CENTRO PER LA FORMAZIONE IN ECONOMIA E POLITICA DELLO SVILUPPO RURALE
UNIVERSITÀ DI NAPOLI FEDERICO II - DIPARTIMENTO DI SCIENZE STATISTICHE
UNIVERSITÀ DI SALERNO - DIPARTIMENTO DI SCIENZE ECONOMICHE E STATISTICHE

Quaderni di STATISTICA

VOLUME 9 - 2007

LIGUORI EDITORE

Volume 9, anno 2007

ISSN 1594-3739 (edizione a stampa)

Registrazione al n. 5264 del 6/12/2001 presso il Tribunale di Napoli
ISBN-13 **978 - 88 - 207 - 4209 - 6**

Direttore responsabile: Gennaro Piccolo

Prima edizione italiana Dicembre 2007
Finito di stampare in Italia nel mese di Dicembre 2007 da OGL - Napoli

Il C.F.E.P.S.R. si avvale per la stampa dei Quaderni di Statistica del contributo dell'Istituto Banco di Napoli - Fondazione.

La carta utilizzata per la stampa di questo volume è inalterabile, priva di acidi, a PH neutro, conforme alle norme UNI EN ISO 9706 ×, realizzata con materie prime fibrose vergini provenienti da piantagioni rinnovabili e prodotti ausiliari assolutamente naturali, non inquinanti e totalmente biodegradabili.

Indice

FORUM

Quaderni di Statistica
Vol. 9, 2007

Estimation of ARIMA models under non-normality

Simona Pacillo
Dipartimento PE.ME.IS, Università degli Studi del Sannio
E-mail: simona.pacillo@unisannio.it

Summary: The present paper deals with the estimation of time series models under departures from normality. The use of flexible models is proposed. These models allow to cope with skewness and/or heavy tails in the distribution of the data while preserving the possibility to carry out inference based on the likelihood function. Normality testing is also discussed. The proposed methodology is applied to estimate an ARIMA model for the energy consumption time series in Queensland.

Keywords: ARIMA models, Departures from normality, Forecasting energy consumption.

1. Introduction

The analysis of real data often leads to reject the hypothesis that they have been generated by a normal distribution (Hill and Dixon, 1982; Azzalini, 1986; Hampel *et al.*, 1986; Azzalini and Della Valle, 1996; Azzalini and Capitanio, 1999). In particular, deviations from normality are frequently observed in time series, where both heavy tails (Tsay, 1986; Denby, 1979) and skewness are often encountered (Wecker, 1981). Martin and Yohai (1985) consider robust methods to cope with non normality in time series data while Ledolter (1979) proposes the use of the symmetric exponential power distribution in ARIMA models for stock price data in order to deal with heavy tails.

Here we are concerned with the possible occurrence of both skewness and leptokurtosis. This paper deals with time series modelling under non-normality and proposes the use of flexible models in the estimation of autoregressive integrated moving average (ARIMA) models to cope with skewness and/or heavy tails.

The flexible models, considered here, are the exponential power (*EP*) distribution (Subbotin, 1923), the skew exponential power (*SEP*) distribution (Azzalini, 1986), the Student t (t_g) distribution, the skew normal (*SN*) distribution (Azzalini, 1985) and the skew t (*St*) distribution (Azzalini and Capitanio, 2003). These models are able to fit distributions of data which are in a neighbourhood of the normal one. Their main benefits are that they allow to carry out inference based on the likelihood function while dealing with departures from normality. Furthermore, the Gaussian assumption can be efficiently verified through a hypothesis test on few model parameters.

Departures from normality are often encountered in electricity consumption time series (Juberias *et al.*, 1999; Papalexopoulos and Hesterberg, 1990). Load forecasting is essential for an efficient planning of the electricity production as it avoids recourse to expensive production techniques to cope with unexpected peaks in demand (Andreotti *et al.*, 2000). Although some regressors (like temperature, numbers of hours of light in the day, etc.) can be used to explain load forecasting, these variables are not always available or recorded with appropriate frequency, so that ARIMA models are widely used. The frequent occurrence of skewness or heavy tails, in the energy consumption time series, points out the need of adequate distributional assumptions.

Section 2 introduces the flexible models, whereas section 3 deals with the estimation of ARIMA model and forecasting under non-normality. Finally, the proposed methodology is applied to model the energy demand in Queensland (Australia), in section 4. Some concluding remarks end the paper.

2. Some flexible models

Consider the location-scale model $X = a + bZ$, where a and b are the location and scale parameter respectively, and Z is a random variable with density function given by one of the distributions considered in this section.

A model, which can be adopted when the deviation from normality is caused by heavy tails, is the *EP* distribution (Subbotin, 1923). The density of X is given by:

$$f_{EP}(x;a,b,\alpha)=\left(\frac{1}{bC_\alpha}\right)exp\left\{-\frac{\left|\frac{x-a}{b}\right|^\alpha}{\alpha}\right\},\ -\infty<x<+\infty, \tag{1}$$

where $C_\alpha = 2\alpha^{\frac{1}{\alpha}-1}\Gamma(1/\alpha)$ and $\alpha>0$. The parameter α controls the thickness of the tails: for $\alpha=2$ the *EP* distribution reduces to a normal model, otherwise for $\alpha\in(1,2)$ it is leptokurtic. This model has been extensively studied by Box (1953), Turner (1960), Lunetta (1963), Mineo and Vianelli (1980), Chiodi (1995) and Capobianco (2000) among others.

Alternatively, the distribution of data with heavy tails can be fitted by the density of a t_g random variable. Under this assumption, the density of X is:

$$f_t(x;a,b,g)=\frac{1}{b}\frac{\Gamma\left(\frac{g+1}{2}\right)}{\sqrt{\pi g}\Gamma(g/2)}\left[1+\frac{\left(\frac{x-a}{b}\right)^2}{g}\right]^{-\frac{g+1}{2}},\ -\infty<x<+\infty. \tag{2}$$

The normal model is obtained as a limit case when the degrees of freedom g go to infinity.

If the deviation from normality is due to a lack of symmetry, a *SN* distribution (Azzalini, 1985) can be assumed for the data. In this case, the density of X is given by:

$$f_{SN}(x;a,b,\lambda)=\frac{2}{b}\phi\left(\frac{x-a}{b}\right)\Phi\left(\frac{x-a}{b}\lambda\right),\ -\infty<x<+\infty, \tag{3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal density and distribution function respectively. The parameter λ regulates the skewness: the normal model is obtained when $\lambda = 0$, for $\lambda > 0$ the skewness is positive, otherwise it is negative. Estimation of the parameters of the skew normal model are discussed in Azzalini (1985), Azzalini and Capitanio (1999), Monti (2003), Chiogna (2005), Liseo and Loperfido (2006) and Sartori (2006).

The *SEP* distribution (Azzalini, 1986) can be adopted when both skewness and heavy tails occur simultaneously. The density of X is given by:

$$f_{SEP}(x;a,b,\lambda,\alpha) = \frac{2}{b}\Phi(w) f_{EP}\left(\frac{x-a}{b};\alpha\right), \ -\infty < x < +\infty, \tag{4}$$

where $w = sign\left(\frac{x-a}{b}\right)\left(\frac{\alpha}{2}\right)^{-1/2} \lambda \left|\frac{x-a}{b}\right|^{\alpha/2}$. In (4) the skewness is regulated by the parameter λ and the kurtosis by the parameter α. The *SEP* distribution reduces to the *EP* distribution when $\lambda = 0$, to the *SN* distribution when $\alpha = 2$, and to the normal distribution when $(\lambda, \alpha) = (0, 2)$. Inference on the *SEP* model is considered by Di Ciccio and Monti (2004).

Finally, another model, which is suitable to handle simultaneous lack of symmetry and kurtosis, is the univariate *St* (Azzalini and Capitanio 2003). Its density is:

$$f_{St}(x;a,b,\lambda,g) = \frac{2}{b} f_t\left(\frac{x-a}{b};g\right) T(w;g+1), \ -\infty < x < +\infty, \tag{5}$$

where $f_t(\cdot)$ and $T(\cdot)$ are the density and distribution function, respectively, of a t_g random variable and

$$w = \lambda\left(\frac{x-a}{b}\right)\left\{(g+1)\Big/\left[\left(\frac{x-a}{b}\right)^2 + g\right]\right\}^{1/2}. \tag{6}$$

As in the previous case, the parameter λ controls the skewness, whereas the tails are regulated by the degrees of freedom g.

When the *EP*, *SN* and *SEP* models are adopted, the hypothesis of normality can be verified through a test on the parameters λ and α. Furthermore symmetry in the *SN* and *St* distributions can be verified by testing the hypothesis $\lambda = 0$. In order to verify normality, when a t_g or *St* distribution is assumed, it has been suggested by Azzalini to construct an asymptotic confidence interval for *g* through the likelihood ratio test. If the upper limit is unbounded, the hypothesis of normality cannot be rejected. For the t_g distribution we consider the profile-likelihood ratio given by:

$$\lambda^*(g_0) = \frac{\sup_{g=g_0} L(\vartheta; x_1,....x_n)}{L(\hat{\vartheta}; x_1,....x_n)}, \tag{7}$$

where $\vartheta = (a,b,g)$ and $\hat{\vartheta}$ is the maximum likelihood estimator. An asymptotic confidence interval, at level $1-\alpha$, is given by $I = \left[g_0 : -2\, ln\left\{ \lambda^*(g_0) \right\} \le \chi^2_{1;1-\alpha} \right]$, where $\chi^2_{1;1-\alpha}$ is the $1-\alpha$ quantile of the χ^2_1 distribution (Barndorff *et al.*, 1994, 89-91).

3. Estimation of ARIMA models and forecasting under non-normality

Consider the seasonal ARIMA model

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t + \theta_0, \tag{8}$$

where ε_t is a Gaussian white noise and the polynomials $\phi(B)$, $\theta(B)$, $\Phi(B^s)$ and $\Theta(B^s)$ are defined according to the standard notation (Box *et al.*, 1994) and satisfy the usual uniqueness and admissibility conditions.

Assume that the error term is $\varepsilon_t = b\eta_t$, where $\eta_t \sim f(\eta,\gamma)$, $f(\eta,\gamma)$ is the density function of one of the models considered in section 2 and

$$\gamma = (\theta_0, \phi_1,..,\phi_p, \Phi_1,..,\Phi_P, \theta_1,..,\theta_q, \Theta_1,...,\Theta_Q, b)'. \tag{9}$$

When η_t has a skew distribution, i.e. the corresponding model is *SN*, *SEP* or *St*, the expected value $E\left[b\eta_t\right] = b\mu_{\eta_t}$ can be different from zero. Thus (8) can also be written as:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Y_t = \tilde{\theta}_0 + \theta(B)\Theta(B^s)(\varepsilon_t - \mu_\varepsilon), \quad (10)$$

where $\tilde{\theta}_0 = \theta_0 + \theta(B)\Theta(B^s)\mu_\varepsilon$ and $\mu_\varepsilon = E\left[\varepsilon_t\right] \neq 0$.

Since the Jacobian of the transformation from $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ to $(Y_1, Y_2, \ldots, Y_n)$ is equal to one (Box *et al*. 1994, 200-215), the log-likelihood function is:

$$l(\gamma; y) = \sum_{t=1}^{n} ln\, f(\eta_t; \gamma). \quad (11)$$

Thus, an estimate $\hat{\gamma}$ of γ can be obtained by numerically maximizing (11) through optimisation routines currently available in many statistical packages (such as R or S-PLUS).

A main purpose, in the construction of ARIMA models, is forecasting future observations. Let $W_t = \nabla^d\nabla_s^D Y_t$, by (8) we obtain:

$$W_t = \phi(B)^{-1}\, \Phi^{-1}(B^s)\theta(B)\Theta(B^s)\varepsilon_t + \phi^{-1}(1)\, \Phi^{-1}(1)\theta_0. \quad (12)$$

Let $t = n + k$, where n is the current time period and $k \geq 1$; the forecast $\hat{W}(k)$ of W_{n+k} is given by:

$$\hat{W}(k) = E(W_{n+k} \left| I_n \right.), \quad (13)$$

where I_n is the information available at time t.

The forecast is obtained by replacing the unknown parameters and the values of ε_t, for $t \leq n$, by the estimates and the residuals respectively. When $f(\eta, \gamma)$ is symmetric, the values of ε_{n+k}, for $k \geq 1$, can be replaced by their expected values, i.e. zero, whereas if $f(\eta, \gamma)$ is a skewed density the values of ε_{n+k} can be replaced by the estimate $\hat{b}\ \hat{\mu}_{\eta_{n+k}}$ of their conditional expectation $E\left[\varepsilon_{n+k} \left| I_n \right.\right] = b\ E[\eta_{n+k}]$. Finally, $\hat{W}(k)$ can be used in the place of the future values W_{n+k}.

4. An application to the study of the energy demand

The internet site *http://www.nemmco.com.au/data/market_data.htm* (section Aggregate Price and Demand Data-Historical; QLD data) makes available the time series of the total energy demand in Queensland (Australia) recorded every 30 minutes. These data were used to build the series of the daily maximum energy demand since January 1, 1999 to February 28, 2001. Figure 1 shows the time series Y_t, for $t = 1, 2, \ldots, 790$.

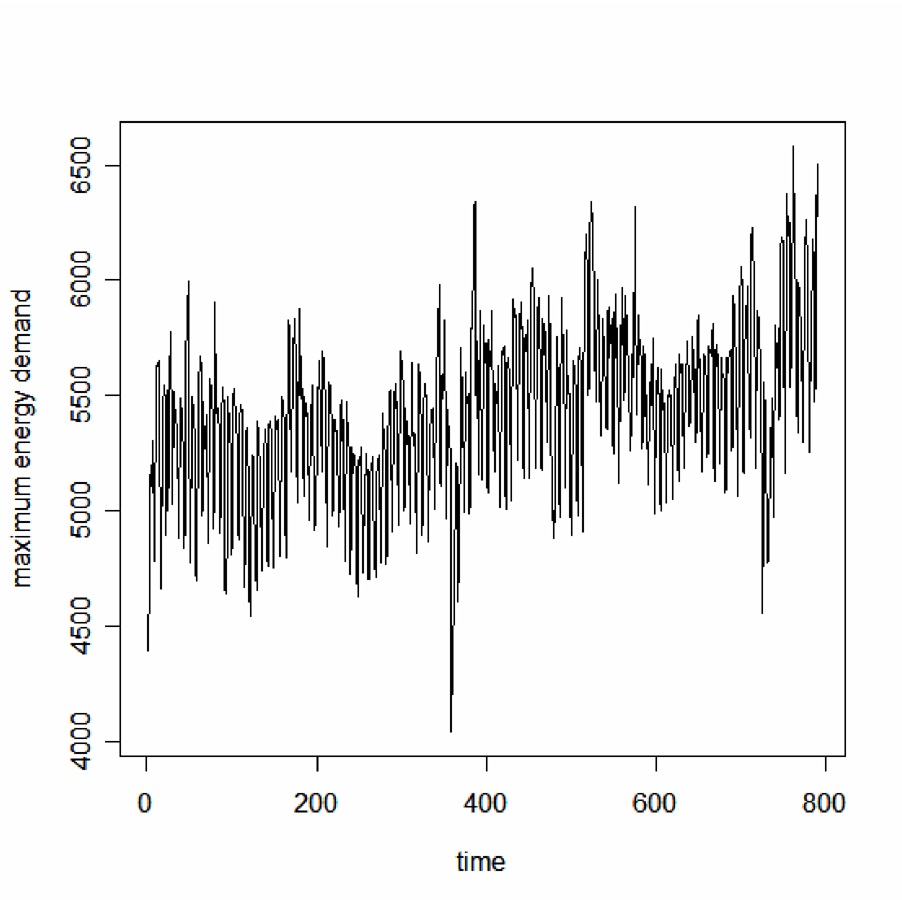

Figure 1. Time series of daily maximum energy demand

Through the analysis of the autocorrelation and partial autocorrelation function, the following model was identified for $X_t = \nabla_7 Y_t$:

$$(1-\phi_1 B - \phi_3 B^3)(1-\Phi_7 B^7)X_t = (1-\Theta_7 B^7)(\varepsilon_t - \mu_\varepsilon) + \theta_0. \qquad (14)$$

Furthermore, two dummy variables were introduced to model the calendar effects due to Christmas and New Year's Day as innovations outliers.

Under the normality assumption, the skewness and kurtosis of the residuals were -0.375 and 8.924 respectively, so that the Bowman and Shenton test rejected the hypothesis of normality at the 5% significance level. The need of a model able to deal with skewness and heavy tails was indeed evident.

Initially both the *SEP* and *St* distributions were considered for η_t. The log-likelihood function was maximized through the *nlminb* routine available in S-PLUS. The constraint $\alpha \in (1,2)$ was imposed in the case of the *SEP* distribution.

The estimates of the parameters, the log-likelihood function and the standard deviations of the error term, $\varepsilon_t = b\eta_t$, are reported in Table 1.

Table 1. Results for the Gaussian distribution and the flexible models

Models	$\hat{\phi}_1$	$\hat{\phi}_3$	$\hat{\Phi}_7$	$\hat{\Theta}_7$	$\hat{\sigma}_{\varepsilon_t}$	Log-likelihood
Gaussian	0.663 (0.028)	0.104 (0.029)	0.163 (0.036)	0.931 (0.001)	151.863	- 4974.554
SEP	0.693 (0.010)	0.056 (0.010)	0.057 (0.014)	0.943 (0.058)	156.179	- 4913.466
St	0.717 (0.028)	0.058 (0.026)	0.050 (0.035)	0.853 (0.018)	157.118	- 4909.414
EP	0.708 (0.016)	0.051 (0.014)	0.077 (0.022)	0.852 (0.013)	156.047	- 4915.431
t_g	0.717 (0.028)	0.059 (0.026)	0.053 (0.035)	0.850 (0.018)	156.536	- 4910.009

Since the *SEP* and *St* models have the same number of parameters, the value of the log-likelihood function at the estimated model can be used to compare the two distributional assumptions. As reported in Table 1 the difference between this two models was small and also the parameter estimates did not differ appreciably. However we preferred the *St* as it fitted slightly better.

Since, under the *St* distribution, the estimate of the parameter λ was close to zero ($\hat{\lambda} = 0.082 \quad (s.e. = 0.018)$), we tested the hypothesis of symmetry, $H_0 : \lambda = 0$, through the likelihood ratio test. The statistic (7) took value 0.30 with corresponding *p*-value 0.59, so that a symmetric model appeared appropriate. Consequently the t_g and the *EP* were considered.

By assuming a t_g distribution, the estimate of the degrees of freedom was $\hat{g}$ = 2.989 (*s.e.* = 0.370) and the value of the log-likelihood function was slightly higher than the value of the log-likelihood function under the *EP* distribution (see Table 1).

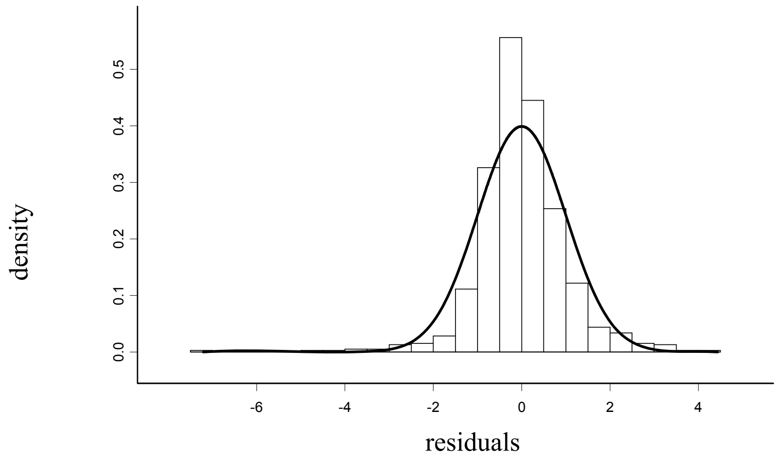

Figure 2. Histogram of the residuals from the Gaussian model and associated density

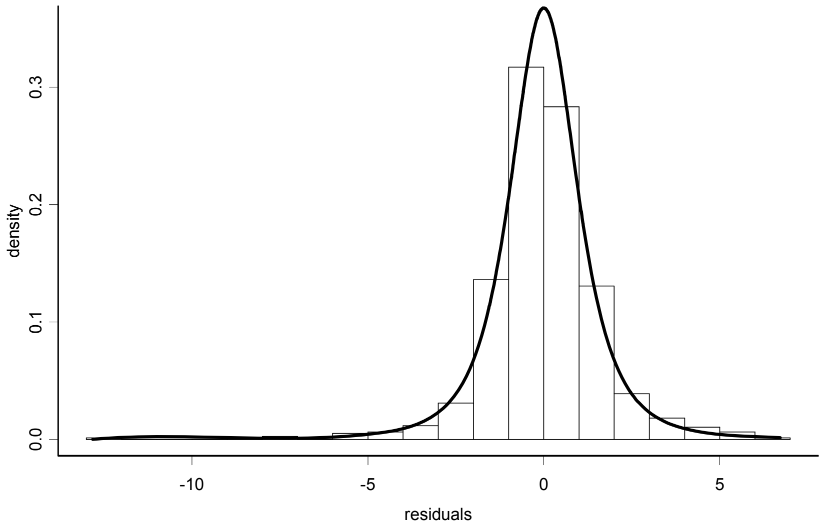

Figure 3. Histogram of the residuals from the t_g *model and associated density*

Figures 2 and 3 show the histograms and the fitted density of the residuals from a Gaussian and a t_g model. The better fit provided by the t_g distribution is outstanding.

In order to test the hypothesis of normality, a 95% confidence interval for the degrees of freedom g was computed. The upper limit of the confidence interval took value 3.77 so that the hypothesis of normality was rejected. This result is consistent with the estimate of the degrees of freedom which indicates a remarkable kurtosis in the data.

In order to evaluate the adequacy of the t_g model, Table 2 compares the forecasts of future observations obtained by (13), when either Gaussian or a t_g model are adopted, with the actual values for the week since March 1 2001 to March 7. Especially for the first three days, the absolute forecast error appears quite small under the t_g model with respect to the one obtained under the normal model.

Table 2. Daily maximum in energy demand (actual and forecasted values)

day	Y_{n+k}	Forecasted values by Gaussian model		Forecasted values by t_g model	
		$\hat{Y}_{n+k}$	$\left\lvert\frac{Y_{n+k}-\hat{Y}_{n+k}}{Y_{n+k}}\right\rvert*100$	$\hat{Y}_{n+k}$	$\left\lvert\frac{Y_{n+k}-\hat{Y}_{n+k}}{Y_{n+k}}\right\rvert*100$
1	6348.81	6233.94	1.81	6451.3	1.61
2	6047.49	6070.59	0.38	6086.31	0.64
3	5346.21	5536.51	3.56	5409.46	1.18
4	5326.19	5488.61	3.05	5455.53	2.43
5	5854.85	6034.00	3.06	6194.36	5.80
6	5902.34	6124.95	3.77	6374.23	7.99
7	6032.34	6025.79	0.11	6010.94	0.35

Since we are comparing the fitting of different models for forecasting purposes, it is important to check how the distributional assumptions about the white noise process modify the forecast functions of each of them. This may be accomplished in the time or the spectral domain.

In the time domain, for instance, it is possible to compute the AR metric, proposed by Piccolo (1984, 1990), which compares the structural dissimilarities among the estimated models by the Euclidean distance of the corresponding forecast functions. In this way, we would be able to assess if the modification induced by a different error distribution affects significantly the predictions.

Alternatively, by exploiting the linearity of the ARMA operators, we could analyse the parameter spectra of the estimated models in order to compare their similarity/dissimilarity over the angular frequency range. To be specific, we define the *parametric spectrum* for an admissible ARMA process, given by $A(B)Z_t = C(B)\varepsilon_t$, as the function:

$$g_Z(\omega) \propto \frac{1}{2\pi}\frac{\left|C\left(e^{-\iota\omega}\right)\right|^2}{\left|A\left(e^{-\iota\omega}\right)\right|^2} = \frac{\left|1-c_1e^{-\iota\omega}-c_2e^{-2\iota\omega}-\cdots\right|^2}{\left|1-a_1e^{-\iota\omega}-a_2e^{-2\iota\omega}-\cdots\right|^2},\ 0<\omega<\pi, \qquad (15)$$

where the proportionality constant is the variance of the white noise process ε_t.

Figure 4 shows the spectra derived from the three fitted models (Gaussian, *St* and t_g) whose estimated parameters, corresponding to the operators in (15), are reported in the following Table:

Models	$a_1 = \hat{\phi}_1$	$a_3 = \hat{\phi}_3$	$a_7 = \hat{\Phi}_7$	$a_8 = -\hat{\phi}_1 \hat{\Phi}_7$	$a_{10} = -\hat{\phi}_3 \hat{\Phi}_7$	$c_7 = \hat{\Theta}_7$
Gaussian	0.663	0.104	0.163	- 0.108	- 0.017	0.931
St	0.717	0.058	0.050	- 0.036	- 0.003	0.853
t_g	0.717	0.059	0.053	- 0.038	- 0.003	0.850

Although the spectra should be normalized for a more accurate comparison, we can add few comments to their plots, mainly based on the observed spectral shapes:

i) the parametric spectra are quite similar confirming that different distributional hypotheses on the errors do not change dramatically the structural components of the estimated models. In particular, the two fitted models obtained under the *St* and the t_g distributional assumption are almost indistinguishable. This result is consistent with the outcome of the test on the hypothesis of symmetry;

ii) most of the variability of the stationary component of the process is explained by an inertial component placed around the fundamental peak (3-4 weeks). This result is consistent with similar result in meteorological and related time series;

iii) the model we prefer in the forecasting experiment (implied by a t_g distribution) has a peak at 23 days instead of a peak at about 28 days which is obtained with the Gaussian model; thus, it gives somewhat more relevance to a shorter period component and should be more sensitive to medium term variations.

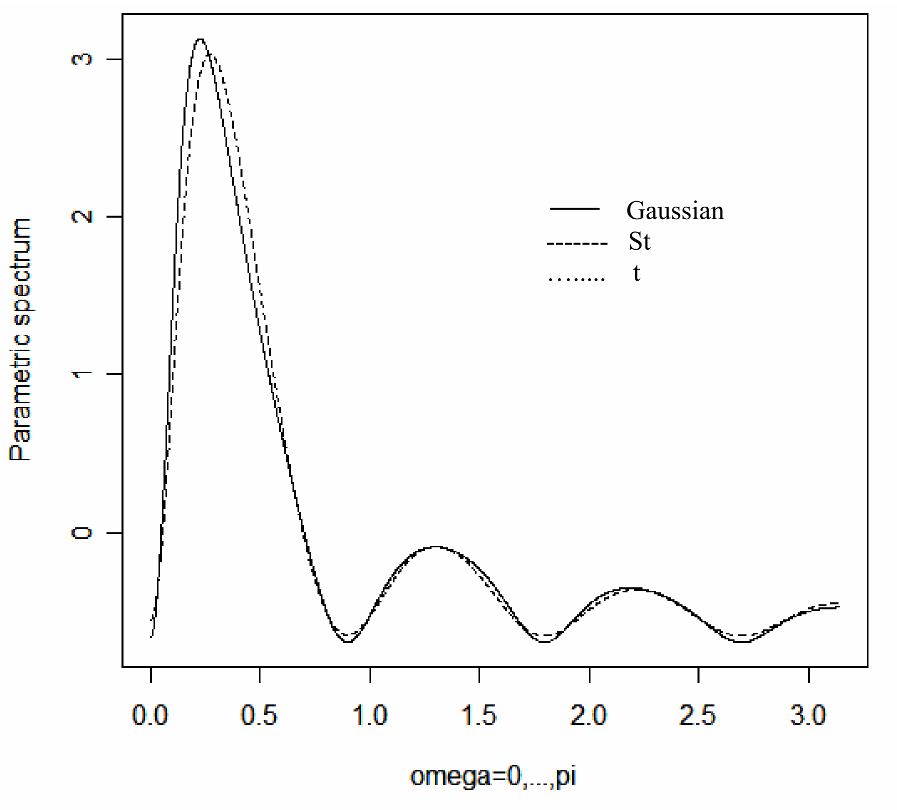

Figure 4. Parametric spectra estimated from the ARMA models

5. Concluding remarks

This paper shows the use of flexible models in time series estimation aimed to cope with skewness and/or heavy tails, and applies the proposed methodology in order to fit an ARIMA model to the energy demand in Queensland.

The flexible models considered here - the *EP*, *SEP,* Student *t* and Skew *t* distribution - are especially appealing since they allow to carry out inference based on the likelihood function while dealing with departures from normality.

Special attention is paid to the forecasting of future observations and to the selection of the distribution of the white noise.

Future research may compare the adoption of flexible models with alternative techniques such as the use of Box and Cox data transformations and robust techniques for parameters estimation.

Acknowledgments: The author wish to thank an anonymous referee, whose constructive comments greatly enhanced the paper. She also thanks Giuseppe Storti for valuable suggestions on an early version of the paper and Adelchi Azzalini for helpful advices. Research partially supported by MUR funds.

References

Andreotti A., Gagliardo F., Lauria D., Villacci D. (2000), Sviluppi nel controllo e nella gestione ottima delle reti di distribuzione, *L'Energia Elettrica*, Sezione "Ricerche", 77, 31-42.

Azzalini A. (1985), A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 12, 171-178.

Azzalini A. (1986), Further results on a class of distributions which includes the normal ones, *STATISTICA,* 46, 199-208.

Azzalini A., Capitanio A. (1999), Statistical applications of the multivariate skew-normal distribution, *Journal of the Royal Statistical Society Series B,* 61, 579-602.

Azzalini A., Capitanio A. (2003), Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistics Society Series B*, 65, 367-389.

Azzalini A., Dalla Valle A. (1996), The multivariate skew-normal distribution, *Biometrika*, 83, 715-726.

Barndorff-Nielsen O.E., Cox D.R. (1994), *Inference and Asymptotics*, Chapman & Hall, London.

Box G.E.P. (1953), A note on regions of kurtosis, *Biometrika,* 40, 465-468.

Box G.E.P., Jenkins G.M., Reinsel G.C. (1994), *Time series analysis: Forecasting and control*, Prentice Hall, New Jersey.

Capobianco R. (2000), Robustness aspects of the generalized normal distribution, *Quaderni di Statistica*, 2, 127-145.

Chiodi M. (1995), Generation of pseudo random variates from a normal distribution of order *p*, *Statistica Applicata*, 7, 401-415.

Chiogna M. (2005), A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution, *Statistical Methods and Applications*, 14, 331-341.

Denby L., Martin R.D. (1979), Robust estimation of the first-order autoregressive parameter, *Journal of the American Statistical Association*, 74, 140-146.

Di Ciccio T.J., Monti A.C. (2004), Inferential aspects of the skew exponential power distribution, *Journal of the American Statistical Association,* 99, 439-450.

Juberias G., Yunta R., Garcia M.J., Mendivil C. (1999), A new ARIMA model for hourly load forecasting, *Transmission and Distribution Conference,* IEEE, 314-319.

Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986), *Robust statistics: The approach based on influence functions,* J. Wiley & Sons, New York.

Hill M.A., Dixon W.J. (1982), Robustness in real life: a study of clinical laboratory data, *Biometrics*, 38, 377-396.