



HAL
open science

The combined INAR() models for time series of counts

Christian H. Weiss

► **To cite this version:**

Christian H. Weiss. The combined INAR() models for time series of counts. *Statistics and Probability Letters*, 2010, 78 (13), pp.1817. 10.1016/j.spl.2008.01.036 . hal-00645372

HAL Id: hal-00645372

<https://hal.science/hal-00645372>

Submitted on 28 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

The combined INAR(p) models for time series of counts

Christian H. Weiß

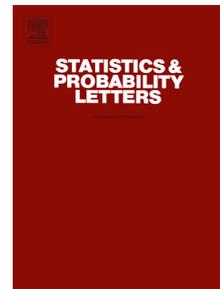
PII: S0167-7152(08)00045-X
DOI: 10.1016/j.spl.2008.01.036
Reference: STAPRO 4909

To appear in: *Statistics and Probability Letters*

Received date: 24 October 2007
Revised date: 14 January 2008
Accepted date: 17 January 2008

Please cite this article as: Weiß, C.H., The combined INAR(p) models for time series of counts. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.01.036

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.



The Combined INAR(p) Models for Time Series of Counts

Christian H. Weiß

Full address:

Christian H. Weiß
University of Würzburg
Institute of Mathematics, Department of Statistics
Am Hubland
D-97074 Würzburg
Germany
Email: christian.weiss@mathematik.uni-wuerzburg.de
Phone: +49-931-888-4968
Fax: +49-931-888-4949

Abstract: A new AR(p) model for time series of counts is investigated, the possible marginal distributions of which are those of the DSD family. We determine the autocorrelation structure of the whole model family and analyze two important special cases. A real-data example demonstrates the practical relevance of the new model family.

Key words: INAR(p) model; EAR(p) model; DSD distributions.

The Combined INAR(p) Models for Time Series of Counts

Christian H. Weiß

University of Würzburg, Inst. of Mathematics, Dep. of Statistics, Germany.

Abstract

A new AR(p) model for time series of counts is investigated, the possible marginal distributions of which are those of the DSD family. We determine the autocorrelation structure of the whole model family and analyze two important special cases. A real-data example demonstrates the practical relevance of the new model family.

Key words: INAR(p) model; EAR(p) model; DSD distributions.

1 Introduction

The first integer-valued ARMA (INARMA) model has been proposed by McKenzie (1985): the INAR(1) model as a counterpart to the usual AR(1) model. It is based on a probabilistic operation called *binomial thinning*, which proved to be an adequate alternative to scalar multiplication for integer-valued time series. If X is a discrete random variable with range $\{0, \dots, n\}$ or \mathbb{N}_0 , then the random variable $\alpha \circ X := \sum_{i=1}^N Y_i$, where Y_i (*counting series*) are independent Bernoulli trials according to $B(1, \alpha)$, independent of X , is said to arise from X by *binomial thinning*. ‘ \circ ’ is called the *binomial thinning operator*. Especially, one obtains $E[\alpha \circ X] = \alpha \cdot E[X]$, which justifies to ‘replace’ the scalar multiplication in the usual ARMA recursion by the probabilistic operation of binomial thinning.

Binomial thinning was originally introduced by Steutel & van Harn (1979) to adapt the terms of *self-decomposability* and *stability* for integer-valued time series. A random variable X with range \mathbb{N}_0 is *discrete self-decomposable (DSD)*

Email address: christian.weiss@mathematik.uni-wuerzburg.de (Christian H. Weiß).

if for any $\alpha \in (0; 1)$, there exists a random variable ϵ_α with range \mathbb{N}_0 , independent of $\alpha \circ X$, such that X and $\alpha \circ X + \epsilon_\alpha$ have the same distribution. Since the INAR(1) model is defined by the recursion $X_t = \alpha \circ X_{t-1} + \epsilon_t$, the DSD distributions are the possible marginal distributions of a stationary INAR(1) process. Many important distributions, including negative binomial, Poisson and generalized Poisson distribution, belong to this class of DSD distributions, see Zhu & Joe (2003).

While the INAR(1) recursion involves one thinning operation only, the counterpart of the general AR(p) model needs p thinning operations:

$$X_t = \alpha_1 \circ_t X_{t-1} + \dots + \alpha_p \circ_t X_{t-p} + \epsilon_t.$$

The time index t below the thinning operation indicates that the corresponding thinning is involved in defining X_t . Since the thinning operations are *probabilistic*, the joint distribution of $(\alpha_1 \circ_{t+1} X_t, \dots, \alpha_p \circ_{t+p} X_t)$ has to be considered, leading to different types of INAR(p) models: Alzaid & Al-Osh (1990) assume a conditional multinomial distribution, Du & Li (1991) require conditional independence. The INAR(p) models are quite complex and difficult to interpret for $p \geq 2$. In addition, the choice of appropriate marginal distributions for $(X_t)_{\mathbb{Z}}$ and $(\epsilon_t)_{\mathbb{Z}}$ is problematic. These difficulties did not arise in the case of the INAR(1) model, where any DSD distribution is a possible marginal process distribution. To overcome these difficulties, Zhu & Joe (2006) recently proposed an alternative definition of a p^{th} order autoregressive model. They combined the EAR(p) model of Lawrance & Lewis (1980) and the INAR(1) model discussed before. Hence, the resulting model will be called *combined INAR(p) (CINAR(p)) model* in this text.

Definition 1 (CINAR(p) Model) Let $(\epsilon_t)_{\mathbb{Z}}$ be an i.i.d. process with range \mathbb{N}_0 , and $\alpha \in (0; 1)$. Let $(\mathbf{D}_t)_{\mathbb{Z}}$ be an i.i.d. process of ‘decision’ random variables $\mathbf{D}_t = (D_{t,1}, \dots, D_{t,p}) \sim \text{MULT}(1; \phi_1, \dots, \phi_p)$, independent of $(\epsilon_t)_{\mathbb{Z}}$. A process $(X_t)_{\mathbb{Z}}$, which follows the recursion

$$X_t = D_{t,1} \cdot (\alpha \circ_t X_{t-1}) + \dots + D_{t,p} \cdot (\alpha \circ_t X_{t-p}) + \epsilon_t,$$

is called an CINAR(p) process if

- the thinnings at time t are performed independently of each other, of $(\epsilon_t)_{\mathbb{Z}}$ and $(\mathbf{D}_t)_{\mathbb{Z}}$, and the thinnings of X_t independent of $(X_s)_{s < t}$,
- ϵ_t and \mathbf{D}_t are independent of all X_s and $\alpha \circ_{s+j} X_s$ with $s < t$, $j = 1, \dots, p$,
- the conditional probability $P(\alpha \circ_{t+1} X_t, \dots, \alpha \circ_{t+p} X_t \mid X_t = x_t, \mathcal{H}_{t-1})$ equals $P(\alpha \circ_{t+1} X_t, \dots, \alpha \circ_{t+p} X_t \mid X_t = x_t)$, where \mathcal{H}_{t-1} abbreviates the process history of all X_s and $\alpha \circ_{s+j} X_s$ for $s \leq t-1$ and $j = 1, \dots, p$.

Zhu & Joe (2006) only analyzed the special CINAR(2) model of Example 4. In the following Section 2, we shall investigate the whole model family. In

particular, we derive a set of Yule-Walker equations for the autocovariance function. These equations simplify in two important special cases, including the results of Zhu & Joe (2006) for $p = 2$. The real-data example of Section 3 demonstrates the practical relevance of the new model family. Finally, we conclude in Section 4.

2 Properties of CINAR(p) Models

The CINAR(p) recursion of Definition 1 states that X_t is either equal to $\alpha \circ_t X_{t-1} + \epsilon_t$ with probability ϕ_1, \dots , or to $\alpha \circ_t X_{t-p} + \epsilon_t$ with probability ϕ_p . Here, the time index t below the thinning operation indicates that the corresponding thinning is involved in defining X_t , but it does *not* necessarily exclude that all thinnings of a process variable X_s are identical, see Section 2.1 below. The main advantage of the CINAR(p) model against the INAR(p) model gets clear considering the marginal process distribution: If $(X_t)_{\mathbb{Z}}$ is a stationary CINAR(p) process, then its probability generating function (pgf) has to fulfill

$$\begin{aligned} p_X(z) &= E[E[z^{X_t} \mid \mathbf{D}_t]] = \sum_{i=1}^p \phi_i \cdot E[z^{\alpha \circ_t X_{t-i} + \epsilon_t}] \\ &= \sum_{i=1}^p \phi_i \cdot p_X(1 - \alpha + \alpha z) \cdot p_\epsilon(z) = p_X(1 - \alpha + \alpha z) \cdot p_\epsilon(z). \end{aligned} \quad (1)$$

Hence, the possible marginal distributions of a stationary CINAR(p) process are indeed those of the DSD family, including the negative binomial and (generalized) Poisson distribution, see Zhu & Joe (2003). Formula (1) furthermore implies that expectation and variance are given by $\mu_X = \mu_\epsilon / (1 - \alpha)$ and $\sigma_X^2 = (\alpha \mu_\epsilon + \sigma_\epsilon^2) / (1 - \alpha^2)$, like in the INAR(1) case. Properties concerning the serial dependence structure of stationary CINAR(p) processes are, however, more difficult to derive. Like in the case of the INAR(p) model, Definition 1 does not specify the CINAR(p) process completely. Therefore, we shall present in the sequel a new and quite general result on the autocorrelation structure first, which simplifies in the two special cases discussed afterwards. Zhu & Joe (2006) presented a similar result for one of these special cases only, and there only for $p = 2$.

Theorem 2 (Autocorrelation Structure of CINAR(p) Models) *Let $(X_t)_{\mathbb{Z}}$ be a stationary CINAR(p) process according to Definition 1. Let $\gamma(k) := \text{Cov}[X_t, X_{t-k}]$ denote the autocovariance function, define*

$$\mu(i, k) := E[(\alpha \circ_t X_{t-i}) \cdot X_{t-k}] - \alpha \cdot E[X_{t-i} \cdot X_{t-k}], \quad k \geq 1.$$

Then the autocovariances can be determined recursively from the equations

$$\gamma(k) = \alpha \cdot \sum_{i=1}^p \phi_i \cdot \gamma(|k - i|) + \sum_{i=k+1}^p \phi_i \cdot \mu(i, k),$$

where $\mu(i, k) = 0$ for $i \leq k$, and otherwise

$$\mu(i, k) = \phi_{i-k} \cdot (\text{Cov}[\alpha \circ_t X_{t-i}, \alpha \circ_{t-k} X_{t-i}] - \alpha^2 \sigma_X^2) + \alpha \cdot \sum_{r=k+1}^{i-1} \phi_{r-k} \cdot \mu(i, r).$$

$$\text{Especially, } \mu(i, i-1) = \phi_1 \cdot (\text{Cov}[\alpha \circ_t X_{t-i}, \alpha \circ_{t-i+1} X_{t-i}] - \alpha^2 \sigma_X^2).$$

The proof of Theorem 2 is provided by Appendix A. The result shows that the autocorrelation structure is determined only if the joint distribution of $(\alpha_1 \circ_{t+1} X_t, \dots, \alpha_p \circ_{t+p} X_t)$ has been specified. Two such possibilities are presented in the subsequent sections.

2.1 CINAR(p) – Identical Thinnings Model

Assume that all thinnings performed to X_t are identical, i. e.,

$$\alpha \circ_{t+1} X_t = \dots = \alpha \circ_{t+p} X_t = \alpha \circ X_t.$$

So the time index t below the thinning operation can be suppressed. Since

$$\text{Cov}[\alpha \circ_t X_{t-i}, \alpha \circ_{t-k} X_{t-i}] = V[\alpha \circ X_{t-i}] = \alpha^2 \sigma_X^2 + \alpha(1 - \alpha) \cdot \mu_X,$$

the recursion for $\mu(i, k)$, $i > k$, simplifies to

$$\mu(i, k) = \phi_{i-k} \cdot \alpha(1 - \alpha) \cdot \mu_X + \alpha \cdot \sum_{r=k+1}^{i-1} \phi_{r-k} \cdot \mu(i, r). \quad (2)$$

Hence, this type of CINAR(p) model has an autocorrelation structure similar to that of an ARMA($p, p-1$) model, i. e., it is closely related to the INAR(p) model of Alzaid & Al-Osh (1990).

Example 3 (CINAR(2) – Identical Thinnings Model) Consider the case $p = 2$. Then one obtains $\mu(2, 1) = \phi_1 \cdot \alpha(1 - \alpha) \cdot \mu_X$, and consequently

$$\begin{aligned} \gamma(1) &= \alpha \cdot \phi_1 \cdot (\sigma_X^2 + \phi_2(1 - \alpha)\mu_X) / (1 - \alpha\phi_2), \\ \gamma(k) &= \alpha \cdot (\phi_1 \cdot \gamma(k-1) + \phi_2 \cdot \gamma(k-2)), \quad k \geq 2. \end{aligned}$$

2.2 CINAR(p) – Independent Thinnings Model

Assume that conditioned on X_t , all thinnings $\alpha \circ_{t+1} X_t, \dots, \alpha \circ_{t+p} X_t$ are independent. So each time $t+j$, $j = 1, \dots, p$, X_t is newly involved in a thinning operation, disregarding the result of previous thinnings. For this model, simply $\text{Cov}[\alpha \circ_t X_{t-i}, \alpha \circ_{t-k} X_{t-i}] = \alpha^2 \sigma_X^2$. So $\mu(i, i-1) = 0$, and the recursion for $\mu(i, k)$, $i > k$, results in $\mu(i, k) = 0$ for all $i > k$. Hence, this type of

CINAR(p) model has an AR(p)-like autocorrelation structure, comparable to the INAR(p) model of Du & Li (1991):

$$\rho(k) = \alpha \cdot (\phi_1 \cdot \rho(|k-1|) + \dots + \phi_p \cdot \rho(|k-p|)). \quad (3)$$

Example 4 (CINAR(2) – Independent Thinnings) *If $p = 2$, then*

$$\rho(1) = \alpha \cdot \phi_1 / (1 - \alpha \phi_2), \quad \rho(k) = \alpha \cdot (\phi_1 \cdot \rho(k-1) + \phi_2 \cdot \rho(k-2)), \quad k \geq 2.$$

This result was also provided by Zhu & Joe (2006).

For both types of CINAR(p) models, model estimation can be done by solving the respective Yule-Walker equations, inserting the empirical instead of the theoretical autocorrelations. In case of the Independent Thinnings model, one can also compute conditional least squares and maximum likelihood estimates, since conditional expectation and distribution can be derived explicitly as

$$\begin{aligned} E[X_t | X_{t-1}, X_{t-2}, \dots] &= \mu_X \cdot (1 - \alpha) + \alpha \cdot \sum_{i=1}^p \phi_i \cdot X_{t-i}, \\ P(X_t = x | X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \dots) & \\ &= \sum_{y=0}^x P(\epsilon_t = y) \cdot \sum_{i=1}^p \phi_i \cdot \binom{x_{t-i}}{x-y} \cdot \alpha^{x-y} \cdot (1 - \alpha)^{x_{t-i}-x+y}. \end{aligned} \quad (4)$$

3 A Real-Data Example

The server of the Department of Statistics of the University of Würzburg collects log data concerning accesses to pages on the server. The data was arranged in such a way that the number of *different* IP addresses (\approx different users) registered within periods of one minute length can be read. We analyzed the data collected in November and December 2005. We restricted ourselves to accesses, which occurred between 10 o'clock in the morning and 6 o'clock in the evening, resulting in daily time series of length 481 each. As an illustrative example, we shall analyze in the following the time series collected on December 2nd, 2005.

Figure 1 (a) shows a run chart of the data. It gets clear that the data exhibits serial dependencies, but does not contradict a stationarity assumption. The histogram in Figure 1 (b) is plotted together with the Poisson distribution $Po(0.711)$, where 0.711 equals the arithmetic mean of the data. Empirical and theoretical distribution are very close to each other, indicating that a process model with Poisson marginals is reasonable. Estimated autocorrelations and partial autocorrelations are plotted in Figures 1 (c) and (d), respectively. The partial autocorrelation function abruptly decreases towards 0 after lag 5, making an autoregressive model of order $p \leq 5$ a reasonable choice. Such

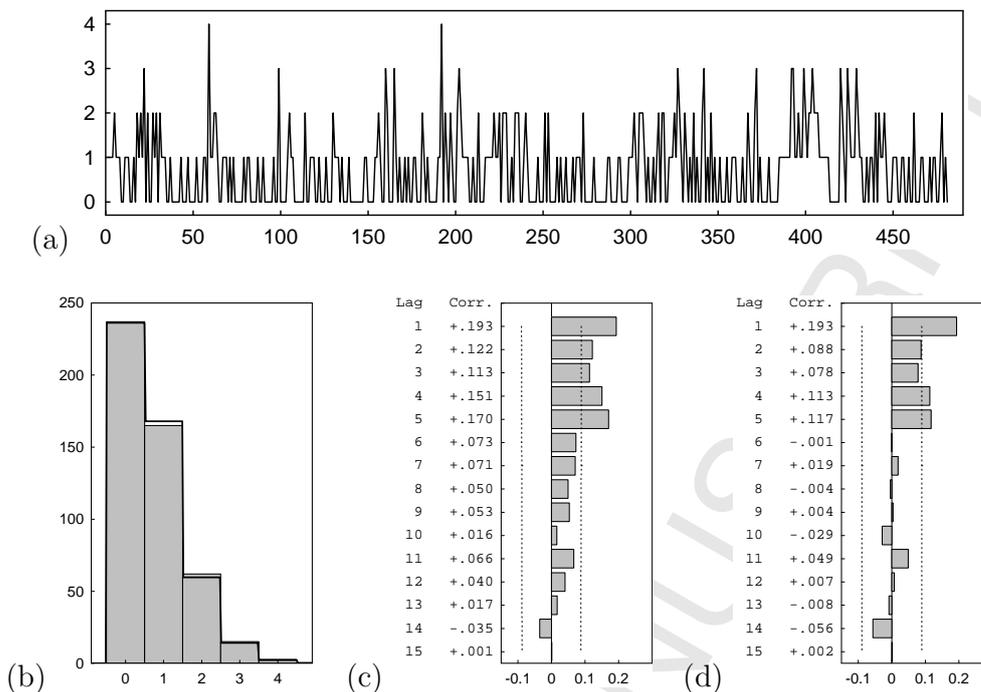


Fig. 1. Run chart, histogram, empirical autocorrelation and partial autocorrelation function of the IP data of Section 3.

a long-term dependence is plausible for the IP data, since users often click through a homepage for more than a minute.

Taking together these observations, we decide to model the data by a *Poisson CINAR(p) – Independent Thinnings* model, $p \leq 5$, which has an autoregressive dependence structure, see formula (3). Because of the Poisson assumption, the model is determined by the $p + 1$ parameters μ_ϵ , α , $\phi_1, \dots, \phi_{p-1}$ ($\phi_p = 1 - \phi_1 - \dots - \phi_{p-1}$) and has Poisson marginals with $\mu_X = \sigma_X^2 = \frac{\mu_\epsilon}{1-\alpha}$. We consider all models with $0 \leq p \leq 5$, i. e., including an i.i.d. model. Since $\hat{\rho}_p(2)$ and $\hat{\rho}_p(3)$ do not deviate significantly from 0, we also consider a reduced CINAR(5) model with $\phi_2 = \phi_3 = 0$.

All CINAR(p) candidate models have been fitted to the data in two steps: Initial estimates for the parameters are obtained from the Yule-Walker equations (3), the final estimates are obtained by numerically maximizing the conditioned likelihood function (with the help of Mathematica 5), which can be computed easily from formula (4). Remember that $\hat{\phi}_p = 1 - \hat{\phi}_1 - \dots - \hat{\phi}_{p-1}$. The results are presented in Table 1, together with the respective values of the information criteria AIC and BIC. Obviously, the i.i.d. model performs worst, so serial dependence has to be considered. Among the models with serial dependence, the reduced fifth order model is the best choice.

Table 1
Maximum-likelihood estimation of candidate models.

Model	$\hat{\mu}_\epsilon$	$\hat{\alpha}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	AIC	BIC
$p = 0$	0.706						1059	1063
$p = 1$	0.566	0.197					1044	1052
$p = 2$	0.524	0.255	0.696				1042	1055
$p = 3$	0.484	0.312	0.554	0.180			1041	1058
$p = 4$	0.442	0.373	0.452	0.120	0.167		1038	1059
$p = 5$	0.412	0.411	0.377	0.075	0.130	0.185	1036	1061
$p = 5, \phi_2, \phi_3 = 0$	0.443	0.369	0.452			0.252	1034	1050

4 Conclusion

In this article, we investigated an autoregressive model for time series of counts based on binomial thinning. In contrast to the standard INAR(p) models, possible marginal distributions of this new model class are easily obtained since they coincide with those of the INAR(1) model: The DSD family, including negative binomial and generalized Poisson distribution. We derived a set of Yule-Walker equations to describe the autocorrelation structure of the whole model family, and showed that these equations simplify in two special cases. We also briefly discussed aspects of model estimation and illustrated such estimation procedures with a real-data example. This example also demonstrated the practical relevance of the new model family.

References

- Alzaid, A.A., Al-Osh, M.A., 1990. An integer-valued p^{th} -order autoregressive structure (INAR(p)) process. *Jour. Appl. Prob.* 27, 314-324.
- Du, J.-G., Li, Y., 1991. The integer-valued autoregressive (INAR(p)) model. *Jour. Time Series Analysis* 12(2), 129-142.
- Lawrance, A.J., Lewis, P.A.W., 1980. The exponential autoregressive-moving average EARMA(p, q) process. *Jour. Royal Stat. Soc. B* 42(2), 150-161.
- McKenzie, E., 1985. Some simple models for discrete variate time series. *Water Resources Bulletin* 21(4), 645-650.
- Steutel, F.W., van Harn, K., 1979. Discrete analogues of self-decomposability and stability. *Ann. Prob.* 7(5), 893-899.
- Zhu, R., Joe, H., 2003. A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models* 19(2), 235-254.
- Zhu, R., Joe, H., 2006. Modelling count data time series with Markov processes based on binomial thinning. *Jour. Time Series Analysis* 27(5), 725-738.

A Proof of Theorem 2

The given Yule-Walker type equations for $\gamma(k)$ follow immediately from

$$\begin{aligned} E[X_t \cdot X_{t-k}] &= \sum_{i=1}^p \phi_i \cdot E[(\alpha \circ_t X_{t-i} + \epsilon_t) \cdot X_{t-k}] \\ k \geq 1 &= \mu_\epsilon \cdot \mu_X + \sum_{i=1}^p \phi_i \cdot E[(\alpha \circ_t X_{t-i}) \cdot X_{t-k}] \\ &= (1 - \alpha) \cdot \mu_X^2 + \sum_{i=1}^p \alpha \phi_i \cdot E[X_{t-i} \cdot X_{t-k}] + \sum_{i=1}^p \phi_i \cdot \mu(i, k) \end{aligned}$$

and $\mu(i, k) = 0$ for $i \leq k$. The latter results from

$$\begin{aligned} E[(\alpha \circ_t X_{t-i}) \cdot X_{t-k}] &= E[E[(\alpha \circ_t X_{t-i}) \cdot X_{t-k} \mid X_{t-i}]] \\ i \leq k &= E[E[\alpha \circ_t X_{t-i} \mid X_{t-i}] \cdot E[X_{t-k} \mid X_{t-i}]] = \alpha \cdot E[X_{t-i} \cdot X_{t-k}]. \end{aligned}$$

So it remains to prove the expression for $\mu(i, k)$ for $i > k$. Then

$$\begin{aligned} E[(\alpha \circ_t X_{t-i}) \cdot X_{t-k}] &= \sum_{j=1}^p \phi_j \cdot E[(\alpha \circ_t X_{t-i}) \cdot (\alpha \circ_{t-k} X_{t-k-j} + \epsilon_{t-k})] \\ k < i &= \alpha(1 - \alpha) \cdot \mu_X^2 + \sum_{j=1}^p \phi_j \cdot E[(\alpha \circ_t X_{t-i}) \cdot (\alpha \circ_{t-k} X_{t-k-j})] \\ &= \alpha(1 - \alpha) \cdot \mu_X^2 + \phi_{i-k} \cdot E[(\alpha \circ_t X_{t-i}) \cdot (\alpha \circ_{t-k} X_{t-i})] \\ &\quad + \sum_{j=1}^{i-k-1} \phi_j \cdot E[E[(\alpha \circ_t X_{t-i}) \cdot (\alpha \circ_{t-k} X_{t-k-j}) \mid X_{t-k-j}]] \\ &\quad + \sum_{j=i-k+1}^p \phi_j \cdot E[E[(\alpha \circ_t X_{t-i}) \cdot (\alpha \circ_{t-k} X_{t-k-j}) \mid X_{t-i}]] \\ &= \alpha(1 - \alpha) \cdot \mu_X^2 + \alpha \cdot \sum_{j=1}^{i-k-1} \phi_j \cdot \mu(i, k + j) \\ &\quad + \phi_{i-k} \cdot (E[(\alpha \circ_t X_{t-i})(\alpha \circ_{t-k} X_{t-i})] - \alpha^2 E[X_{t-i}^2]) \\ &\quad + \alpha \cdot \sum_{j=1}^{i-k} \phi_j \cdot (\alpha \cdot E[X_{t-k-j} \cdot X_{t-i}]) \\ &\quad + \alpha \cdot \sum_{j=i-k+1}^p \phi_j \cdot E[(\alpha \circ_{t-k} X_{t-k-j}) \cdot X_{t-i}]. \end{aligned}$$

Inserting

$$\text{Cov}[\alpha \circ_t X_{t-i}, \alpha \circ_{t-k} X_{t-i}] - \alpha^2 \sigma_X^2 = E[(\alpha \circ_t X_{t-i})(\alpha \circ_{t-k} X_{t-i})] - \alpha^2 E[X_{t-i}^2],$$

$$\alpha \cdot \sum_{j=1}^p \phi_j \cdot E[\epsilon_{t-k} \cdot X_{t-i}] = \alpha(1 - \alpha) \cdot \mu_X^2, \quad \text{and}$$

$$E[(\alpha \circ_{t-k} X_{t-k-j}) \cdot X_{t-i}] = \alpha \cdot E[X_{t-k-j} \cdot X_{t-i}] \quad \text{for } j \leq i - k$$

into this equation, we obtain

$$\begin{aligned}
E[(\alpha \circ_t X_{t-i}) \cdot X_{t-k}] &= \\
&= \alpha \cdot \sum_{j=1}^{i-k-1} \phi_j \cdot \mu(i, k+j) + \phi_{i-k} \cdot (Cov[\alpha \circ_t X_{t-i}, \alpha \circ_{t-k} X_{t-i}] - \alpha^2 \sigma_X^2) \\
&\quad + \alpha \cdot \sum_{j=1}^p \phi_j \cdot E[\epsilon_{t-k} \cdot X_{t-i}] + \alpha \cdot \sum_{j=1}^p \phi_j \cdot E[(\alpha \circ_{t-k} X_{t-k-j}) \cdot X_{t-i}] \\
&= \alpha \cdot \sum_{j=1}^{i-k-1} \phi_j \cdot \mu(i, k+j) + \phi_{i-k} \cdot (Cov[\alpha \circ_t X_{t-i}, \alpha \circ_{t-k} X_{t-i}] - \alpha^2 \sigma_X^2) \\
&\quad + \alpha \cdot E[X_{t-k} \cdot X_{t-i}].
\end{aligned}$$

This completes the proof.