

Accepted Manuscript

Analytic expressions for predictive distributions in mixture autoregressive models

Georgi N. Boshnakov

PII: S0167-7152(09)00158-8
DOI: [10.1016/j.spl.2009.04.009](https://doi.org/10.1016/j.spl.2009.04.009)
Reference: STAPRO 5405

To appear in: *Statistics and Probability Letters*

Received date: 27 February 2009
Revised date: 15 April 2009
Accepted date: 17 April 2009

Please cite this article as: Boshnakov, G.N., Analytic expressions for predictive distributions in mixture autoregressive models. *Statistics and Probability Letters* (2009), doi:10.1016/j.spl.2009.04.009

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Analytic expressions for predictive distributions in mixture autoregressive models

Georgi N. Boshnakov^a

^a*School of Mathematics, The University of Manchester, Oxford Road,
Manchester M13 9PL, UK*

Abstract

We show that the distributions of the multi-step predictors in mixture autoregressive models are also mixtures and specify them analytically. In the case of mixtures of Gaussian or stable distributions the multi-step distributions can be obtained by simple arithmetic manipulation on components' parameters.

Key words: mixture autoregression, prediction, time series, stable distributions

1. Introduction

Mixture autoregressive models are studied by Wong and Li (2000) and Wong (1998). This is a relatively simple class of models having the attractive property that the shape of the conditional distribution of a forecast depends on the recent history of the process. In particular, it may have a varying number of modes over time.

Wong and Li (2000) note that the multi-step conditional distributions of predictors in MAR models are intractable analytically and resort to Monte Carlo simulations. We show that the distributions of the multi-step predictors in MAR models are also mixtures and specify them analytically.

We also demonstrate that (conditional) characteristic functions are the natural tool for calculations in this type of model. Conditional means and variances are not sufficient for prediction (even when they exist) in the presence of severe deviation from normality. Natural alternatives to them are the conditional characteristic functions. A characteristic function contains the entire distributional information. If needed, conditional mean, variance or other moments can be obtained from it. Moreover, unlike variances and means, characteristic functions always exist and hence provide more general results. In the important class of α -stable distributions with $0 < \alpha < 2$, moments of order greater than or equal to α do not exist (see Zolotarev

(1986)). Time series models based on stable distributions are studied actively, see Samorodnitsky and Taqqu (1994), Rachev and Mittnik (2000) and the references therein. Also, it is difficult to obtain manageable forms of the probability densities of the stable distributions (except in some special cases). On the other hand, their characteristic functions have a remarkably simple form and thus are the natural tool to use. Some of the most efficient methods for simulation of α -stable distributions are based on their characteristic functions, not densities.

2. The MAR model

Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_g)$ be a discrete distribution such that $\pi_k > 0$ for $k = 1, \dots, g$ and $\sum_{k=1}^g \pi_k = 1$. A process $\{y(t)\}$ is said to be a mixture autoregressive process with g components if the conditional distribution function of $y(t+1)$ given the information from the past of the process is a mixture of the following form:

$$\begin{aligned} F_{t+1|t}(x) &\equiv \Pr(y(t+1) \leq x | \mathcal{F}_t) \\ &= \sum_{k=1}^g \pi_k F_k \left(\frac{x - \phi_{k,0} - \sum_{i=1}^{p_k} \phi_{k,i} y(t+1-i)}{\sigma_k} \right), \end{aligned} \quad (1)$$

where for each $k = 1, \dots, g$, F_k is a distribution function, $\sigma_k > 0$, and $\phi_{k,i}$, $i = 0, 1, \dots, p_k$, are the autoregressive coefficients of the k th component. \mathcal{F}_t is the sigma field generated by the process $\{y(t)\}$ up to and including time t . It is convenient to set $p = \max_{1 \leq k \leq g} p_k$ and $\phi_{k,i} = 0$ for $i > p_k$. We assume also that $t > p$. We will denote by f_k and φ_k the probability density function and the characteristic function of F_k . The “location” and “scale” parameters of F_k are assumed to be zero and one, respectively;

The conditional density corresponding to $F_{t+1|t}(x)$ is

$$f_{t+1|t}(x) = \sum_{k=1}^g \frac{\pi_k}{\sigma_k} f_k \left(\frac{x - \phi_{k,0} - \sum_{i=1}^{p_k} \phi_{k,i} y(t+1-i)}{\sigma_k} \right). \quad (2)$$

For many applications it is sufficient to consider standard normal noise components, where $\varphi_k(s) = e^{-s^2/2}$. Another useful class is the class of α -stable distributions, e.g., $\varphi_k(s) = e^{-|s|^\alpha}$ is the characteristic function of a symmetric α -stable noise component.

At each time $t + 1$ one of g autoregressive-like equations is picked up at random to generate $y(t+1)$. So, an alternative description of the dynamics of

$\{y(t)\}$ is the following:

$$y(t+1) = \begin{cases} \phi_{1,0} + \sum_{i=1}^{p_1} \phi_{1,i}y(t+1-i) + \sigma_1\varepsilon_1(t+1), & \text{with probability } \pi_1, \\ \phi_{2,0} + \sum_{i=1}^{p_2} \phi_{2,i}y(t+1-i) + \sigma_2\varepsilon_2(t+1), & \text{with probability } \pi_2, \\ \dots \\ \phi_{g,0} + \sum_{i=1}^{p_g} \phi_{g,i}y(t+1-i) + \sigma_g\varepsilon_g(t+1), & \text{with probability } \pi_g, \end{cases} \quad (3)$$

where the distribution function of $\varepsilon_k(t)$, the k th noise component, is F_k for $k = 1, \dots, g$. With respect to the dependence structure we assume that $\varepsilon_k(t)$ are jointly independent and are also independent of past y s in the sense that for each t the σ -field generated by the set of random variables $\{\varepsilon_k(t+n), n \geq 1, 1 \leq k \leq g\}$ is independent of \mathcal{F}_t . Further, we assume that the choice of the component at time t does not depend on \mathcal{F}_{t-1} and $\{\varepsilon_k(t), t \geq 1, 1 \leq k \leq g\}$.

Equation (3) can be written in a more compact form, useful for calculations. Let for $k = 1, \dots, g$,

$$\mu_k(t+1) = \phi_{k,0} + \sum_{i=1}^{p_k} \phi_{k,i}y(t+1-i) \quad \text{and} \quad y_k(t+1) = \mu_k(t+1) + \sigma_k\varepsilon_k(t+1).$$

Let $\{z_t\}$ be an i.i.d. sequence of random variables with distribution $\boldsymbol{\pi}$, such that $\Pr\{z_t = k\} = \pi_k$ for $k = 1, \dots, g$. Then $y(t)$ is given by component k in Equation (3) if $z_t = k$. Now Equation (3) can be written as

$$y(t+1) = \phi_{z_{t+1},0} + \sum_{i=1}^p \phi_{z_{t+1},i}y(t+1-i) + \sigma_{z_{t+1}}\varepsilon_{z_{t+1}}(t+1) \quad (4)$$

$$= \mu_{z_{t+1}}(t+1) + \sigma_{z_{t+1}}\varepsilon_{z_{t+1}}(t+1). \quad (5)$$

3. Main Results

Equation (4) is well suited for one-step prediction since the random elements in it are either in \mathcal{F}_t or are independent of it. For longer horizons Equation (4) can be applied recursively to eliminate unobserved values of the process. The method is a natural extension of the similar procedure for autoregressive models. For example, for an AR(2) model we might get equations suitable for prediction two lags ahead by eliminating the unknown value at time $t+1$ as follows:

$$\begin{aligned} x(t+2) &= \phi_1x(t+1) + \phi_2x(t) + \varepsilon(t+2) \\ &= \phi_1(\phi_1x(t) + \phi_2x(t-1) + \varepsilon(t+1)) + \phi_2x(t) + \varepsilon(t+2) \\ &= (\phi_1^2 + \phi_2)x(t) + \phi_1\phi_2x(t-1) + \varepsilon(t+2) + \phi_1\varepsilon(t+1). \end{aligned}$$

The computations for the MAR model are given in Appendix B. The result is formulated in the following lemma.

Lemma 1. *The following representation holds for any $h \geq 1$:*

$$y(t+h) = \mu_{z_{t+h}, \dots, z_{t+1}}(t+h) + \sum_{i=0}^{h-1} \theta_i^{(z_{t+h}, \dots, z_{t+1})} \varepsilon_{z_{t+h-i}}(t+h-i), \quad (6)$$

where

$$\mu_{z_{t+h}, \dots, z_{t+1}}(t+h) = \beta_0^{(z_{t+h}, \dots, z_{t+1})} + \sum_{i=1}^p \beta_i^{(z_{t+h}, \dots, z_{t+1})} y_{(t+1-i)}, \quad (7)$$

with coefficients obtained from (4)–(5) for $h = 1$ and defined recursively for $h \geq 2$ by

$$\beta_i^{(z_{t+h}, \dots, z_t)} = \begin{cases} \beta_0^{(z_{t+h}, \dots, z_{t+1})} + \beta_1^{(z_{t+h}, \dots, z_{t+1})} \phi_{z_t, 0} & \text{for } i = 0, \\ \beta_{i+1}^{(z_{t+h}, \dots, z_{t+1})} + \beta_1^{(z_{t+h}, \dots, z_{t+1})} \phi_{z_t, i} & \text{for } i = 1, \dots, p-1, \\ \beta_1^{(z_{t+h}, \dots, z_{t+1})} \phi_{z_t, p} & \text{for } i = p. \end{cases} \quad (8)$$

$$\theta_i^{(z_{t+h}, \dots, z_t)} = \begin{cases} \theta_i^{(z_{t+h}, \dots, z_{t+1})} & \text{for } i = 0, \dots, h-1, \\ \beta_1^{(z_{t+h}, \dots, z_{t+1})} \sigma_{z_t} & \text{for } i = h. \end{cases} \quad (9)$$

The coefficients $\theta_i^{(z_{t+h}, \dots, z_{t+1})}$ and $\beta_i^{(z_{t+h}, \dots, z_{t+1})}$ are functions of z_{t+h}, \dots, z_{t+1} , and the parameters $\phi_{k,i}$ and σ_k of the MAR model. For example, for $h = 2$ we have

$$\beta_i^{(k,l)} = \begin{cases} \phi_{k,0} + \phi_{k,1} \phi_{l,0} & \text{for } i = 0, \\ \phi_{k,i+1} + \phi_{k,1} \phi_{l,i} & \text{for } 1 \leq i \leq p-1, \\ \phi_{k,1} \phi_{l,p} & \text{for } i = p, \end{cases} \quad \theta_i^{(k,l)} = \begin{cases} \sigma_k & \text{for } i = 0, \\ \phi_{k,1} \phi_{l,i} & \text{for } i = 1. \end{cases} \quad (10)$$

The quantity $\mu_k(t+1)$ has the meaning of a location parameter of the (conditional) distribution of the k th component of the MAR model. In particular, $\mu_k(t+1) = E(y_k(t+1)|\mathcal{F}_t)$ when the mean $E\varepsilon_k(t+1)$ exists (see Appendix A). The variables $y_k(t+1)$ are somewhat artificial but an important consequence of the last equation is that $\mu_k(t+1) = E(y(t+1)|\mathcal{F}_t, z_{t+1} = k)$, for $k = 1, \dots, g$. More generally, for any horizon $h \geq 1$ and any h -tuple k_1, \dots, k_h such that $1 \leq k_i \leq g$, for $1 \leq i \leq h$,

$$\begin{aligned} \mu_{k_1, \dots, k_h}(t+h) &= E(y(t+h)|\mathcal{F}_t, z_{t+h} = k_1, z_{t+h-1} = k_2, \dots, z_{t+1} = k_h) \\ &= \sum_{i=1}^p \beta_i^{(k_1, \dots, k_h)} y_{(t+h-i)} + \beta_0^{(k_1, \dots, k_h)}, \end{aligned}$$

when the means of the noise components exist but we should stress again that our results do not rely on the existence of any moments.

The (one-step) conditional characteristic function of $y(t)$ is

$$\varphi_{t+1|t}(s) \equiv E(e^{isy(t+1)}|\mathcal{F}_t) = \sum_{k=1}^g \pi_k e^{is\mu_k(t+1)} \varphi_k(\sigma_k s). \quad (11)$$

The conditional characteristic functions for longer horizons can be obtained recursively using Equation (6), Lemma 1, and the iterated rule for conditional expectations (see Appendix B)

$$\begin{aligned} \varphi_{t+h|t}(s) &\equiv E(e^{isy(t+h)}|\mathcal{F}_t) = E(E(e^{isy(t+h)}|\mathcal{F}_t, z_{t+h}, \dots, z_{t+1})|\mathcal{F}_{t-h}) \\ &= \sum_{k_1, \dots, k_h=1}^g (\pi_{k_1} \cdots \pi_{k_h}) e^{is(\mu_{k_1, \dots, k_h}(t+h))} \prod_{i=0}^{h-1} \varphi_{k_{h-i}}(\theta_{h-i}^{(k_1, \dots, k_h)} s), \end{aligned} \quad (12)$$

where

$$\mu_{k_1, \dots, k_h}(t+h) = \sum_{i=1}^p \beta_i^{(k_1, \dots, k_h)} y_{(t+1-i)} + \beta_0^{(k_1, \dots, k_h)}.$$

For $h = 2$ this gives

$$\varphi_{t+2|t}(s) = \sum_{k=1}^g \sum_{l=1}^g \pi_k \pi_l e^{is\mu_{k,l}(t+2)} \varphi_k(\sigma_k s) \varphi_l(\phi_{k,1} \sigma_l s), \quad (13)$$

where $\mu_{k,l}(t+2)$ is a linear combination of $y_{(t+2-i)}$, $i \geq 2$ (see Equation (20)). Thus, the conditional distribution of $y_{(t+2)}$ given \mathcal{F}_t is a mixture of g^2 components with mixing probabilities $\pi_k \pi_l$. For Gaussian noise components we have

$$\varphi_k(\sigma_k s) \varphi_l(\phi_{k,1} \sigma_l s) = e^{-(\sigma_k^2 + \sigma_l^2 \phi_{k,1}^2) s^2 / 2},$$

which together with Equation (13) shows that the conditional distribution of the two-step predictor is a mixture of normal distributions with variances $\sigma_k^2 + \sigma_l^2 \phi_{k,1}^2$ and means $\mu_{k,l}(t+2)$.

It is worth noting that the calculations for the general case are not more complicated than that for the Gaussian case. Joint conditional distributions can be found using the same technique.

The above results are summarised in the following theorem. Its interpretation and some corollaries are discussed in the remarks after it. Recall that $\varepsilon_k(t)$ and $y_k(t)$ are the k th noise component and the k th component of the MAR model, respectively.

Theorem 1. *For each $h \geq 1$ the conditional characteristic function, $\varphi_{t+h|t}(s) \equiv E(e^{isy(t+h)}|\mathcal{F}_t)$, of the h -step predictor at time t of the MAR process (3) is given by Equation (12).*

Remark 1 The expression on the right-hand side of (12) represents a mixture. Hence, the conditional distribution of the h -step predictor at time t of the MAR process (3) is a mixture of g^h components. The probability associated with the (k_1, \dots, k_h) th component is $\pi_{k_1} \cdots \pi_{k_h}$ and its conditional characteristic function is $e^{is\mu_{k_1, \dots, k_h}(t+h)} \prod_{i=0}^{h-1} \varphi_{k_{h-i}}(\theta_{h-i}^{(k_1, \dots, k_h)} s)$.

Remark 2 The exponent, $e^{is(\dots)}$, is the only term in the characteristic function of the (k_1, \dots, k_h) th component that depends on the past of the process $\{y(t)\}$. Since such a factor corresponds to a shift of the distribution, it follows that only the location parameter of the conditional distribution of the (k_1, \dots, k_h) th component of the h -step predictor depends on the past of the process $\{y(t)\}$.

It should be stressed however that this qualitative property holds for the conditional distributions of the individual components, not for the conditional distribution of $y(t+h)$ itself. The shape and other properties of the latter (such as $\text{Var}(y(t+h)|\mathcal{F}_t)$) do depend on $y(t)$, $y(t-1)$, etc.

Since we have assumed that the location parameters of the noise components are zero, it is tempting to state that the location parameter of the distribution of the (k_1, \dots, k_h) th component of the h -step predictor at time t is $\mu_{k_1, \dots, k_h}(t+h)$, but this is not always the case (see Equation (15) below for an example).

Remark 3 In general, the distributions of the components of the h -step predictors depend both on h and on the noise distributions in the specification of the MAR model and do not necessarily belong to a common class. However, the product of the noise characteristic functions in Equation (12) shows that if the components of the MAR model (3) are normal, then for each $h \geq 1$ the components of the h -step ahead predictor at time t are also normal. The conditional mean and variance of the (k_1, \dots, k_h) th component are $\mu_{k_1, \dots, k_h}(t+h)$ and $\sum_{i=0}^{h-1} (\theta_{h-i}^{(k_1, \dots, k_h)})^2$, respectively. For the two-step predictors (i.e., for $h = 2$) these simplify to $\mu_{k,l}(t+2)$ and $\sigma_k^2 + \sigma_l^2 \phi_{k,1}^2$, respectively.

Remark 4 If $\varepsilon_k(t)$ are α -stable, $S_\alpha(1, \beta_k, 0)$, $k = 1, \dots, g$, then the product $\prod_{i=0}^{h-1} \varphi_{k_{h-i}}(\theta_{h-i}^{(k_1, \dots, k_h)} s)$ is the characteristic function of an α -stable distribution. So, if the components of the MAR model (3) are α -stable, $S_\alpha(1, \beta_i, 0)$, $i = 1, \dots, g$, $0 < \alpha \leq 2$, with characteristic functions (see e.g. Lukacs, 1979, Theorem 5.7.3),

$$\varphi_k(s) = e^{-|s|^\alpha (1 - i\beta_k \text{sign}(s)w(|s|, \alpha))}, \text{ where } w(|s|, \alpha) = \begin{cases} \tan \frac{\pi\alpha}{2}, & \alpha \neq 1, \\ \frac{2}{\pi} \log|s|, & \alpha = 1, \end{cases} \quad (14)$$

then for each $h \geq 1$ the components of the h -step ahead predictor at time t are α -stable. The distribution of the (k_1, \dots, k_h) th component is $S_\alpha(\sigma, \beta, \mu)$,

where

$$\mu = \begin{cases} \mu_{k_1, \dots, k_h}(t+h) & \text{if } \alpha \neq 1, \\ \mu_{k_1, \dots, k_h}(t+h) + \frac{2}{\pi} \sum_{i=0}^{h-1} \beta_{k_{h-i}} \theta_{h-i}^{(k_1, \dots, k_h)} \log |\theta_{h-i}^{(k_1, \dots, k_h)}| & \text{if } \alpha = 1. \end{cases} \quad (15)$$

$$\sigma = \left(\sum_{i=0}^{h-1} |\theta_{h-i}^{(k_1, \dots, k_h)}|^\alpha \right)^{1/\alpha}, \quad \beta = \frac{\sum_{i=0}^{h-1} \beta_{k_{h-i}} |\theta_{h-i}^{(k_1, \dots, k_h)}|^\alpha \text{sign}(\theta_{h-i}^{(k_1, \dots, k_h)})}{\sum_{i=0}^{h-1} |\theta_{h-i}^{(k_1, \dots, k_h)}|^\alpha}. \quad (16)$$

4. Example

Wong and Li (2000) built a mixture autoregressive model for the daily IBM stock closing price data (Box and Jenkins, 1976) with the following parameters: $g = 3$; $\pi = (0.5439, 0.4176, 0.0385)$; $\sigma_1 = 4.8227$, $\sigma_2 = 6.0082$, $\sigma_3 = 18.1716$; $p_1 = 2$, $p_2 = 2$, $p_3 = 1$; $\phi_{k,0} = 0$, $k = 1, 2, 3$; $\phi_{1,1} = 0.6792$, $\phi_{1,2} = 0.3208$, $\phi_{2,1} = 1.6711$, $\phi_{2,2} = -0.6711$, $\phi_{3,1} = 1$. Putting these parameters into (2) gives the one-step conditional density,

$$\begin{aligned} f_{t+1|t}(x) = & 0.000845235e^{-0.0015142(x-y_t)^2} \\ & + 0.0449924e^{-0.0214976(x-0.6792y_t-0.3208y_{t-1})^2} \\ & + 0.0277285e^{-0.013851(x-1.6711y_t+0.6711y_{t-1})^2}. \end{aligned} \quad (17)$$

From Theorem 1 (see also Equation (13)) the two step conditional density is

$$\begin{aligned} f_{t+2|t}(x) = & 0.0000230104e^{-0.000757101(x-y_t)^2} \\ & + 0.000530972e^{-0.00201992(x-1.y_t)^2} \\ & + 0.000264531e^{-0.000850474(x-1.y_t)^2} \\ & + 0.0104655e^{-0.00666972(x-0.463911y_t-0.536089y_{t-1})^2} \\ & + 0.000444341e^{-0.00141457(x-0.6792y_t-0.3208y_{t-1})^2} \\ & + 0.0188846e^{-0.0128023(x-0.782113y_t-0.217887y_{t-1})^2} \\ & + 0.0131094e^{-0.0104654(x-1.45581y_t+0.455811y_{t-1})^2} \\ & + 0.000335127e^{-0.00136498(x-1.6711y_t+0.6711y_{t-1})^2} \\ & + 0.00708503e^{-0.00518551(x-2.12148y_t+1.12148y_{t-1})^2}. \end{aligned} \quad (18)$$

As an illustration, for this series we have $y(258) = 361$, $y(257) = 399$ and Table 4 gives the modes of the predictive densities, $f_{258+h|258}(x)$, for horizons 2–5 and the values (or “heights”) of the densities at the modes. The prefix

Table 1: Modes of the predictive distributions discussed in the example. (M1, M2, M3) and (WLM1, WLM2, WLM3) are modes obtained by the analytical formulae and the Monte Carlo method of Wong-Li, respectively. “M1 height”, etc., stand for the values of the predictive pdf’s at the corresponding modes.

Horizon	2	3	4	5
Number of Modes	3	2	2	1
M1	370.136	371.174	370.974	370.690
WLM1	370.089	371.495	372.399	369.426
M2	343.48	342.25	343.870	
WLM2	342.994	342.238	343.048	
M3	318.854			
WLM3	317.728			
M1 height	0.0260764	0.0208125	0.0180173	0.0162017
WLM1 height	0.0257348	0.0227771	0.0195596	0.0164599
M2 height	0.0158331	0.0119786	0.0111955	
WLM2 height	0.0156722	0.0099772	0.0110598	
M3 height	0.0062759			
WLM3 height	0.0061026			

WL is used for quantities derived from the approximate densities computed by the method of Wong-Li based on 10^3 simulated values. In this example the dominant mode remains quite stable for several prediction horizons. The “weakest” mode (M3) disappears after $h = 2$ and the second mode after $h = 4$. Obviously, the results from the Wong-Li method vary slightly from run to run but they agree well with the exact results and increasing the number of simulations to 10^5 gives predictive densities (not shown here) essentially identical to the exact ones. The second mode, which is not very well pronounced, was not present on some runs of the Wong-Li method.

5. Conclusion

We have shown that the multi-step predictors for a mixture autoregressive model are mixtures. Moreover, when the noise components are normal or stable the predictors remain mixtures of normal, respectively stable, distributions for all horizons. We have also demonstrated that the conditional characteristic function is a useful and intuitive instrument for analysis of MAR models.

6. Acknowledgements

I thank an anonymous referee for the constructive comments and suggestions.

A. Derivation of the representation of $y(t)$ given by Lemma 1

To clarify the ideas we start with the case $h = 2$. It is necessary to eliminate $y(t-1)$ from the expression for $y(t)$. From Equations (4)-(5) we get

$$\begin{aligned} y(t) &= \mu_{z_t}(t) + \sigma_{z_t} \varepsilon_{z_t}(t) \\ &= (\mu_{z_t}(t) - \phi_{z_t,1} y(t-1)) + \phi_{z_t,1} y(t-1) + \sigma_{z_t} \varepsilon_{z_t}(t) \\ &= \mu_{z_t, z_{t-1}}(t) + \phi_{z_t,1} \sigma_{z_{t-1}} \varepsilon_{z_{t-1}}(t-1) + \sigma_{z_t} \varepsilon_{z_t}(t), \end{aligned} \quad (19)$$

where

$$\begin{aligned} \mu_{k,l}(t) &\equiv \mu_k(t) - \phi_{k,1} y(t-1) + \phi_{k,1} \mu_l(t-1) \\ &= \phi_{k,0} + \phi_{k,1} \phi_{l,0} + \sum_{i=1}^{p-1} (\phi_{k,i+1} + \phi_{k,1} \phi_{l,i}) y(t-1-i) + \phi_{k,1} \phi_{l,p} y(t-1-p). \end{aligned} \quad (20)$$

Equations (19)-(20) provide the required representation of $y(t)$ in the form (6) with coefficients given by Equation (10).

For higher lags we continue by induction. Assume that $y(t)$ is represented in the form (6)-(7) for some $h \geq 2$. From Equation (6) we get

$$\begin{aligned} y(t) &= \mu_{z_{t:t-h+1}}(t) + \sum_{i=0}^{h-1} \theta_i^{(z_{t:t-h+1})} \varepsilon_{z_{t-i}}(t-i) \\ &= (\mu_{z_{t:t-h+1}}(t) - \beta_1^{(z_{t:t-h+1})} y(t-h)) + \beta_1^{(z_{t:t-h+1})} y(t-h) \\ &\quad + \sum_{i=0}^{h-1} \theta_i^{(z_{t:t-h+1})} \varepsilon_{z_{t-i}}(t-i) \\ &= (\mu_{z_{t:t-h+1}}(t) - \beta_1^{(z_{t:t-h+1})} y(t-h)) + \beta_1^{(z_{t:t-h+1})} \mu_{z_{t-h}}(t-h) \\ &\quad + \beta_1^{(z_{t:t-h+1})} \sigma_{z_{t-h}} \varepsilon_{z_{t-h}}(t-h) + \sum_{i=0}^{h-1} \theta_i^{(z_{t:t-h+1})} \varepsilon_{z_{t-i}}(t-i). \end{aligned} \quad (21)$$

Our aim is to eliminate $y(t-h)$. From Equation (7) we get

$$\mu_{z_{t:t-h+1}}(t) - \beta_1^{(z_{t:t-h+1})} y(t-h) = \beta_0^{(z_{t:t-h+1})} + \sum_{i=1}^{p-1} \beta_{i+1}^{(z_{t:t-h+1})} y(t-h-i). \quad (22)$$

From Equation (5) we get

$$\beta_1^{(z_{t:t-h+1})} y(t-h) = \beta_1^{(z_{t:t-h+1})} (\mu_{z_{t-h}}(t-h) + \sigma_{z_{t-h}} \varepsilon_{z_{t-h}}(t-h)) \quad (23)$$

$$\mu_{z_{t-h}}(t-h) = \phi_{z_{t-h},0} + \sum_{i=1}^p \phi_{z_{t-h},i} y(t-h-i). \quad (24)$$

From Equation (21) we can now see that the required representation for $y(t)$ for lag $h+1$ is

$$y(t) = \mu_{z_t, \dots, z_{t-h}}(t) + \beta_1^{(z_{t:t-h+1})} \sigma_{z_{t-h}} \varepsilon_{z_{t-h}}(t-h) + \sum_{i=0}^{h-1} \theta_i^{(z_{t:t-h+1})} \varepsilon_{z_{t-i}}(t-i), \quad (25)$$

where

$$\mu_{z_t, \dots, z_{t-h}}(t) = (\mu_{z_t, \dots, z_{t-h+1}}(t) - \beta_1^{(z_{t:t-h+1})} y(t-h)) + \beta_1^{(z_{t:t-h+1})} \mu_{z_{t-h}}(t-h),$$

since the last expression is a linear combination of $y(t-i)$, $i \geq h+1$. To obtain formulae (8)–(9) we use Equations (21), (23) and (24) to simplify the last equation to

$$\begin{aligned} \mu_{z_t, \dots, z_{t-h}}(t) &= \beta_0^{(z_{t:t-h+1})} + \sum_{i=1}^{p-1} \beta_{i+1}^{(z_{t:t-h+1})} y(t-h-i) \\ &\quad + \beta_1^{(z_{t:t-h+1})} \phi_{z_{t-h},0} + \sum_{i=1}^p \beta_1^{(z_{t:t-h+1})} \phi_{z_{t-h},i} y(t-h-i). \end{aligned} \quad (26)$$

Comparing Equations (25) and (26) with (6)–(9) we get the desired result.

When the mean $E \varepsilon_k(t)$ exists we have

$$E(y_k(t) | \mathcal{F}_{t-1}) = E(\mu_k(t) + \sigma_k \varepsilon_k(t) | \mathcal{F}_{t-1}) = \mu_k(t).$$

B. Derivation of the conditional characteristic functions

Using Equation (6) we get for any $h \geq 2$

$$\begin{aligned} \varphi_{t|t-h}(s) &\equiv E(e^{isy(t)} | \mathcal{F}_{t-h}) = E(E(e^{isy(t)} | \mathcal{F}_{t-h}, z_t, \dots, z_{t+1-h}) | \mathcal{F}_{t-h}) \\ &= E(e^{is\mu_{z_t:t+1-h}}(t) E(e^{is(\sum_{m=0}^{h-1} \theta_{h-m}^{(z_{t:t+1-h})} \varepsilon_{z_{t-m}}(t-m))} | \mathcal{F}_{t-h}, z_t, \dots, z_{t+1-h}) | \mathcal{F}_{t-h}) \\ &= \sum_{k_1, \dots, k_h=1}^g (\pi_{k_1} \cdots \pi_{k_h}) e^{is\mu_{k_1:h}}(t) \prod_{m=0}^{h-1} \varphi_{k_{h-m}}(\theta_{h-m}^{(k_1:h)} s). \end{aligned}$$

References

- Box, G. E., Jenkins, G. M., 1976. Time series analysis: Forecasting and control. Rev. ed. Holden-Day Series in Time Series Analysis. San Francisco etc.: Holden-Day.
- Lukacs, E., 1979. Characteristic functions. (Kharakteristicheskie funktsii). Transl. from the English by V. M. Zolotarev. Moskva: Nauka.
- Rachev, S., Mittnik, S., 2000. Stable Paretian models in finance. Chichester: Wiley.
- Samorodnitsky, G., Taqqu, M. S., 1994. Stable non-Gaussian random processes: stochastic models with infinite variance. Stochastic Modeling. New York, NY: Chapman & Hall.
- Wong, C. S., 1998. Statistical inference for some nonlinear time series models. Ph.D. thesis, University of Hong Kong, Hong Kong.
- Wong, C. S., Li, W. K., 2000. On a mixture autoregressive model. J. R. Stat. Soc., Ser. B, Stat. Methodol. 62 (1), 95–115.
- Zolotarev, V., 1986. One-dimensional stable distributions. Transl. from the Russian by H. H. McFaden, ed. by Ben Silver. Translations of Mathematical Monographs, Vol. 65. Providence, Rhode Island: American Mathematical Society.