $g(e_t) = \Theta_{e_t} + \gamma[(|e_t| - E|e_t|)]$ is the news impact function and Θ, $\gamma \in \mathbb{R}$. As already indicated by Nelson (1991) the EGARCH can be extended for fractional orders of integration. Following **bollerslev1996modeling** the extension to the FIEGARCH is given by

$$\log(h_t) = \alpha_0 + \tilde{\phi}(B)^{-1}(1 - B)^{-d}[1 + \psi(B)]g(e_{t-1}).FIXME \tag{15}$$

However, for practical purposes we use the following definition for the FIEGARCH

$$\tilde{\phi}(B)(1 - B)^d \log(h_t) = \alpha_0 + \sum_{i=1}^{p} (\alpha_j |e_{t-i}| + \gamma_i e_{t-i}), \tag{16}$$

where $\gamma_i \neq 0$ controls for the existence of leverage effects. Please note that there are various slightly different definitions of a FIEGARCH process. Definition (16) is also considered in the software S-Plus. It can be shown that, under certain mild conditions, these definitions are equivalent to each other (**lopes2012theoretical**). Moreover, the FIEGARCH nests the integrated EGARCH (IEGARCH) and the EGARCH $d = 1$ and $d = 0$ respectively. With regard to the FARIMA, $\log(h_t)$ is second-order stationary and invertible for $d \in \{-0.5; 0.5\}$, see, e.g. **hosking1981fractional**. Another type of GARCH-type model is the so called Log-GARCH model which was independently introduced by Geweke (1986), Pantula (1986) and Milhoj (1987). The Log-GARCH model is given by

$$\log(h_t) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \log(\epsilon_{t-i}^2) + \sum_{j=1}^{q} \beta_j \log(h_{t-j}). \tag{17}$$

Pantula (1986) among others showed that (17) admits an ARMA-type representation. Analogously to (10) we can rearrange the terms in (17) and it follows that

$$[1-\alpha(B)-\beta(B)]log(\epsilon_t) = \alpha_0 + [1-\beta(B)][\log(\epsilon_t^2)-\log(h_t)]. \tag{18}$$

In correspondence to (11) we then have

$$\tilde{\phi}(B)(1-B)^d log(\epsilon_t) = \alpha_0 + [1-\beta(B)]\tilde{u}, \tag{19}$$

where $\tilde{u} = [\log(\epsilon_t^2)-\log(h_t)]$. (19) is called a long-memory Log-GARCH (LM-Log-GARCH) model.

2.3.1 Semiparametric estimation

Let x_t denote the stock prices of some stock index and r_t their corresponding (log-) returns. Consider the equidistant time series model

$$r_t = \mu + \sigma(\tau)\epsilon_t \tag{20}$$

where μ is an unknown constant, $\sigma(\tau)$ a smooth scale function and ϵ_t^2 is assumed to follow a long-memory GARCH process, i.e. a FIGARCH, FIEGARCH or LM-GARCH process defined by (11), (13)and (28). We assume that there are strictly stationary solutions for all these long-memory GARCH processes such that $E(\epsilon_t^8) < \infty$. This assumption is needed for the nonparametric estimation of $\sigma(\tau)$ (**feng2004simultaneously**). Moreover, it is assumed that $var(\epsilon_t) = E(\epsilon_t^2) = 1$. Now define $Z_t = (r_t - \mu)$ then model (20) can be rewritten as:

$$Z_t = \sigma(\tau)\epsilon_t. \tag{21}$$

Please note that model (21) is a semiparametric MEM and is equivalent to model (1). By taking the square we can formulate model (21) as

$$X_t = \sigma(\tau)^2\epsilon_t^2, \tag{22}$$

where $X_t = Z_t^2$. Following **feng2004simultaneously** we can transfer model (32) to the following nonparametric regression problem:

$$X_t = \sigma(\tau)^2 + \sigma(\tau)^2\xi_t, \tag{23}$$

where $\xi_t = (\epsilon_t^2 - 1) \geq -1$ are zero mean stationary time series errors and $\sigma(\tau)^2$ is at the same time trend function and part of the error term. Please see **feng2004simultaneously** for further discussion about the estimation of model (23) in the short memory context. Another way is to simply take the logs of model (22) and then we have

$$Y_t = g(\tau) + \eta_t, \tag{24}$$

where $Y_t = log(X_t)$, $g(\tau) = log(\sigma(\tau)^2)$ and $\eta_t = log(\epsilon_t^2)$. Model (23) and (24) define a semiparametric LM-GARCH model. Please note that model (24) is analogue to the SEMI-FARIMA model defined in (3) which enables us to estimate the scale function $\gamma(\tau)$ in model (24) by means of ESEMIFAR-algorithms. In the case that $\sigma(\tau)^2 \equiv$ is constant over time (FIXME), models (20), (21) and (22) reduce to a standard LM-GARCH model. If $\sigma(\tau)^2$ is a time variant function then the specification of these models as a LM-GARCH is not suitable and would result in an inconsistent estimation (**feng2004simultaneously**). It is proposed to estimate $\sigma(\tau)$ by means of the ESEMIFAR algorithm in order to calculate

$$\hat{\epsilon}_{tt} = X_t / \hat{\sigma}(\tau_t)^2. \tag{25}$$

$\hat{\epsilon}_t$ is approximately stationary. To achieve unit variance we simply divide $\hat{\epsilon}_t$ by its standard deviation σ_ϵ.

3 Conclusion

4 Appendix

```
> ... test
```

PGARCH

Succarat 2016

Francq 2013