Forecast error is not only caused by the randomness of the data-generating process but also by the uncertainty due to estimated model parameters. We investigate these different sources of forecast error for a popular type of count process, the Poisson first-order integer-valued autoregressive (INAR(1)) process. However, many of our analytical derivations also hold for the more general family of conditional linear AR(1) (CLAR(1)) processes. In addition, results from a simulation study are presented, to verify and complement our asymptotic approximations.

**KEYWORDS**
count time series, estimation uncertainty, point forecasting, Poisson INAR(1) model

1 | **INTRODUCTION**

If analyzing stationary real-world processes \((X_t)_{t \in \mathbb{Z}} = \ldots, -1, 0, 1, \ldots\)
, one is often concerned with an (approximately) exponentially decaying autocorrelation function (ACF), that is, an ACF of the form \(\rho(k) = \text{Corr}[X_t, X_{t-k}] = \alpha^k\ (k \in \mathbb{N} = \{1, 2, \ldots\})\) for some \(|\alpha| < 1\). Such an ACF is observed for the well-known (Gaussian) first-order autoregressive (AR(1)) model, but also for many other types of time series model. In particular, several models with such an AR(1)-like ACF have been proposed for count time series, that is, where the range consists of nonnegative integer values from \(\mathbb{N}_0 = \{0, 1, \ldots\}\) (Weiβ, 2018). Besides the famous integer-valued AR(1) (INAR(1)) model by McKenzie (1985), these include, among others, the binomial AR(1) model by McKenzie (1985), the INARCH(1) model (with conditional heteroscedasticity) studied in Weiβ (2010), or the new geometric INAR(1) (NGINAR(1)) model by Ristić, Bakouch, and Nastić (2009). Because
of this variety of different AR(1)-like models, Grunwald, Hyndman, Tedesco, and Tweedie (2000) defined and discussed a general class of time series models, which are characterized by such an autocorrelation structure, namely, the conditional linear AR(1) (CLAR(1)) models.

**Definition 1.** A stationary real-valued (either discrete- or continuous-valued) process \((X_t)\) is said to be a CLAR(1) process if its conditional mean is given by

\[
E[X_t | X_{t-1}, ...] = \alpha \cdot X_{t-1} + \beta \quad \text{with} \quad \alpha \in (-1; 1) \quad \text{and} \quad \beta \in \mathbb{R},
\]

and if the variance exists, \(\sigma^2 = \text{Var}[X_t] < \infty\).

A CLAR(1) process according to Definition 1 satisfies \(\mu = E[X_t] = \beta / (1 - \alpha) \) and \(\rho(k) = \alpha^k \) (Grunwald et al., 2000). If being concerned with a count process, then we additionally have to require that \(\beta > 0 \) and \(\alpha \in [0; 1)\).

**Example 1.** The INAR(1) model by McKenzie (1985) is defined by the model recursion \(X_t = \alpha \cdot X_{t-1} + \epsilon_t\), where “\(\alpha \cdot \)” denotes the binomial thinning operation (as an integer-valued substitute of the ordinary multiplication in the AR(1)’s model recursion), and where \((\epsilon_t)\) constitutes an independent and identically distributed (i.i.d.) count process of innovations. Here, binomial thinning is defined by requiring that \(\alpha \cdot X_t \sim \text{Bin}(X, \alpha)\), and it is assumed that these thinnings are executed independently of other thinnings and innovations. A comprehensive survey of properties and methods concerning the INAR(1) model is presented in chapter 2 of Weiß (2018). Among others, it holds that Poisson-distributed innovations lead to Poisson-distributed observations, namely, \(\epsilon_t \sim \text{Poi}(\beta)\) and \(X_t \sim \text{Poi}(\mu)\); we refer to this specific instance as the Poisson INAR(1) model.

For a CLAR(1) process with its conditional-mean-based definition, it appears obvious to do point forecasting based on the conditional mean. So given the parameter values of \(\alpha, \beta\), the \(h\)-step-ahead point forecast is defined by

\[
\hat{X}_{T+h} = E[X_{T+h} | X_T, ...] = a^h \cdot X_T + \frac{1 - a^h}{1 - a} \beta.
\]

In case of a discrete-valued data-generating process (DGP), this does not lead to a coherent point forecast in the sense of Freeland and McCabe (2004) and Jung and Tremayne (2006), so an additional discretization (e.g., rounding) would be required in practice here (Homburg, Weiß, Alwan, Frahm, & Göb, 2019). However, for the sake of a unique treatment, let us concentrate on the basic mean-based forecasting approach. In addition, the mean-based forecasting of count processes has been considered by many other authors as well, for example, by Sutradhar (2008), Mohammadipour and Boylan (2012), and Maiti, Biswas, and Das (2016).

In practice, \(\alpha, \beta\) are not known but have to be estimated from the available time series data \(X_1, ..., X_S\) with \(S \leq T\); denote the corresponding estimators by \(\hat{\alpha}, \hat{\beta}\). A possible choice for \(\hat{\alpha}, \hat{\beta}\), which is unique across all CLAR(1) processes, are the moment estimators \(\hat{\rho}(1) \) and \(\hat{\mu}(1) \), respectively (whereas maximum likelihood (ML) estimators would be model-specific). Forecasting with estimated parameters is then done according to

\[
\hat{X}_{T+h} = \hat{\alpha}^h \cdot X_T + \frac{1 - \hat{\alpha}^h}{1 - \hat{\alpha}} \hat{\beta}.
\]

In particular, the 1-step-ahead forecast \((h = 1)\) is computed as \(\hat{X}_{T+1} = \hat{\alpha} \cdot X_T + \hat{\beta}\), where \(\hat{\alpha}, \hat{\beta}\) are computed from \(X_1, ..., X_S\) with \(S \leq T\).
Inspired by the works of Phillips (1979) and Fuller and Hasza (1981), let us now look at stochastic properties of the forecast error

\[ \hat{X}_{T+1} - X_{T+1} = (\alpha X_T + \beta - X_{T+1}) + (\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta) . \]  

(1)

The situation \( h = 1 \) in (1) is not only most important for applications in practice, it is also the only situation where an asymptotic analytic treatment turned out to be feasible (because the order of the involved moments increases with \( h \)). Nevertheless, the case \( h > 1 \) is later investigated via simulations.

Phillips (1979) considered the special case of a zero-mean Gaussian AR(1) process (so only \( \alpha \) has to be estimated). There, it was possible to derive an approximate distribution for this forecast error. This, however, will not be possible for more general CLAR(1) processes. In particular, it will not be feasible if considering CLAR(1) count processes like the Poisson INAR(1) model (Weiß, 2018). Therefore, in the present article, we concentrate on the probably most important measure of forecast uncertainty, the variance of \( \hat{X}_{T+1} - X_{T+1} \), like it was also done by Fuller and Hasza (1981) for Gaussian AR processes. Knowledge of the variance will also allow to construct simple \( k\sigma \) prediction intervals. Here, \( \mathbb{V}[\hat{X}_{T+1} - X_{T+1}] \) is influenced by

- the estimation uncertainty of \( \hat{\alpha} - \alpha \) and \( \hat{\beta} - \beta \);
- the cross-dependence of \( X_T \) and \( \hat{\alpha} - \alpha, \hat{\beta} - \beta \);
- the randomness of the DGP as expressed by \( X_{T+1} - \alpha X_T - \beta \).

For the ordinary AR(1) process \( X_t = \alpha \cdot X_{t-1} + \epsilon_t \), the latter difference reduces to \( \epsilon_{T+1} - \beta \), so uncertainty is only caused by the error term. For an INAR(1) process, in contrast, we have the binomial thinning operation as an additional source of uncertainty, \( X_{T+1} - \alpha X_T - \beta = (\alpha \circ X_T - \alpha X_T) + (\epsilon_{T+1} - \beta) \). Note that explicit asymptotic formulas for the case of moment estimators \( \hat{\alpha}, \hat{\beta} \) have been studied by Weiß and Schweer (2016) for the Poisson INAR(1) model (and also the INARCH(1) one). It should also be emphasized that the dependence between \( X_T \) and \( \hat{\alpha}, \hat{\beta} \) depends on the actual value of the difference \( T - S \).

To derive an analytic approximation for the variance of the forecast error, \( \mathbb{V}[\hat{X}_{T+1} - X_{T+1}] \), we start with a linear approximation of the involved estimators (under a general CLAR(1) assumption) in Section 2. Then, again under CLAR(1) assumption, we derive a general expression for the forecast-error variance in Section 3. This expression depends on higher order mixed moments, which can only be computed under full model specification. This is illustrated in Section 4, where we derive a simple closed-form approximate formula for the case of a Poisson INAR(1) DGP. Also, approximations for large \( T - S \) or \( h \) are discussed. The approximations are evaluated in the simulation study presented in Section 5. Finally, we conclude in Section 6.

## 2  \ |  LINEAR APPROXIMATION OF MOMENT ESTIMATORS

Let us denote the \( r \)th-order mixed moments

\[ \mu_{(s_1, \ldots, s_{r-1})} := \mathbb{E}[X_t \cdot X_{t+s_1} \cdot \ldots \cdot X_{t+s_{r-1}}], \]

(2)

with \( 0 \leq s_1 \leq \cdots \leq s_{r-1} \) and \( r \in \mathbb{N} \). Such moments up to order 4 are readily provided for the Poisson INAR(1) model introduced in example 1 by Weiß (2012), but also for the Poisson
INARCH(1) model by Weiß (2010), or for the binomial AR(1) model by Weiß and Kim (2013). Hence, although we concentrate on a Poisson INAR(1) DGP later in Section 4, analogous derivations would be possible for these alternative types of count CLAR(1) process, too.

Defining the functions

\[
g_a(y_1, y_2, y_3) := \frac{y_3 - y_1^2}{y_2 - y_1^2}, \quad g_\beta(y_1, y_2, y_3) := -\frac{y_1 y_2 - y_3}{y_2 - y_1^2},
\]

and recalling that \(\mu(k) = \mathbb{E}[X_t | X_{t-k}]\), we have \(g_a(\mu, \mu(0), \mu(1)) = \alpha\) and \(g_\beta(\mu, \mu(0), \mu(1)) = \beta\), also see Weiß and Schweer (2016). Furthermore, with the sample counterparts \(\hat{\mu} = \bar{X} = \frac{1}{S}\sum_{t=1}^{S} X_t\), \(\hat{\mu}(0) = \frac{1}{S}\sum_{t=1}^{S} X_t^2\), and \(\hat{\mu}(1) = \frac{1}{S}\sum_{t=1}^{S} X_t X_{t-1}\) of these raw moments, we obtain the moment estimators \(\hat{\alpha} = g_a(\hat{\mu}, \hat{\mu}(0), \hat{\mu}(1))\) and \(\hat{\beta} = g_\beta(\hat{\mu}, \hat{\mu}(0), \hat{\mu}(1))\). These agree, except for asymptotically negligible terms, with the ordinary moment estimators \(\hat{\rho}(1)\) and \(\overline{X} (1 - \hat{\rho}(1))\), respectively.

To be able to approximate the variance of the forecast error (1), we approximate the deviations \(\hat{\alpha} - \alpha, \hat{\beta} - \beta\) by their first-order Taylor approximations. Then we compute the asymptotic variance of the linearized version of \((\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta)\) as an approximation of the corresponding true variance. For this purpose, we require the higher order joint moments of the form (2). Generally, these depend on the specific model assumption for the DGP. However, a few relations hold for CLAR(1) processes in general.

**Proposition 1.** Let \((X_t)_{t \in \mathbb{Z}}\) be a stationary CLAR(1) process according to Definition 1.

(i) For any \(k \geq 0\), we have \(\mu(k) = \alpha^k \sigma^2 + \mu^2\).

(ii) For any \(l \geq k \geq 0\), we have \(\mu(k,l) = \alpha^{l-k} (\mu(k,k) - \mu \mu(k)) + \mu \mu(k)\).

**Proof.** An explicit expression for the case \(\mu(k)\) is easily derived by using that the ACF equals \(\rho(k) = \alpha^k\):

\[\mu(k) = \text{Cov}[X_t, X_{t+k}] + \mu^2 = \sigma^2 \alpha^k + \mu^2.\]

For \(\mu(k,l)\) with \(l > k\), the CLAR(1) assumption implies that

\[\mu(k,l) = \mathbb{E}[X_t \cdot X_{t+k} \cdot \mathbb{E}[X_{t+l} | X_{t+l-1}, \ldots]]\]

\[= \alpha \cdot \mu(k,l-1) + \mu(1 - \alpha) \cdot \mu(k)\]

\[= \cdots = \alpha^{l-k} \cdot \mu(k,k) + \mu(1 - \alpha) \cdot \mu(k) \sum_{j=0}^{l-k-1} \alpha^j\]

\[= \alpha^{l-k} (\mu(k,k) - \mu \cdot \mu(k)) + \mu \cdot \mu(k).\]

This completes the proof of Proposition 1.

The first-order Taylor approximation of \(\hat{\alpha} - \alpha\) or \(\hat{\beta} - \beta\), respectively, takes the form

\[
\frac{\partial}{\partial y_1} g_a(\mu, \mu(0), \mu(1)) (\hat{\mu} - \mu) + \frac{\partial}{\partial y_2} g_a(\mu, \mu(0), \mu(1)) (\hat{\mu}(0) - \mu(0))
\]

\[+ \frac{\partial}{\partial y_3} g_a(\mu, \mu(0), \mu(1)) (\hat{\mu}(1) - \mu(1)),\]

\[4\]
where the functions $g_\alpha, g_\beta$ are defined by (3). These linear approximations are computed as follows.

**Proposition 2.** Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary CLAR(1) process according to Definition 1. Recall the moment estimators $\hat{\alpha} = g_\alpha(\hat{\mu}, \hat{\mu}_0, \hat{\mu}_1)$ and $\hat{\beta} = g_\beta(\hat{\mu}, \hat{\mu}_0, \hat{\mu}_1)$ defined by (3).

Then, the linear approximation of $\hat{\alpha} - \alpha$ is

$$\hat{\alpha} - \alpha \approx \frac{1}{\sigma^2} \frac{1}{S} \sum_{t=1}^{S} (a \cdot X_t - \alpha \cdot X_t^2 + X_tX_{t-1} + b),$$

and the linear approximation of $\hat{\beta} - \beta$ is

$$\hat{\beta} - \beta \approx \frac{1}{\sigma^2} \frac{1}{S} \sum_{t=1}^{S} (A \cdot X_t + \alpha \mu \cdot X_t^2 - \mu \cdot X_tX_{t-1} + B),$$

where

$$a = -2 \mu (1 - \alpha), \quad b = \mu^2 (1 - \alpha),$$
$$A = (1 - \alpha) (\sigma^2 + 2 \mu^2), \quad B = -\mu (1 - \alpha) (\sigma^2 + \mu^2).$$

The proof of Proposition 2 is provided by Appendix A.1.

Before considering the variance of the forecast error $\hat{X}_{T+1} - X_{T+1}$, let us first have a brief look at the mean forecast error $\mathbb{E} [\hat{X}_{T+1} - X_{T+1}]$ under linear approximation. According to (1), the mean forecast error $\mathbb{E} [\hat{X}_{T+1} - X_{T+1}]$ decomposes into three parts. The mean $\mathbb{E} [X_{T+1} - \alpha X_T - \beta] = 0$ by the definition of a CLAR(1) process. Recalling that $\hat{\alpha}$ is computed from $X_1, \ldots, X_S$ with $S \leq T$, and by applying the law of total expectation, we further conclude that

$$\mathbb{E} [(\hat{\alpha} - \alpha) X_T] = \mathbb{E} [(\hat{\alpha} - \alpha) \mathbb{E}[X_T | X_S, \ldots]]$$
$$= \mathbb{E} [(\hat{\alpha} - \alpha) \left( \alpha^{T-S} X_S + \frac{1-\alpha^{T-S}}{1-\alpha} \beta \right)].$$

So,

$$\mathbb{E} [\hat{X}_{T+1} - X_{T+1}] = \alpha^{T-S} \cdot \mathbb{E} [(\hat{\alpha} - \alpha) X_S] + (1 - \alpha^{T-S}) \mu \cdot \mathbb{E} [\hat{\alpha} - \alpha] + \mathbb{E} [\hat{\beta} - \beta]. \quad (5)$$

Under linear approximation, both means $\mathbb{E} [\hat{\alpha} - \alpha]$ and $\mathbb{E} [\hat{\beta} - \beta]$ become zero. Therefore, we approximate the mean of $\hat{X}_{T+1} - X_{T+1}$ by

$$\mathbb{E} [\hat{X}_{T+1} - X_{T+1}] \approx \alpha^{T-S} \cdot \mathbb{E} [(\hat{\alpha} - \alpha) X_S],$$

that is, by a term expressing the crossdependence of $X_S$ and $\hat{\alpha} - \alpha$. This mean can be approximated for a general CLAR(1) process as follows.

**Proposition 3.** Let $(X_t)_{t\in\mathbb{Z}}$ be a CLAR(1) process according to Definition 1. Then, under linear approximation,

$$\mathbb{E} [\hat{X}_{T+1} - X_{T+1}] \approx \alpha^{T-S} \cdot \mathbb{E} [(\hat{\alpha} - \alpha) X_S].$$
where
\[\mathbb{E}[(\hat{\alpha} - \alpha) X_S] \approx -\frac{1 - \alpha^S}{S} \left( \mu + \frac{1}{\sigma^2 (1 - \alpha)} \left( \mu_{(0,1)} - \mu_{(1,1)} \right) \right).\]

The proof of Proposition 3 is provided by Appendix A.2. Note that \(\mu_{(0,1)} - \mu_{(1,1)} = \mathbb{E}[X_0^2 X_1] - \mathbb{E}[X_0 X_1^2]\), that is, this difference becomes zero if the considered CLAR(1) process is time-reversible. Thus, for a time-reversible CLAR(1) process, Proposition 3 simplifies to
\[\mathbb{E}\left[\hat{X}_{T+1} - X_{T+1}\right] \approx \alpha^{T-S} \cdot \mathbb{E}[(\hat{\alpha} - \alpha) X_S] \approx -\frac{1}{S} \mu \alpha^{T-S}(1 - \alpha^S).\] (6)

Examples of such time-reversible CLAR(1) processes are the continuous-valued Gaussian AR(1) process, and the discrete-valued Poisson INAR(1) process (unbounded counts) as well as binomial AR(1) process (bounded counts), see McKenzie (1985).

The formulas derived in Proposition 3 and Equation (6) are not expected to serve as a reasonable approximation of the actual mean forecast error because the biases of \(\hat{\alpha}, \hat{\beta}\) in (5) are ignored due to the linear approximation. However, Proposition 3 and Equation (6) are relevant to us for deriving the variance of \(\hat{X}_{T+1} - X_{T+1}\). As we have seen, \(\mathbb{E}[(\hat{\alpha} - \alpha) X_S] \approx 0 + O(S^{-1})\) under linear approximation. This result is used in the subsequent section to approximate the forecast-error variance \(\mathbb{V} [\hat{X}_{T+1} - X_{T+1}]\).

### 3 VARIANCE OF CLAR(1) FORECAST ERROR

For the variance of \(\hat{X}_{T+1} - X_{T+1}\), formula (1) first implies the decomposition
\[\mathbb{V} [\hat{X}_{T+1} - X_{T+1}] = \mathbb{V} [X_{T+1} - \alpha X_T - \beta] + \mathbb{V} [(\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta)] - 2 \text{Cov} [(\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta), X_{T+1} - \alpha X_T - \beta].\]

However, it should be noted that for any expression \(f(X_1, \ldots, X_T)\), the law of total expectation and the definition of a CLAR(1) process imply that
\[\mathbb{E} [f(X_1, \ldots, X_T) (X_{T+1} - \alpha X_T - \beta)] = \mathbb{E} [f(X_1, \ldots, X_T) \mathbb{E}[X_{T+1} - \alpha X_T - \beta \mid X_T, \ldots, X_1]] = 0.\]

Thus,
\[\mathbb{V} [\hat{X}_{T+1} - X_{T+1}] = \mathbb{V} [X_{T+1} - \alpha X_T - \beta] + \mathbb{V} [(\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta)].\]

This can be further decomposed into
\[\mathbb{V} [\hat{X}_{T+1} - X_{T+1}] = \mathbb{V} [X_{T+1} - \alpha X_T - \beta] + \mathbb{V} [(\hat{\alpha} - \alpha) X_T] + \mathbb{V} [\hat{\beta} - \beta] + 2 \text{Cov} [(\hat{\alpha} - \alpha) X_T, \hat{\beta} - \beta],\] (7)

where the “known-parameter variance” (unavoidable variance if doing forecasting) computes as
\[ \forall \left[ X_{T+1} - \alpha \, X_T - \beta \right] = \forall \left[ X_{T+1} \right] + \alpha^2 \, \forall \left[ X_T \right] - 2 \, \alpha \, \text{Cov} \left[ X_{T+1}, X_T \right] = \sigma^2 + \alpha^2 \, \sigma^2 - 2 \, \alpha \, \sigma^2 = (1 - \alpha^2) \, \sigma^2. \] (8)

**Remark 1.** In (7), it is again possible to account for the possible discrepancy between \( S \) and \( T \). The law of total covariance implies that

\[ \text{Cov} \left[ \left( \hat{\alpha} - \alpha \right) \, X_T, \hat{\beta} - \beta \right] = \text{Cov} \left[ \left( \hat{\alpha} - \alpha \right) \, \mathbb{E} \left[ X_T \mid X_S, \ldots \right], \hat{\beta} - \beta \right] = \alpha^{T-S} \, \text{Cov} \left[ \left( \hat{\alpha} - \alpha \right) \, X_S, \hat{\beta} - \beta \right] + (1 - \alpha^{T-S}) \, \mu \, \text{Cov} \left[ \hat{\alpha} - \alpha, \hat{\beta} - \beta \right]. \]

Analogously, the law of total variance yields

\[ \forall \left[ \left( \hat{\alpha} - \alpha \right) \, X_T \right] = \forall \left[ \left( \hat{\alpha} - \alpha \right) \, \left( \alpha^{T-S} \, X_S + (1 - \alpha^{T-S}) \, \mu \right) \right] + \mathbb{E} \left[ \left( \hat{\alpha} - \alpha \right)^2 \, \forall \left[ X_T \mid X_S, \ldots \right] \right]. \]

However, for general CLAR(1) processes, the conditional variance \( \forall \left[ X_T \mid X_S, \ldots \right] \) is not specified. Consequently, we cannot derive a general expression for \( S \leq T \) this time. Therefore, in the sequel, we concentrate on the case \( S = T \). However, later in Section 4, when considering the special case of a Poisson INAR(1) process, we explain how the results could be extended to the case \( S < T \).

To obtain an approximation for the “estimated-parameter variance add-on” \( \forall \left[ \left( \hat{\alpha} - \alpha \right) \, X_T + (\hat{\beta} - \beta) \right] \) of order \( T^{-1} \), we need \( T^{-1} \)-approximations for \( \forall \left[ \left( \hat{\alpha} - \alpha \right) \, X_T \right] \), \( \forall [\hat{\beta} - \beta] \), and \( \text{Cov} \left[ \left( \hat{\alpha} - \alpha \right) \, X_T, \hat{\beta} - \beta \right] \), respectively. Since the considered linear approximations of \( \hat{\alpha} - \alpha, \hat{\beta} - \beta \) imply that the means of \( (\hat{\alpha} - \alpha) \, X_T \) and \( \hat{\beta} - \beta \) are 0 + \( O(T^{-1}) \), it actually suffices to derive \( T^{-1} \)-approximations for \( \mathbb{E} \left[ \left( \hat{\alpha} - \alpha \right)^2 \, X_T^2 \right], \mathbb{E} \left[ (\hat{\beta} - \beta)^2 \right], \) and \( \mathbb{E} \left[ (\hat{\alpha} - \alpha) \, (\hat{\beta} - \beta) \, X_T \right] \), respectively. In a first step, this is done for general CLAR(1) processes. Later in Section 4, the obtained formulas are used to derive explicit closed-form expressions for the illustrative case of a Poisson INAR(1) process.

**Proposition 4.** Let \( (X_t) \) be a CLAR(1) process according to Definition 1, let \( S = T \). Then,

\[ \forall \left[ \left( \hat{\alpha} - \alpha \right) \, X_T + (\hat{\beta} - \beta) \right] \approx \frac{1}{T} \left[ \frac{A^2}{\sigma^2} + \frac{1 + \alpha}{1 - \alpha} \, \frac{2}{\sigma^4} \left( \mu A \left( \mu_{(0,1)} - \mu_{(1,1)} \right) + \sigma^2 \, a \, B \right) + O(T^{-2}) \right. \]
\[ + \frac{\mu \, b}{\sigma^4} \left( A \, \mu - 4 \, B \right) + \frac{2 \, b}{\sigma^4} \left( \frac{1}{T} \sum_{k=0}^{T-1} \left( \mu_{(k,k)} - \alpha \, \mu_{(0,k,k)} + \mu_{(1,k+1,k+1)} \right) \right) \]
\[ + \frac{1}{T^2} \sum_{s,t=1}^{T} \left( 2 \, a \, A \, \mathbb{E}[X_t \mathbf{X}_s] + a^2 \, \mathbb{E}[X_t X_s X_t^2] \right) \]
\[ + \alpha^2 \, \mu^2 \, \mathbb{E}[X_t^2 X_s^2] - 2 \, \alpha^2 \, \mu \, \mathbb{E}[X_t^2 X_s^2 X_t^2] + a^2 \, \mathbb{E}[X_t^2 X_s X_t^2] \]
\[ + \mu^2 \, \mathbb{E}[X_t X_{t-1} X_s X_{s-1}] - 2 \mu \, \mathbb{E}[X_t X_{t-1} X_s X_{s-1} X_t] + \mathbb{E}[X_t X_{t-1} X_s X_{s-1} X_t^2] \]
\[ + 2 \, \alpha \, A \, \mu \, \mathbb{E}[X_t^2 X_s] - 2 \, (A + 2 \, b) \, \alpha \, \mathbb{E}[X_t^2 X_s X_t] - 2 \, a \, a \, \mathbb{E}[X_t^2 X_s^2] \]
\[ - 2 \, \mu \, A \, \mathbb{E}[X_t X_s X_{s-1}] + (2 \, A + 2 \, b) \, \mathbb{E}[X_t X_s X_{s-1} X_t] + 2 \, a \, \mathbb{E}[X_t X_s X_{s-1} X_t^2] \]
\[ - 2 \, a \, a \, \mu \, \mathbb{E}[X_t^2 X_s X_{s-1}] - 4 \, a \, \mu \, \mathbb{E}[X_t^2 X_s X_{s-1} X_t] - 2 \, \alpha \, \mathbb{E}[X_t^2 X_s X_{s-1} X_t^2] \right]. \]

where \( a, b, A, B \) are defined in Proposition 2.

The proof of Proposition 4 is provided by Appendix A.3. Since general closed-form expressions for the higher order moments of a CLAR(1) process are not possible, the approximate variance as
In this section, we consider a special type of CLAR(1) process, the Poisson INAR(1) process. Up to order 6 are required for this purpose. We shall exemplify the required computations for the special case of a Poisson INAR(1) process.

4 \ VARIANCE OF POISSON INAR(1) FORECAST ERROR

In this section, we consider a special type of CLAR(1) process, the Poisson INAR(1) process introduced in Example 1. To be able to compute the approximate variance according to Proposition 4, we need expressions for the higher order joint moments.

**Proposition 5.** Let \((X_t)_{t \geq 0}\) be a stationary Poisson INAR(1) process according to Example 1. Then the following results hold:

(i) For any \(k \geq 0\), we have \(\mu(k) = \mu a^k + \mu^2\).

(ii) For any \(l \geq k \geq 0\), we have

\[
\mu(k,l) = \mu^2 (a^k + a^{l-k}) + \mu (\mu + 1) a^l + \mu^3.
\]

(iii) For any \(m \geq l \geq k \geq 0\), we have

\[
\mu(k,l,m) = \mu^3 (a^k + a^{l-k} + a^{m-l}) + \mu^2 (a^{k-l+m} + 2 a^{l+m-k}) + \mu^2 (\mu + 1) (a^{m-k} + a^l) + \mu (\mu + 1)^2 a^m + \mu^4.
\]

(iv) For any \(n \geq m \geq l \geq k \geq 0\), we have

\[
\mu(k,l,m,n) = \mu^4 (a^k + a^{l-k} + a^{m-l} + a^{n-m}) + 2 \mu^2 (\mu + 2) a^{m-n-k} + \mu^3 (a^{k+m-l} + a^{l+n-k-m} + a^{k-m+n} + a^{l-m+n} + 2 a^{l+m-k} + 2 a^{m+n-k-l}) + \mu^2 a^{l-m+n} + \mu^2 (\mu + 1) (a^{k-l+n} + 2 a^{l+n-k} + 2 a^{m+n-l}) + \mu (\mu + 1)^3 a^n + \mu^3 (\mu + 1) (a^{m-k} + a^l + a^{n-l}) + \mu^2 (\mu + 1)^2 (a^{n-k} + a^m) + \mu^5.
\]

(v) For any \(o \geq n \geq m \geq l \geq k \geq 0\), we have

\[
\mu(k,l,m,n,o) = \mu^5 (a^k + a^{l-k} + a^{m-l} + a^{n-m} + a^{o-n}) + \mu^4 (a^{k-n-o} + a^{l-o-k-n} + a^{k-l+m} + a^{k-m+n} + a^{l+n-k-m} + a^{m+o-l-n} + 2 a^{l+m-k} + 2 a^{m+n-k-l} + 2 a^{n+o-l-m} + 4 a^{l+n+o-k-m} + 6 a^{m+n+o-k-l}) + \mu^3 (1 + \mu)^2 (a^m + a^{n-k} + a^{o-l}) + \mu^2 (1 + \mu)^3 (a^n + a^{o-k}) + \mu (1 + \mu)^4 a^o + \mu^3 (1 + \mu) (a^{k-l+n} + a^{l-m+n} + a^{k-m+o} + a^{l+o-k-m} + a^{l+n+o} + a^{m+o-k-n} + 2 a^{l+n-k} + 2 a^{m+n-o} + 2 a^{m+o-k-l} + 2 a^{n+o-k-m}) + 2 \mu^2 (2 + \mu)^2 a^{n+o-k} + \mu^2 (1 + \mu)^2 (a^{k+o-l} + a^{l-m+o} + a^{m-n+o} + 2 a^{l+o-k} + 2 a^{m+n-o} + 2 a^{n+o-m}) + 2 \mu^3 (2 + \mu) (a^{m+n-k} + a^{n+o-k-l}) + 2 \mu^2 (1 + \mu) (2 + \mu) (a^{m+o-k} + a^{n+o-l} + \mu^6.
\]

The expressions (i)–(iii) for the moments of order \( \leq 4 \) have been derived by Weiß (2012). The remaining proof of Proposition 5 is provided by Appendix A.4.

Now we apply Proposition 5 to Equations (7) and (8), as well as to Proposition 4.
Theorem 1. Let $(X_t)_t$ be a Poisson INAR(1) process according to Example 1, let $S = T$. Then the variance of the forecast error, $\mathbb{V} [\hat{X}_{T+1} - X_{T+1}]$, is the sum of

\[
\mathbb{V} [X_{T+1} - \alpha X_T - \beta] = (1 - \alpha^2) \mu \quad \text{and} \\
\mathbb{V} [\hat{\alpha} - \alpha X_T + (\hat{\beta} - \beta)] \approx \frac{1 - \alpha}{T} (1 + \alpha)(1 + 2\mu) - 1 + O(T^{-2}).
\]

The proof of Theorem 1 is provided by Appendix A.5. It should be pointed out that the given known-parameter variance could also be computed in a different way, recall the discussion in Section 1:

\[
\mathbb{V} [X_{T+1} - \alpha X_T - \beta] = \mathbb{V} [\alpha \circ X_T - \alpha X_T] + \mathbb{V} [\epsilon_{T+1} - \beta].
\]

Note that $\epsilon_{T+1}$ is independent of $X_T$ and the thinning thereof. By the law of total variation and the binomial variance, we have

\[
\mathbb{V} [\alpha \circ X_T - \alpha X_T] = \mathbb{E} [\mathbb{V}[\alpha \circ X_T - \alpha X_T | X_T]] = \mathbb{E} [\alpha(1 - \alpha) X_T].
\]

So altogether,

\[
\mathbb{V} [X_{T+1} - \alpha X_T - \beta] = (\alpha(1 - \alpha) \mu + \mu(1 - \alpha)).
\]

Note that for the ordinary AR(1) process, the first term would vanish. Hence, INAR(1) forecasting suffers from stronger uncertainty than AR(1) forecasting does. Actually, in Fuller and Hasza (1981), we find the approximate expression $\mathbb{V} [\hat{X}_{T+1} - X_{T+1}] \approx \sigma_e^2 \left(1 + \frac{2}{T}\right)$ for the Gaussian AR(1) process. Setting $\sigma_e^2 = \mu(1 - \alpha)$ like in the Poisson INAR(1) case ("e-method" in the sense of Homburg et al. (2019)), then the Poisson INAR(1)'s variance according to Theorem 1 is larger than the Gaussian AR(1)'s one by the summand $\alpha(1 - \alpha) \left( \mu + \frac{14\sigma_\alpha^2}{T} \right)$. This summand covers the additional randomness of the DGP as well as the additional uncertainty of its parameter estimators.

Remark 2. Let us recall Remark 1 concerning the case $S < T$. There, it was shown that the variance $\mathbb{V} [(\hat{\alpha} - \alpha) X_T]$ computes as a sum consisting of two terms:

1. $\mathbb{E} [(\hat{\alpha} - \alpha)^2 \mathbb{V}[X_T | X_S, \ldots]]$: Under a Poisson INAR(1) assumption, it is known (Freeland & McCabe, 2004) that

\[
\mathbb{V}[X_T | X_S, \ldots] = \alpha^{T-S} (1 - \alpha^{T-S}) \cdot X_S + \mu (1 - \alpha^{T-S}).
\]

So, in this case, we can compute $\mathbb{E} [(\hat{\alpha} - \alpha)^2 \mathbb{V}[X_T | X_S, \ldots]]$ as

\[
\alpha^{T-S} (1 - \alpha^{T-S}) \cdot \mathbb{E} [(\hat{\alpha} - \alpha)^2 X_S] + \mu (1 - \alpha^{T-S}) \cdot \mathbb{E} [(\hat{\alpha} - \alpha)^2].
\]

2. $\mathbb{V} [(\hat{\alpha} - \alpha) (\alpha^{T-S} X_S + (1 - \alpha^{T-S}) \mu)]$: This variance can be split into the three summands $\alpha^{2(T-S)} \mathbb{V} [(\hat{\alpha} - \alpha) X_S], (1 - \alpha^{T-S})^2 \mu^2 \mathbb{V} [\hat{\alpha} - \alpha],$ and $2 \alpha^{T-S} (1 - \alpha^{T-S}) \mu \mathbb{Cov} [(\hat{\alpha} - \alpha) X_S, \hat{\alpha} - \alpha].$
Hence, it becomes clear that additional effort is required if intending to derive an asymptotic approximation for the case \( S < T \): the moment \( \mathbb{E} [(\hat{\alpha} - \alpha)^2 X_S] \) needs to be computed in analogy to Lemma 3, and an approximation for \( \nabla [(\hat{\alpha} - \alpha)] \) is required, where the latter is provided by Weiß and Schweer (2016).

Although a general extension to \( T > S \) would require additional effort, some assertions for \( T \gg S \) immediately follow. Since then \( a^{T-S} \approx 0 \), the mean forecast error, \( \mathbb{E} [\hat{X}_{T+1} - X_{T+1}] \), follows from Equation (5) as

\[
\mathbb{E} [\hat{X}_{T+1} - X_{T+1}] \approx \mu \cdot \mathbb{E} [\hat{\alpha} - \alpha] + \mathbb{E} [\hat{\beta} - \beta].
\]

The required bias expressions are known from Weiß and Schweer (2016) as

\[
\mathbb{E} [\hat{\alpha} - \alpha] \approx -\frac{1}{S} \left( 1 + 3 \alpha + \frac{\alpha}{\mu} \right), \quad \mathbb{E} [\hat{\beta} - \beta] \approx \frac{1}{S} \left( 1 + 3 \alpha \right) \mu.
\]

So altogether,

\[
\mathbb{E} [\hat{X}_{T+1} - X_{T+1}] \approx -\frac{\alpha}{S} \quad \text{for} \quad T \gg S,
\]

which becomes more pronounced if \( \alpha \) increases. Regarding \( \mathbb{V} [\hat{X}_{T+1} - X_{T+1}] \) for \( T \gg S \), Remark 2 implies the approximation \( \mathbb{V} [(\hat{\alpha} - \alpha) X_T] \approx \mu (1 + \mu) \mathbb{V} [\hat{\alpha}] \), and Remark 1 implies \( \text{Cov} [(\hat{\alpha} - \alpha) X_T, \hat{\beta} - \beta] \approx \mu \text{Cov} [\hat{\alpha} - \alpha, \hat{\beta} - \beta] \). The asymptotic cross-covariance matrix of \( \hat{\alpha}, \hat{\beta} \) equals

\[
\frac{1}{S} \begin{pmatrix} 1 - \alpha^2 + \frac{a(1-a)}{\mu} & -(1 - \alpha^2) \mu \\ -(1 - \alpha^2) \mu & (1 - \alpha) \mu + (1 - \alpha^2) \mu^2 \end{pmatrix},
\]

see Weiß and Schweer (2016). Thus, (7) implies that

\[
\mathbb{V} [(\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta)] \approx \frac{1}{S} \left( \mu (1 + \mu) \left( 1 - \alpha^2 + \frac{a(1-a)}{\mu} \right) \right.
\]

\[
\left. + (1 - \alpha) \mu + (1 - \alpha^2) \mu^2 - 2 \left( 1 - \alpha^2 \right) \mu^2 \right)
\]

\[
= \frac{1 - \alpha}{S} \left( (1 + \alpha)(1 + 2\mu) - 1 \right)
\]

like in Theorem 1. Hence, this approximation formula seems to be appropriate for many situations \( S \leq T \). In contrast to the forecast bias (10), increasing \( \alpha \) has a decreasing effect on the variance, while it is increased with increasing \( \mu \).

Finally, for \( h \to \infty \), we approximately predict \( \hat{X}_{T+h} \approx \hat{\mu} \), so

\[
\mathbb{V} [\hat{X}_{T+h} - X_{T+h}] \approx \mathbb{V} [\hat{\mu}] + \mathbb{V} [X_{T+h}] \approx \mu + \frac{\frac{1 + \alpha}{S} \mu}{1 - \alpha} \quad \text{for} \quad h \to \infty,
\]

where the last expression follows from Weiß and Schweer (2016).

**Example 2.** Let us discuss a time series consisting of monthly counts of claims caused by burn related injuries in the heavy manufacturing industry (1987–1994, so \( T = 96 \)). The dataset was taken from example 2.5.1 in Freeland (1998), where it was used to illustrate the Poisson
FIGURE 1  Claim counts data from Example 2: time series plot (left) and plot of sample ACF against time lag k (right)

INAR(1) model. A plot of the data and their sample ACF are shown in Figure 1; the moment estimates compute as \( \hat{\mu} \approx 8.60 \) and \( \hat{\alpha} \approx 0.452 \). If decomposing the variance of the forecast error, \( \sqrt{\frac{\hat{X}_{T+1} - X_{T+1}}{\sqrt{6.99}}} \), for the fitted Poisson INAR(1) model, then we obtain from Theorem 1 and Equation (9) that

- the innovations’ variance equals \( \approx 4.71 \),
- the additional variance caused by the binomial thinning operation equals \( \approx 2.13 \), and
- the additional variance caused by the parameter estimation equals \( \approx 0.15 \).

If one would use a Gaussian AR(1) model with innovations’ variance \( \approx 4.71 \) as a base for forecasting instead, a scenario critically discussed by Homburg et al. (2019), then one would approximate \( \sqrt{\frac{\hat{X}_{T+1} - X_{T+1}}{\sqrt{4.81}}} \) (Fuller & Hasza, 1981). Hence, a notable part of the forecast uncertainty would be neglected in this way.

However let us return to the fitted Poisson INAR(1) model. The derived results can now be used to evaluate the uncertainty of the point forecast value \( \hat{x}_{T+1} = \hat{\alpha} x_T + \hat{\mu} (1 - \hat{\alpha}) \) in the given data example, where \( x_T = x_{96} = 11 \) according to Figure 1. So \( \hat{x}_{T+1} = \hat{x}_{97} \approx 9.69 \). An approximate 90% interval is constructed by adding \( \pm 1.645 \) times the forecast standard deviation to \( \hat{x}_{97} \). If ignoring the estimation uncertainty and, thus, using \( \sqrt{6.84} \) for the standard deviation, then we obtain the interval \( \approx [5.39; 13.99] \) for \( \hat{x}_{97} \), which includes the counts \( \{6, \ldots, 13\} \). If considering the estimation uncertainty by using Theorem 1, in contrast, then we construct the approximate interval by adding \( \pm 1.645 \sqrt{6.99} \). This leads to the interval \( \approx [5.34; 14.04] \) for \( \hat{x}_{97} \), now covering the counts \( \{6, \ldots, 14\} \). Hence, accounting for estimation uncertainty leads to a different inference in this particular example. If using Equation (10) for an additional approximate bias correction, then we have to shift this interval by \( +0.005 \), still yielding \( \approx [5.34; 14.04] \).

5 | RESULTS FROM A SIMULATION STUDY

To evaluate the approximations of the forecast-error variance according to Theorem 1 and Equations (11), (12), as well as that of the mean forecast error according to Equation (10), we did a comprehensive simulation study, where Poisson INAR(1) processes with different parametrizations were simulated (always 1 million replications per scenario). The moment estimates of \( \alpha, \beta \)
were computed based on the simulated time series \( x_1, \ldots, x_S \), and the subsequent observations were used to determine the forecast error \( \hat{x}_{T+h} - x_{T+h} \) for different choices of \( T - S \geq 0 \) and \( h \geq 1 \). For each scenario, the one million values of forecast error were used to compute the sample mean and variance thereof. The full simulation results are available from the authors upon request. In what follows, we present representative excerpts thereof to illustrate our general findings.

Although our main interest is in the forecast-error variance, let us start with a brief discussion of the mean forecast error. Probably most important, see the excerpt in Table 1, the mean forecast error is generally very close to 0 such that it might be judged negligible for applications in practice. Nevertheless, a few further conclusions can be drawn. We see that with increasing gap \( T - S \) (first block of columns in Table 1), the mean forecast errors quickly stabilize, namely, at a value close to the approximation (10). Actually, it appears that (10) approximates \( \mathbb{E} [\hat{x}_{T+1} - X_{T+1}] \) reasonably well for any choice of \( T \geq S \), at least if \( S \geq 100 \). Regarding increasing values of the forecast horizon \( h \) (last block of columns in Table 1), we recognize the expected tendency that \( \mathbb{E} [\hat{x}_{T+h} - X_{T+h}] \to 0 \) for increasing \( h \), but the actual mean value is affected by decreasing sample size \( S \) and increasing dependence parameter \( \alpha \).

Next, consider the forecast-error variance \( \mathbb{V} [\hat{x}_{T+h} - X_{T+h}] \). For the 1-step-ahead predictions, the variance approximation according to Theorem 1 or Equation (11), respectively, can be used. From Table 2, we see that this approximation generally works quite well, and the approximation quality improves for decreasing \( \alpha \) and increasing \( S \). Increasing \( T - S \) deteriorates the approximation quality, but at least for \( S \geq 200 \), we again have a close agreement. Although the variance increases with increasing mean \( \mu \) (see Equation 11), the actual approximation quality does not seem to be affected by varying \( \mu \). Thus, only the case \( \mu = 2.5 \) is shown in the plotted tables.

**Table 1** Simulated mean forecast error, \( \mathbb{E} [\hat{x}_{T+h} - X_{T+h}] \), for Poisson INAR(1) DGP with \( \mu = 2.5 \)

| \( \alpha \) | \( \beta \) | \( S \) | \( h = 1, T - S = 0,1,2,5,10 \) | Approx. \( T = S, h = 5 \) |
|---|---|---|---|---|
| 0.25 | 1.875 | 50 | -0.003 -0.006 -0.003 -0.007 -0.005 | -0.003 -0.002 -0.003 -0.001 |
| 100 | -0.003 -0.001 0.001 -0.003 -0.003 | -0.003 0.000 -0.002 0.002 |
| 200 | -0.001 -0.001 0.000 0.000 -0.002 | -0.001 -0.001 0.000 0.000 |
| 500 | 0.002 -0.001 -0.001 -0.001 0.001 | -0.002 -0.001 -0.002 -0.002 |
| 1,000 | 0.001 -0.002 0.001 -0.003 0.000 | 0.001 -0.002 -0.001 0.000 |
| 0.5 | 1.25 | 50 | -0.007 -0.010 -0.010 -0.006 -0.006 | -0.010 -0.007 -0.009 0.000 0.002 |
| 100 | -0.005 -0.003 -0.005 -0.007 -0.004 | -0.005 -0.003 0.002 -0.003 |
| 200 | -0.004 -0.002 -0.004 -0.003 -0.005 | -0.004 -0.002 -0.002 -0.002 |
| 500 | 0.001 -0.003 -0.003 -0.003 -0.002 | 0.001 -0.002 -0.001 0.000 |
| 1,000 | -0.003 -0.001 0.000 -0.002 -0.001 | -0.003 -0.002 0.001 -0.001 |
| 0.75 | 0.625 | 50 | -0.010 -0.011 -0.012 -0.013 -0.012 | -0.010 -0.013 -0.011 -0.004 |
| 100 | -0.006 -0.007 -0.008 -0.008 -0.008 | -0.006 -0.008 -0.012 -0.005 |
| 200 | -0.002 -0.002 -0.003 -0.005 -0.005 | -0.002 -0.002 -0.006 -0.001 |
| 500 | -0.001 -0.001 -0.002 -0.001 0.000 | -0.001 -0.001 -0.001 0.000 |
| 1,000 | 0.000 -0.001 0.000 0.000 -0.001 | 0.000 -0.001 0.000 0.002 |
### TABLE 2  Simulated forecast-error variance, $\sqrt{V[\hat{X}_{T+1} - X_{T+1}]}$, for Poisson INAR(1) DGP with $\mu = 2.5$

| $\alpha$ | $\beta$ | $T - S =$ | $S$ | 0 | 1 | 2 | 5 | 10 | (11) | (9) |
|---------|---------|-----------|-----|----|----|----|----|----|-----|-----|
| 0.25    | 1.875   | 50        | 2.441 | 2.452 | 2.449 | 2.452 | 2.452 | 2.441 | $\sqrt{V[X_{T+1} - \alpha X_T - \beta]}$: |
|         |         | 100       | 2.391 | 2.391 | 2.389 | 2.386 | 2.396 | 2.393 | 2.344 |
|         |         | 200       | 2.375 | 2.361 | 2.370 | 2.368 | 2.370 | 2.368 | Innovations’ variance: |
|         |         | 500       | 2.352 | 2.350 | 2.352 | 2.356 | 2.349 | 2.354 | 1.875 |
|         |         | 1,000     | 2.352 | 2.345 | 2.353 | 2.354 | 2.351 | 2.349 |
| 0.5     | 1.25    | 50        | 1.965 | 1.978 | 1.974 | 1.974 | 1.980 | 1.955 | $\sqrt{V[X_{T+1} - \alpha X_T - \beta]}$: |
|         |         | 100       | 1.920 | 1.920 | 1.919 | 1.919 | 1.917 | 1.915 | 1.875 |
|         |         | 200       | 1.895 | 1.896 | 1.900 | 1.893 | 1.900 | 1.895 | Innovations’ variance: |
|         |         | 500       | 1.884 | 1.882 | 1.886 | 1.884 | 1.883 | 1.883 | 1.250 |
|         |         | 1,000     | 1.880 | 1.882 | 1.880 | 1.880 | 1.882 | 1.879 |
| 0.75    | 0.625   | 50        | 1.164 | 1.169 | 1.173 | 1.182 | 1.183 | 1.141 | $\sqrt{V[X_{T+1} - \alpha X_T - \beta]}$: |
|         |         | 100       | 1.122 | 1.126 | 1.127 | 1.128 | 1.130 | 1.118 | 1.094 |
|         |         | 200       | 1.112 | 1.105 | 1.107 | 1.111 | 1.107 | 1.106 | Innovations’ variance: |
|         |         | 500       | 1.101 | 1.100 | 1.102 | 1.102 | 1.098 | 1.099 | 0.625 |
|         |         | 1,000     | 1.096 | 1.098 | 1.094 | 1.095 | 1.098 | 1.096 |

The necessity for accounting for the estimation uncertainty becomes clear if considering the known-parameter variance $\sqrt{V[\hat{X}_{T+1} - \alpha X_T - \beta]}$ and its decomposition according to (9), see the last column in Table 2. It can be clearly seen that the simulated variance $\sqrt{V[\hat{X}_{T+1} - X_{T+1}]}$ is always considerably larger than $\sqrt{V[X_{T+1} - \alpha X_T - \beta]}$, and this discrepancy further increases with decreasing sample size. Hence, the newly derived approximation formula Equation (11) should be used in practice. If computing, for example, a $k$-$\sigma$ prediction interval without considering the estimation uncertainty, there is a higher risk of falling below the intended coverage level. It is also insightful to compare with the pure innovations’ variance (this would be the only source of variation in the AR(1)-case), which is much smaller than $\sqrt{V[X_{T+1} - \alpha X_T - \beta]}$. Hence, for a Poisson INAR(1) process, a major part of the forecast uncertainty is caused by the thinning mechanism.

Finally, let us discuss the $h$-step-ahead forecast-error variance (see Table 3). For $h = 1$, also see the previous discussion, the approximation according to Theorem 1 or Equation (11), respectively, works quite well. Then, for increasing forecast horizon $h$, also the variance quickly increases, namely, toward the value given by Equation (12). Actually, this approximation can also be used for finite values of $h$. As an example, for $h = 10$, a rather good agreement can be observed.

### 6 CONCLUSIONS

We analyzed the distribution of error in mean-based point forecasting of CLAR(1) processes. Forecast error is not only caused by the randomness of the DGP itself, but it is notably affected
TABLE 3  Simulated forecast-error variance, $\mathbb{V} [\hat{X}_{T+h} - X_{T+h}]$, for Poisson INAR(1) DGP with $\mu = 2.5$ and with $T = S$

| $\alpha$ | $\beta$ | $S$ | Approx. ($h = 1$) | 2 | 3 | 4 | 5 | 10 | Approx. ($h = 12$) |
|----------|---------|-----|-----------------|---|---|---|---|----|-----------------|
| 0.25     | 1.875   | 50  | 2.441          | 2.441 | 2.579 | 2.582 | 2.578 | 2.575 | 2.580 | 2.583 |
| 100      | 2.393   |     | 2.391          | 2.530 | 2.536 | 2.538 | 2.551 | 2.538 | 2.542 |
| 200      | 2.368   |     | 2.375          | 2.505 | 2.521 | 2.522 | 2.522 | 2.523 | 2.521 |
| 500      | 2.354   |     | 2.352          | 2.497 | 2.508 | 2.507 | 2.507 | 2.507 | 2.513 | 2.508 |
| 1,000    | 2.349   |     | 2.352          | 2.493 | 2.506 | 2.500 | 2.504 | 2.510 | 2.504 |
| 0.5      | 1.25    | 50  | 1.955          | 1.965 | 2.468 | 2.593 | 2.625 | 2.634 | 2.642 | 2.650 |
| 100      | 1.915   |     | 1.920          | 2.403 | 2.526 | 2.554 | 2.562 | 2.583 | 2.575 |
| 200      | 1.895   |     | 1.895          | 2.376 | 2.500 | 2.528 | 2.534 | 2.539 | 2.538 |
| 500      | 1.883   |     | 1.884          | 2.356 | 2.481 | 2.512 | 2.514 | 2.511 | 2.515 |
| 1,000    | 1.879   |     | 1.880          | 2.355 | 2.467 | 2.494 | 2.502 | 2.507 | 2.508 |
| 0.75     | 0.625   | 50  | 1.141          | 1.164 | 1.855 | 2.250 | 2.471 | 2.603 | 2.791 | 2.850 |
| 100      | 1.118   |     | 1.122          | 1.777 | 2.154 | 2.367 | 2.487 | 2.648 | 2.675 |
| 200      | 1.106   |     | 1.112          | 1.742 | 2.101 | 2.309 | 2.423 | 2.573 | 2.588 |
| 500      | 1.099   |     | 1.101          | 1.727 | 2.080 | 2.274 | 2.385 | 2.531 | 2.535 |
| 1,000    | 1.096   |     | 1.096          | 1.713 | 2.062 | 2.259 | 2.371 | 2.506 | 2.518 |

by the estimation uncertainty of the model parameters. We derived approximate expressions for the mean and variance of the forecast error $\hat{X}_{T+h} - X_{T+h}$, first some general results for CLAR(1) processes, then some closed-form formulas for the special case of a Poisson INAR(1) process. While the forecast bias $\mathbb{E} [\hat{X}_{T+h} - X_{T+h}]$, though slightly negative, is often negligible, the variance becomes quite large (much larger than in the Gaussian AR(1) case). In particular, the decomposition of $\mathbb{V} [\hat{X}_{T+h} - X_{T+h}]$ showed that the estimation uncertainty causes a notable increase in variance compared with the known-parameter case. Thus, for applications in practice, it is highly recommended to account for the effect of parameter estimation if evaluating the uncertainty in forecasting.

ACKNOWLEDGEMENTS
The authors thank the associate editor and the two referees for their useful comments on an earlier draft of this article. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 394832307.

ORCID
Christian H. Weiβ https://orcid.org/0000-0001-8739-6631

REFERENCES
Freeland, R.K. (1998). Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data (PhD thesis). University of British Columbia, Canada. https://open.library.ubc.ca/cIRcle/collections/ubctheses/831/items/1.0088709
Freeland, R. K., & McCabe, B. P. M. (2004). Forecasting discrete valued low count time series. *International Journal of Forecasting, 20*(3), 427–434.

Fuller, W. A., & Hasza, D. P. (1981). Properties of predictors for autoregressive time series. *Journal of the American Statistical Association, 76*(373), 155–161.

Grunwald, G., Hyndman, R. J., Tedesco, L., & Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics, 42*(4), 479–495.

Homburg, A., Weiß, C. H., Alwan, L. C., Frahm, G., & Göb, R. (2019). Evaluating approximate point forecasting of count processes. *Econometrics, 7*(3), 30.

Jung, R. C., & Tremayne, A. R. (2006). Coherent forecasting in integer time series models. *International Journal of Forecasting, 22*(2), 223–238.

Maiti, R., Biswas, A., & Das, S. (2016). Coherent forecasting for count time series using Box-Jenkins’s AR(p) model. *Statistica Neerlandica, 70*(2), 123–145.

Phillips, P. C. B. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics, 9*(3), 241–261.

Ristić, M. M., Bakouch, H. S., & Nastić, A. S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference, 139*(7), 2218–2226.

Sutradhar, B. C. (2008). On forecasting counts. *Journal of Forecasting, 27*(2), 109–129.

Weiß, C. H. (2012). Process capability analysis for serially dependent processes of Poisson counts. *Journal of Statistical Computation and Simulation, 82*(3), 383–404.

Weiß, C. H. (2018). *An introduction to discrete-valued time series*. Chichester, NH: John Wiley & Sons, Inc.

Weiß, C. H., & Schweer, S. (2016). Bias corrections for moment estimators in Poisson INAR(1) and INARCH(1) processes. *Statistics and Probability Letters, 112*, 124–130.

**APPENDIX PROOFS**

**A.1 Proof of Proposition 2**
The first-order partial derivatives of $g_a(y_1, y_2, y_3)$ are

\[
\frac{\partial}{\partial y_1} g_a = \frac{2y_1(y_3 - y_2)}{(y_2 - y_1^2)^2}, \quad \frac{\partial}{\partial y_2} g_a = \frac{y_1^2 - y_3}{(y_2 - y_1^2)^2}, \quad \frac{\partial}{\partial y_3} g_a = \frac{1}{y_2 - y_1^2},
\]

also see Weiß and Schweer (2016). Evaluating the derivatives at $(\mu, \mu(0), \mu(1)) = (\mu, \sigma^2 + \mu^2, \alpha \sigma^2 + \mu^2)$ yields

\[
\frac{\partial}{\partial y_1} g_a(\mu, \mu(0), \mu(1)) = -\frac{2\mu(1 - \alpha)}{\sigma^2}, \quad \frac{\partial}{\partial y_2} g_a(\mu, \mu(0), \mu(1)) = -\frac{\alpha}{\sigma^2},
\]

How to cite this article: Nik S, Weiß CH. CLAR(1) point forecasting under estimation uncertainty. *Statistica Neerlandica*. 2020;74:489–516. https://doi.org/10.1111/stan.12206
\[ \frac{\partial}{\partial y_3} g_{\sigma}(\mu, \mu_0, \mu_1) = \frac{1}{\sigma^2}. \]

Utilizing this for the Taylor approximation (4) leads to

\[ \hat{\alpha} - \alpha \approx \frac{1}{\sigma^2} \sum_{t=1}^{S} \left( -2\mu(1-\alpha)X_t - \alpha X_t^2 + X_t X_{t-1} - \left( -2\mu^2(1-\alpha) - \alpha(\sigma^2 + \mu^2) + \alpha\sigma^2 + \mu^2 \right) \right). \]

Defining \( a := -2\mu(1-\alpha) \) and \( b := -\left( -2\mu^2(1-\alpha) - \alpha(\sigma^2 + \mu^2) + \alpha\sigma^2 + \mu^2 \right) = \mu^2(1-\alpha) \) leads to the linear approximation given in Proposition 2.

Next, we are going to compute the first-order Taylor approximation of \( \hat{\beta} - \beta \) analytically. The first-order partial derivatives of \( g_\beta(y_1, y_2, y_3) \) are

\[ \frac{\partial}{\partial y_1} g_\beta = \frac{y_2 - y_3(y_2 + y_1^2)}{(y_2 - y_1^2)^2}, \quad \frac{\partial}{\partial y_2} g_\beta = \frac{y_1(y_3 - y_1^2)}{(y_2 - y_1^2)^2}, \quad \frac{\partial}{\partial y_3} g_\beta = -\frac{y_1}{y_2 - y_1^2}, \]

also see Weiß and Schweer (2016). Evaluating the derivatives at \((\mu, \mu_0, \mu_1) = (\mu, \sigma^2 + \mu^2, \alpha\sigma^2 + \mu^2)\) yields

\[ \frac{\partial}{\partial y_1} g_\beta(\mu, \mu_0, \mu_1) = \frac{(1-\alpha)(\sigma^2 + 2\mu^2)}{\sigma^2}, \quad \frac{\partial}{\partial y_2} g_\beta(\mu, \mu_0, \mu_1) = \frac{\alpha \mu}{\sigma^2}, \quad \frac{\partial}{\partial y_3} g_\beta(\mu, \mu_0, \mu_1) = -\frac{\mu}{\sigma^2}. \]

Utilizing this for the Taylor approximation (4) leads to

\[ \hat{\beta} - \beta \approx \frac{1}{\sigma^2} \sum_{t=1}^{S} \left( (1-\alpha)(\sigma^2 + 2\mu^2) \cdot X_t + \alpha \mu \cdot X_t^2 - \mu \cdot X_t X_{t-1} - (\mu \cdot (1-\alpha)(\sigma^2 + 2\mu^2) + (\sigma^2 + \mu^2) \cdot \alpha \mu - \mu \cdot (\alpha\sigma^2 + \mu^2)) \right). \]

Defining \( A := (1-\alpha)(\sigma^2 + 2\mu^2) \) and \( B := -\left( \mu \cdot (1-\alpha)(\sigma^2 + 2\mu^2) + (\sigma^2 + \mu^2) \cdot \alpha \mu - \mu \cdot (\alpha\sigma^2 + \mu^2) \right) = -\mu(1-\alpha) \cdot (\sigma^2 + \mu^2) \) leads to the linear approximation given in Proposition 2. This completes the proof.

### A.2 Proof of Proposition 3

To compute \( \mathbb{E}[(\hat{\alpha} - \alpha)X_s] \) analytically, we are going to use the Taylor approximation according to Proposition 2. This yields

\[ \mathbb{E}[(\hat{\alpha} - \alpha)X_s] \approx \mathbb{E} \left[ \left( \frac{1}{\sigma^2} \sum_{t=1}^{S} \left( a \cdot X_t - \alpha \cdot X_t^2 + X_t X_{t-1} + b \right) \right) X_s \right] \]

\[ = \frac{1}{\sigma^2} \sum_{t=1}^{S} \left( a \cdot \mathbb{E}[X_t] - \alpha \cdot \mathbb{E}[X_t^2] + \mathbb{E}[X_t X_{t-1}] + b \cdot \mathbb{E}[X_s] \right). \]

Using the moment notations (2) leads to

\[ \frac{1}{\sigma^2} \sum_{t=1}^{S} \left( a \cdot \mu_{(S-t)} - \alpha \cdot \mu_{(0,S-t)} + \mu_{(1,S+1-t)} + b \cdot \mu \right). \]
Utilizing Proposition 1 yields

\[
\frac{1}{\sigma^2} \frac{1}{S} \sum_{t=1}^{S} \left( a \cdot (\alpha^{S-t} \sigma^2 + \mu^2) - a \cdot \left( \alpha^{S-t} (\mu_{(0,0)} - \mu_{(0)} + \mu \cdot \mu_{(0)}) \right) \\
+ \alpha^{S-t} \left( \mu_{(1,1)} - \mu \cdot \mu_{(1)} + \mu \cdot \mu_{(1)} + b \cdot \mu \right) \\
= \frac{1}{\sigma^2} \frac{1}{S} \sum_{t=1}^{S} \left( \alpha^{S-t} (\mu_{(1,1)} - \alpha \cdot \mu_{(0,0)} + a \cdot \sigma^2 + \mu \cdot (\alpha \cdot \mu_{(0)} - \mu_{(1)})) \right) \\
+ \mu \cdot \left( a \cdot \mu + b + \mu_{(1)} - \alpha \cdot \mu_{(0)} \right) \right).
\]

Inserting \( \mu_{(0)} \) and \( \mu_{(1)} \) as well as the definitions of \( a \) and \( b \) leads to

\[
\frac{1}{\sigma^2} \frac{1}{S} \sum_{t=1}^{S} \left( \alpha^{S-t} (\mu_{(1,1)} - \alpha \cdot \mu_{(0,0)} - 2\mu(1-\alpha) \cdot \sigma^2 - \mu \cdot \mu^2(1 - \alpha)) \right) \\
+ \mu \cdot \left( -2\mu^2(1 - \alpha) + \mu^2(1 - \alpha) + \mu^2(1 - \alpha) \right) \right)
\]

\[
= \frac{1}{\sigma^2} \frac{1}{S} \left( \mu_{(1,1)} - \alpha \cdot \mu_{(0,0)} - \mu(1 - \alpha)(\mu^2 + 2\sigma^2) \right) \frac{1}{1 - \alpha} \sum_{t=1}^{S} \alpha^{S-t} \\
= \frac{1}{\sigma^2} \frac{1}{S} \left( \mu_{(1,1)} - \alpha \cdot \mu_{(0,0)} - \mu(1 - \alpha)(\sigma^2 + \mu_{(0)})) \right) \frac{1}{1 - \alpha} \sum_{t=1}^{S} \alpha^{S-t} \\
= -\frac{1}{\sigma^2} \mu(1 - \alpha^S) - \frac{1}{S} \frac{1}{1 - \alpha} \left( \alpha \cdot \mu_{(0,0)} - \mu_{(1,1)} + \mu(1 - \alpha) \cdot \mu_{(0)} \right).
\]

Using the fact that \( \mu_{(0,1)} = \alpha \cdot \mu_{(0,0)} + \mu(1 - \alpha) \cdot \mu_{(0)} \), see Proposition 1, yields the formula given in Proposition 3.

### A.3 Proof of Proposition 4

To prove Proposition 4, we first derive the three auxiliary results given in Lemmas 1–3.

**Lemma 1.** Let \( (X_t)_{t \in \mathbb{N}} \) be a CLAR(1) process according to Definition 1, let \( S = T \). Then,

\[
\mathbb{E}[(\hat{\beta} - \beta)^2] \approx \frac{\mu b}{\sigma^4} (A \cdot \mu - B) + \frac{1}{T} \frac{A^2}{\sigma^2} \frac{1 + \alpha}{1 - \alpha} - \frac{1 - \alpha T}{T^2} \frac{A^2}{\sigma^2} \frac{2\alpha}{(1 - \alpha)^2} \\
+ \frac{1}{\sigma^2} \frac{1}{T^2} \sum_{s,t=1}^{T} (\alpha^2 \mu^2 \mathbb{E}[X_t^2 X_s^2] + \mu^2 \mathbb{E}[X_t X_{t-1} X_s X_{s-1}]) \\
+ 2\mu A \left( \alpha \mathbb{E}[X_t^2 X_s] - \mathbb{E}[X_t X_{t-1} X_{s-1}] \right) - 2\mu \mathbb{E}[X_t^2 X_s X_{s-1}]),
\]

where \( A, B \) are defined in Proposition 2.

**Proof.** Using the linear approximations according to Proposition 2, we obtain

\[
\mathbb{E}[(\hat{\beta} - \beta)^2] \approx \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \mathbb{E} \left[ (A \cdot X_t + \alpha \mu \cdot X_t^2 - \mu \cdot X_{t-1} + B) \right. \\
\left. \cdot (A \cdot X_s + \alpha \mu \cdot X_s^2 - \mu \cdot X_{s-1} + B) \right].
\]
Expanding this product yields

$$\mathbb{E}[(\hat{\beta} - \beta)^2] \approx \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} (B^2 + 2AB \cdot \mathbb{E}[X_t] + 2\mu B \cdot \mathbb{E}[X_t^2]$$

$$- 2\mu B \cdot \mathbb{E}[X_t X_{t-1}] + A^2 \cdot \mathbb{E}[X_t^2] + 2\mu A \cdot \mathbb{E}[X_t^2 X_{t-1}] - 2\mu A \cdot \mathbb{E}[X_t X_{t-1}]$$

$$+ a^2 \mu^2 \cdot \mathbb{E}[X_t^2 X_{t-1}]^2 - 2a \mu^2 \cdot \mathbb{E}[X_t^2 X_{t-1}] + \mu^2 \cdot \mathbb{E}[X_t X_{t-1} X_{t-1}]$$.

This sum can be split as follows

$$\mathbb{E}[(\hat{\beta} - \beta)^2] \approx \frac{1}{\sigma^4} B \cdot (B + 2\mu(A + a \cdot (\mu(0) - \mu(1))))$$

$$+ \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} (A^2 \cdot \mathbb{E}[X_t X_s] + 2\mu A \cdot (a \cdot \mathbb{E}[X_t^2 X_s] - \mathbb{E}[X_t X_{t-1} X_{t-1}]))$$

$$+ a^2 \mu^2 \cdot \mathbb{E}[X_t^2 X_s^2] - 2a \mu^2 \cdot \mathbb{E}[X_t^2 X_{t-1} X_{t-1}] + \mu^2 \cdot \mathbb{E}[X_t X_{t-1} X_{t-1} X_{t-1}]$$

$$= -\frac{B^2}{\sigma^4} + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \left( A^2 \mathbb{E}[X_t X_s] + a^2 \mu^2 \mathbb{E}[X_t^2 X_s^2] + \mu^2 \mathbb{E}[X_t X_{t-1} X_{t-1} X_{t-1}] \right)$$

$$+ 2\mu A \left( a \mathbb{E}[X_t^2 X_s] - \mathbb{E}[X_t X_{t-1} X_{t-1}]) - 2a \mu^2 \mathbb{E}[X_t^2 X_{t-1} X_{t-1}] \right),$$

where we used the expressions for $\mu(0)$ and $\mu(1)$ according to Proposition 1. Finally, we can explicitly compute

$$\sum_{s,t=1}^{T} \mathbb{E}[X_t X_s] = \sum_{t=1}^{T} \mathbb{E}[X_t^2] + 2 \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \mathbb{E}[X_t X_s] = T \mu(0) + 2 \sum_{s=1}^{T-1} \sum_{k=1}^{T-s} \mu(k)$$

$$= T \mu(0) + T(T - 1) \mu^2 + 2\sigma^2 \sum_{s=1}^{T-1} \sum_{k=1}^{T-s} a^k$$

$$= T \sigma^2 + T^2 \mu^2 + 2\sigma^2 \frac{a}{1 - a} \sum_{s=1}^{T-1} (1 - a^{T-s})$$

$$= T \sigma^2 + T^2 \mu^2 + a \left( T - \frac{1 - a^T}{1 - a} \right)$$

$$= T \sigma^2 \cdot \frac{1 + a}{1 - a} + T^2 \mu^2 - \sigma^2 \left( \frac{2a}{(1 - a)^2} \right) \left( 1 - a^T \right).$$

Thus,

$$-\frac{B^2}{\sigma^4} + \frac{A^2}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \mathbb{E}[X_t X_s] = \frac{1}{\sigma^4} \left( A^2 \mu^2 - B^2 \right) + \frac{A^2}{\sigma^2} \frac{1}{T} \frac{1 + a}{1 - a} - \frac{1 - a^T}{T^2} \frac{2a}{\sigma^2 (1 - a)^2}.$$

Using that $A \mu + B = \mu^3 (1 - a) = \mu b$, we complete the proof of Lemma 1.

\[\text{Lemma 2.}\ \text{Let} \ (X_t)_{0} \text{ be a CLAR}(1) \text{ process according to Definition 1, let} \ S = T. \text{ Then},\]

$$\mathbb{E}[(\hat{\alpha} - a)X_T(\hat{\beta} - \beta)] \approx -\frac{1}{\sigma^4} \mu b B + \frac{1}{\sigma^4} \frac{1}{T} \frac{1 - a^T}{1 - a} \left( \mu A \left( \mu(0,1) - \mu(1,1) \right) + \sigma^2 aB \right)$$
Using the linear approximations according to Proposition 2, we obtain
\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} (aA \cdot \mathbb{E}[X_tX_sX_T] - \alpha^2 \mu \cdot \mathbb{E}[X_t^2X_s^2X_T] - \mu \cdot \mathbb{E}[X_tX_{s-1}X_sX_{s-1}X_T] \]
\[ + (A + 2b) \cdot \left( \mathbb{E}[X_tX_sX_{s-1}X_T] - \alpha \cdot \mathbb{E}[X_t^2X_sX_T] \right) + 2\alpha \mu \cdot \mathbb{E}[X_tX_{s-1}X_s^2X_T] \].

where \( a, b, A, B \) are defined in Proposition 2.

Proof. Using the linear approximations according to Proposition 2, we obtain
\[ \mathbb{E}[(\hat{\alpha} - \alpha) X_T \ (\hat{\beta} - \beta)] \approx \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \mathbb{E} \left[ (a \cdot X_t - \alpha \cdot X_t^2 + X_{s-1} + b) \cdot X_T \cdot \left( A \cdot X_s + \alpha \mu \cdot X_s^2 - \mu \cdot X_{s-1} + B \right) \right]. \]

Expanding this product yields
\[ \mathbb{E}[(\hat{\alpha} - \alpha) X_T \ (\hat{\beta} - \beta)] \approx \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} (bB \cdot \mathbb{E}[X_T] + (bA + aB) \cdot \mathbb{E}[X_sX_T] + aA \cdot \mathbb{E}[X_sX_T] \]
\[ + (A + 2b) \cdot \mathbb{E}[(X_tX_sX_{s-1}X_T)] \]
\[ + (A - \alpha \mu) \cdot \left( \mathbb{E}[X_tX_sX_{s-1}X_T] - \alpha \cdot \mathbb{E}[X_t^2X_sX_T] \right) - \alpha^2 \mu \cdot \mathbb{E}[X_t^2X_s^2X_T] \]
\[ - \mu \cdot \mathbb{E}[X_tX_{s-1}X_sX_{s-1}X_T] + 2\alpha \mu \cdot \mathbb{E}[X_tX_{s-1}X_s^2X_T] \].

Note that \( A - \alpha \mu = A + 2b \) and \( \mu b - B = \mu A \). Then, the sum can be split as follows
\[ \mathbb{E}[(\hat{\alpha} - \alpha) X_T \ (\hat{\beta} - \beta)] \approx \frac{1}{\sigma^4} bB \cdot \mu + \frac{1}{\sigma^4} \frac{1}{T} \sum_{t=1}^{T} \left( (bA + aB) \cdot \mathbb{E}[X_sX_T] + \mu A \right) \]
\[ \left( \mathbb{E}[X_t^2X_s^2X_T] - \mathbb{E}[X_sX_{s-1}X_T] \right) \]
\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \left( aA \cdot \mathbb{E}[X_sX_T] + (A + 2b) \right) \]
\[ \cdot \left( \mathbb{E}[X_tX_sX_{s-1}X_T] - \alpha \cdot \mathbb{E}[X_t^2X_sX_T] \right) \]
\[ - \alpha^2 \mu \cdot \mathbb{E}[X_t^2X_s^2X_T] - \mu \cdot \mathbb{E}[X_tX_{s-1}X_sX_{s-1}X_T] + 2\alpha \mu \cdot \mathbb{E}[X_tX_{s-1}X_s^2X_T] \]
\[ = \frac{1}{\sigma^4} bB \cdot \mu + \frac{1}{\sigma^4} \frac{1}{T} \sum_{t=0}^{T-1} \left( (bA + aB) \cdot \mu(k) + \mu A \cdot \left( \alpha \cdot \mu(k) - \mu(1,k+1) \right) \right) \]
\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \left( aA \cdot \mathbb{E}[X_sX_T] - \alpha^2 \mu \cdot \mathbb{E}[X_t^2X_s^2X_T] \right) \]
\[ - \mu \cdot \mathbb{E}[X_tX_{s-1}X_sX_{s-1}X_T] \]
\[ + (A + 2b) \cdot \left( \mathbb{E}[X_tX_sX_{s-1}X_T] - \alpha \cdot \mathbb{E}[X_t^2X_sX_T] \right) \]
\[ + 2\alpha \mu \cdot \mathbb{E}[X_tX_{s-1}X_s^2X_T] \].

Now, we can apply Proposition 1 in the same way as in the proof of Proposition 3:
\[ \alpha \mu_{(0,k)} - \mu_{(1,k+1)} = \alpha^k \left( \alpha \mu_{(0,0)} - \mu_{(1,1)} - \mu (\alpha \mu_{(0)} - \mu_{(1)}) \right) + \mu (\alpha \mu_{(0)} - \mu_{(1)}) \]
\[ a^k \left( \alpha \mu_{(0,0)} - \mu_{(1,1)} + \mu^3 (1 - \alpha) \right) - \mu^3 (1 - \alpha). \]

Using that \( \alpha \cdot \mu_{(0,0)} = \mu_{(0,1)} - \mu(1 - \alpha) \cdot \mu(0), \) we get
\[
\alpha \mu_{(0,k)} - \mu_{(1,k+1)} = a^k \left( \mu_{(0,1)} - \mu_{(1,1)} - \mu(1 - \alpha) \cdot \sigma^2 \right) - \mu^3 (1 - \alpha)
= a^k \left( \mu_{(0,1)} - \mu_{(1,1)} \right) - \mu(1 - \alpha) \cdot \mu(k).
\]

Thus, using that \( b = \mu^2 (1 - \alpha), \) we get
\[
(b \ A + a \ B) \cdot \mu(k) + \mu A \cdot \left( \alpha \cdot \mu_{(0,k)} - \mu_{(1,k+1)} \right)
= (b \ A + a \ B - \mu^2 (1 - \alpha) A) \cdot \mu(k) + \mu A \left( \mu_{(0,1)} - \mu_{(1,1)} \right) \cdot a^k
= \mu A \left( \mu_{(0,1)} - \mu_{(1,1)} \right) \cdot a^k + \sigma^2 \ a B \cdot a^k + \mu^2 \ a B.
\]

Using that \( \mu a = -2b \) and applying the geometric sum formula completes the proof. \( \blacksquare \)

**Lemma 3.** Let \((X_t)_{t \geq 0}\) be a CLAR(1) process according to Definition 1, let \( S = T. \) Then,
\[
\mathbb{E}[(\hat{\alpha} - \alpha)^2 X_T^2] \approx \frac{1}{\sigma^4} \mu b B + \frac{2b}{\sigma^4} \frac{1}{T} \sum_{k=0}^{T-1} \left( a \mu_{(k,k)} - \alpha \mu_{(0,k,k)} + \mu_{(1,k+1,k+1)} \right)
+ \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \left( a^2 \mathbb{E}[X_t X_s X_T^2] + a^2 \mathbb{E}[X_t^2 X_s^2 X_T^2] + \mathbb{E}[X_t X_{t-1} X_s X_{s-1} X_T^2] \right)
- 2 \alpha a \mathbb{E}[X_t X_s X_T^2] + 2 a \mathbb{E}[X_t X_{t-1} X_T^2] - 2 a \mathbb{E}[X_t X_{t-1} X_s X_{s-1} X_T^2],
\]
where \( a, b \) are defined in Proposition 2.

**Proof.** Using the linear approximations according to Proposition 2, the squared moment computes as
\[
\mathbb{E}[(\hat{\alpha} - \alpha)^2 X_T^2] \approx \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{t,s=1}^{T} \mathbb{E} \left[ \left( a \cdot X_t - a \cdot X_t^2 + X_t X_{t-1} + b \right) \cdot \left( a \cdot X_s - a \cdot X_s^2 + X_s X_{s-1} + b \right) X_T^2 \right].
\]

Expanding this product yields
\[
\mathbb{E}[(\hat{\alpha} - \alpha)^2 X_T^2] \approx \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^{T} \left( b^2 \cdot \mathbb{E}[X_T^2] + 2ab \cdot \mathbb{E}[X_t X_T^2] + a^2 \cdot \mathbb{E}[X_t X_s X_T^2] - 2ab \cdot \mathbb{E}[X_t^2 X_T^2] \right)
+ 2b \cdot \mathbb{E}[X_t X_{t-1} X_T^2] - 2aa \cdot \mathbb{E}[X_T X_s X_T^2] + 2a \cdot \mathbb{E}[X_t X_{t-1} X_s^2]
+ a^2 \cdot \mathbb{E}[X_t^2 X_s^2 X_T^2] - 2a \cdot \mathbb{E}[X_t X_{t-1} X_s^2 X_T^2] + \mathbb{E}[X_t X_{t-1} X_s X_{s-1} X_T^2].
\]

This sum can be split as follows
\[
\mathbb{E}[(\hat{\alpha} - \alpha)^2 X_T^2] \approx \frac{1}{\sigma^4} b^2 \cdot \mathbb{E}[X_T^2] + \frac{1}{\sigma^4} \frac{1}{T} \sum_{t=1}^{T} \left( 2ab \cdot \mathbb{E}[X_t X_T^2] - 2ab \cdot \mathbb{E}[X_T^2 X_T^2] + 2b \cdot \mathbb{E}[X_t X_{t-1} X_T^2] \right)
\]
\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^T \left( a^2 \cdot \mathbb{E}[X_sX_sT] - 2aa \cdot \mathbb{E}[X_sX_{s-1}X_T] + 2a \cdot \mathbb{E}[X_sX_{s-1}X_T^2] \right) \]

\[ + \frac{b^2}{\sigma^4} \mu_0 + \frac{2b}{\sigma^4} \frac{1}{T} \sum_{t=1}^T \left( a \cdot \mu_{(T-t,T-t)} - \mu_0 \cdot \mu_{(0,T-t)} + \mu_{(1,T-t+1,T-t+1)} \right) \]

\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^T \left( a^2 \cdot \mathbb{E}[X_sX_sT] + \sigma^2 \cdot \mathbb{E}[X_sX_sT] + \mathbb{E}[X_sX_{s-1}X_sX_{s-1}X_T] \right) \]

\[ - 2aa \cdot \mathbb{E}[X_sX_sT] + 2a \cdot \mathbb{E}[X_sX_{s-1}X_T] - 2a \cdot \mathbb{E}[X_sX_{s-1}X_T^2] \].

Using that \( b \mu_0 = -\mu B \), we complete the proof of Lemma 3.

Now we combine the expressions derived in Lemmas 1–3 according to

\[ \mathbb{E} \left[ (\hat{\alpha} - \alpha)^2 X_T^2 \right] + \mathbb{E} \left[ (\hat{\beta} - \beta)^2 \right] + 2 \cdot \mathbb{E} \left[ (\hat{\alpha} - \alpha) X_T (\hat{\beta} - \beta) \right] . \]

This yields

\[ \forall \left[ (\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta) \right] \approx \frac{\mu b}{\sigma^4} (A \mu - B) + \frac{1}{T} \frac{A^2}{\sigma^2} \frac{1 + \alpha}{1 - \alpha} + O(T^{-2}) \]

\[ - 2 \frac{\mu b}{\sigma^4} B + \frac{2}{\sigma^4} \frac{1}{T} \frac{1 - \alpha^T}{1 - \alpha} \left( A \left( \mu_{(0,1)} - \mu_{(1,1)} \right) + \sigma^2 a B \right) \]

\[ - \frac{\mu b}{\sigma^4} B + \frac{2b}{\sigma^4} \frac{1}{T} \sum_{k=0}^{T-1} \left( a \mu_{(k,k)} - \alpha \mu_{(0,k,k)} + \mu_{(1,k+1,k+1)} \right) \]

\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^T \left( a^2 \cdot \mathbb{E}[X_sX_sT] + \sigma^2 \cdot \mathbb{E}[X_sX_{s-1}X_sX_{s-1}] \right) \]

\[ + 2 \frac{\mu A}{\sigma^4} \left( \alpha \mathbb{E}[X_sX_s] - \mathbb{E}[X_sX_{s-1}] \right) - 2 \alpha \mu^2 \mathbb{E}[X_sX_{s-1}] \]

\[ + 2aA \cdot \mathbb{E}[X_sX_T] - 2a^2 \mu \cdot \mathbb{E}[X_sX_sX_T] - 2 \mu \cdot \mathbb{E}[X_sX_{s-1}X_sX_{s-1}X_T] \]

\[ + 2(A + 2b) \left( \mathbb{E}[X_sX_{s-1}X_T] - \alpha \cdot \mathbb{E}[X_sX_T] \right) + 4a \mu \cdot \mathbb{E}[X_sX_{s-1}X_sX_{s-1}] \]

\[ + a^2 \mathbb{E}[X_sX_sX_T^2] + a^2 \mathbb{E}[X_sX_sX_T] + \mathbb{E}[X_sX_{s-1}X_sX_{s-1}] \]

\[ - 2a \alpha \mathbb{E}[X_sX_sX_sX_{s-1}] + 2a \mathbb{E}[X_sX_{s-1}X_sX_{s-1}] - 2 \alpha \mathbb{E}[X_sX_{s-1}X_sX_{s-1}] . \]

Collecting terms of equal type leads to

\[ \forall \left[ (\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta) \right] \approx \frac{1}{T} \frac{A^2}{\sigma^2} \frac{1 + \alpha}{1 - \alpha} + \frac{2}{T} \frac{1}{1 - \alpha} \left( \mu A \left( \mu_{(0,1)} - \mu_{(1,1)} \right) + \sigma^2 a B \right) + O(T^{-2}) \]

\[ + \frac{\mu b}{\sigma^4} (A \mu - 4 B) + \frac{2b}{\sigma^4} \frac{1}{T} \sum_{k=0}^{T-1} \left( a \mu_{(k,k)} - \alpha \mu_{(0,k,k)} + \mu_{(1,k+1,k+1)} \right) \]

\[ + \frac{1}{\sigma^4} \frac{1}{T^2} \sum_{s,t=1}^T \left( 2aA \cdot \mathbb{E}[X_sX_sX_T] + a^2 \mathbb{E}[X_sX_sX_T] \right) \]

\[ + a^2 \mu^2 \mathbb{E}[X_sX_sX_T^2] - 2a^2 \mu \cdot \mathbb{E}[X_sX_sX_T^2] + a^2 \mathbb{E}[X_sX_sX_T^2] \]

\[ + \mu^2 \mathbb{E}[X_sX_{s-1}X_sX_{s-1}] - 2 \mu \cdot \mathbb{E}[X_sX_{s-1}X_sX_{s-1}X_T] \]
+ \mathbb{E}[X_t X_{t-1} X_{t-2} X_{T}] \\
+ 2 \mu A \alpha \mathbb{E}[X_t^2 X_{t}] - 2(A + 2b) \alpha \mathbb{E}[X_t^2 X_{X_t}] - 2 \alpha a \mathbb{E}[X_t^2 X_{X_t}^2] \\
- 2 \mu A \mathbb{E}[X_t X_{X_t} X_{t-1}] + 2(A + 2b) \mathbb{E}[X_t X_{X_{t-1}} X_{T}] \\
+ 2 \alpha \mathbb{E}[X_t X_{X_t} X_{t-1} X_{T}^2] \\
- 2 \alpha \mu^2 \mathbb{E}[X_t^2 X_{X_t} X_{t-1}] + 4\alpha \mu \cdot \mathbb{E}[X_t^2 X_{X_{t-1}} X_{T}] \\
- 2 \alpha \mathbb{E}[X_t^2 X_{X_{t-1}} X_{T}^2] \). \\

This concludes the proof of Proposition 4.

A.4 Proof of Proposition 5

It remains to prove the expressions for the fifth- and sixth-order moment given in parts (iv) and (v), respectively. Both formulas are derived in a completely analogous way. To save some space, we only present the proof of part (iv) for illustration.

To derive an explicit expression for the fifth-order moment \( \mu_{(k,l,m,n)} \) with \( 0 \leq k \leq l \leq m \leq n \), we need the following conditional moments:

\[
\mathbb{E}[X_t | X_{t-1}, \ldots] = \alpha X_{t-1} + \mu(1 - \alpha), \\
\mathbb{E}[X_t^2 | X_{t-1}, \ldots] = \alpha^2 X_{t-1}^2 + \alpha(1 - \alpha) \mu X_{t-1} + \mu(1 - \alpha)(1 + \alpha(1 - \alpha)), \\
\mathbb{E}[X_t^3 | X_{t-1}, \ldots] = \alpha^3 X_{t-1}^3 + 3\alpha^2(1 - \alpha)(1 + \mu) X_{t-1}^2 \\
+ \alpha(1 - \alpha) \left(1 - 2\alpha + 3(2 - \alpha) \mu + 3(1 - \alpha) \mu^2\right) X_{t-1} \\
+ \mu(1 - \alpha) \left(1 + 3\mu(1 - \alpha) + \mu^2(1 - \alpha)^2\right), \\
\mathbb{E}[X_t^4 | X_{t-1}, \ldots] = \alpha^4 X_{t-1}^4 + 2\alpha^3(1 - \alpha)(2\mu + 3) X_{t-1}^3 \\
+ \alpha^2(1 - \alpha) \left(7 + 18\mu + 6\mu^2 - \alpha(11 + 12\mu + 6\mu^2)\right) X_{t-1}^2 \\
+ \alpha(1 - \alpha) \left(1 + 14\mu + 18\mu^2 + 4\mu^3 - 2\alpha \left(4\mu^3 + 12\mu^2 + 9\mu + 3\right) \\
+ \alpha^2 \left(4\mu^3 + 6\mu^2 + 8\mu + 6\right)\right) X_{t-1} \\
+ \mu(1 - \alpha) \left(1 + 7(1 - \alpha) \mu + 6(1 - \alpha)^2 \mu^2 + (1 - \alpha)^3 \mu^3\right).
\]

These expressions are easily derived from moments of the binomial and the Poisson distribution, since the conditional distribution of \( \alpha \circ X_{t-1} \) given \( X_{t-1}, \ldots \) is Bin(\( X_{t-1}, \alpha \)), that of \( \epsilon_1 \) given \( X_{t-1}, \ldots \) simply Poi(\( \mu(1 - \alpha) \)).

For the conditional moments given before, we introduce the notation

\[
\mathbb{E} \left[ X_t^n | X_{t-1}, \ldots \right] = \alpha^n \cdot X_t^n + A_{n-1,n} \cdot X_{t-1}^{n-1} + A_{n-2,n} \cdot X_{t-1}^{n-2} + \cdots + A_{0,n} \tag{A1}
\]

for \( n \geq 1 \). Using the law of total expectation, we now distinguish between the following five cases:

Case 1: \( n > m \). Then we have

\[
\mu_{(k,l,m,n)} = \mathbb{E} \left[ X_t \cdot X_{t+k} \cdot X_{t+l} \cdot X_{t+m} \cdot \mathbb{E} \left[ X_t+n | X_{t+n-1}, \ldots \right] \right] \\
= \alpha \cdot \mu_{(k,l,m,n-1)} + \mu(1 - \alpha) \cdot \mu_{(k,l,m)} \\
= \cdots = \alpha^{n-m} \cdot \mu_{(k,l,m,m)} + \mu(1 - \alpha) \cdot \mu_{(k,l,m,m)} \sum_{j=0}^{n-m-1} \alpha^j.
\]
\[ = a^{n-m} \cdot \mu(k,l,m,m) + \mu(1 - a^{n-m}) \cdot \mu(k,l,m). \]

Case 2: \( n = m > l \). Then we have
\[
\mu(k,l,m,m) = \mathbb{E} \left[ X_t \cdot X_{t+k} \cdot \mathbb{E} \left[ X_{t+l} \cdot X_{t+m} \mid X_{t+m-1}, \ldots \right] \right] \\
= a^2 \cdot \mu(k,l,m-1,m-1) + A_{1,2} \cdot \mu(k,l,m-1) + A_{0,2} \cdot \mu(k,l) = \ldots \\
= a^{2(m-l)} \cdot \mu(k,l,l,l) + A_{1,2} \cdot \sum_{j=0}^{m-l-1} a^{2j} \mu(k,l,m-1-j) + A_{0,2} \cdot \mu(k,l) \sum_{j=0}^{m-l-1} a^{2j}.
\]

Part (iii) of Proposition 5 implies that
\[ a^{2j} \mu(k,l,m-1-j) = a_{(k,l)} \cdot a^{2j} + a^{l-1} \left( \mu(k,l,m) - a_{(k,l)} \right), \]
with \( a_{(k,l)} := \mu^4 + \mu^3 \left( a^k + a^{l-k} \right) + \mu^2 (1 + \mu) a^l \). Hence, we get
\[
\mu(k,l,m,m) = a^{2(m-l)} \cdot \mu(k,l,l,l) + \left( a_{(k,l)} \cdot A_{1,2} + A_{0,2} \cdot \mu(k,l) \right) \sum_{j=0}^{m-l-1} a^{2j} \\
+ \frac{A_{1,2}}{\alpha} \cdot \left( \mu(k,l,m) - a_{(k,l)} \right) \sum_{j=0}^{m-l-1} a^j \\
= a^{2(m-l)} \cdot \mu(k,l,l,l) + \left( a_{(k,l)} \cdot A_{1,2} + A_{0,2} \cdot \mu(k,l) \right) \sum_{j=0}^{m-l-1} a^{2j} \\
+ (1 + 2\mu)(1 - a^{m-l}) \left( \mu(k,l,m) - a_{(k,l)} \right).
\]

Case 3: \( n = m = l > k \). Then we have
\[
\mu(k,l,l,l) = \mathbb{E} \left[ X_t \cdot X_{t+k} \cdot \mathbb{E} \left[ X_{t+l} \cdot X_{t+l-1}, \ldots \right] \right] \\
= a^3 \cdot \mu(k,l-1,l-1,l-1) + A_{2,3} \cdot \mu(k,l-1,l-1) + A_{1,3} \cdot \mu(k,l-1) + A_{0,3} \cdot \mu(k) \\
= \ldots = a^{3(l-k)} \cdot \mu(k,k,k,k) + A_{2,3} \cdot \sum_{j=0}^{l-k-1} a^{3j} \mu(k,l-1-j,l-1-j) \\
+ A_{1,3} \cdot \sum_{j=0}^{l-k-1} a^{3j} \mu(k,l-1-j) + A_{0,3} \cdot \mu(k) \sum_{j=0}^{l-k-1} a^{3j}.
\]

Parts (ii) and (iii) of Proposition 5 imply that
\[ a^{3j} \mu(k,l-1-j) = d_{(k)} \cdot a^{3j} + \left( \mu(k,l) - d_{(k)} \right) \cdot a^{2j-1} \]
\[ a^{3j} \mu(k,l-1-j,l-1-j) = b_{(k)} \cdot a^{3j} + \left( \mu(k,l,l) - c_{(k)} \right) \cdot a^{2j-1} + 2\mu^2 a^{2(l-1)-k} \cdot a^j, \]
with
\[ b_{(k)} := \mu^4 + \mu^3 (1 + a^k) + \mu^2 a^k, \quad c_{(k,l)} := b_{(k)} + 2\mu^2 a^{2l-k}, \quad d_{(k)} := \mu^3 + \mu^2 a^k.\]
This yields
\[
\mu_{(k,l,l,l)} = \alpha^3(l-k) \cdot \mu_{(k,k,k,k)} + \left( b_k \cdot A_{2,3} + d_k \cdot A_{1,3} + A_{0,3} \cdot \mu_k \right) \sum_{j=0}^{l-1} \alpha^j
\]
\[
+ \frac{\left( A_{2,3} \cdot \left( \mu_{(k,l,l)} - c_l \right) + A_{1,3} \cdot \left( \mu_{(k,k)} - d_k \right) \right)}{\alpha} \sum_{j=0}^{l-1} \alpha^j
\]
\[
+ \frac{\left( A_{2,3} \cdot \left( \mu_{(k,l,l)} - c_l \right) + A_{1,3} \cdot \left( \mu_{(k,k)} - d_k \right) \right)}{\alpha} \sum_{j=0}^{l-1} \alpha^j
\]
\[
= \alpha^3(l-k) \cdot \mu_{(k,k,k,k)} + 6(1 + \mu) \mu^2 \left( \alpha^{2l-k} - \alpha^{2l-2k} \right)
\]
\[
+ \frac{1 - \alpha^{3l-k}}{\alpha(1 - \alpha^2)} \left( A_{2,3} \cdot \left( \mu_{(k,l,l)} - c_l \right) + A_{1,3} \cdot \left( \mu_{(k,k)} - d_k \right) \right)
\]
\[
+ \frac{1 - \alpha^{3l-k}}{1 - \alpha^3} \left( b_k \cdot A_{2,3} + d_k \cdot A_{1,3} + A_{0,3} \cdot \mu_k \right).
\]

Case 4: \( n = m = l = k > 0 \). Then we have
\[
\mu_{(k,k,k,k)} = \mathbb{E} \left[ X_t \cdot \mathbb{E} \left[ X_{l+k}^4 | X_{l+k-1}, \ldots \right] \right]
\]
\[
= \alpha^4 \cdot \mu_{(k-1,k-1,k-1,k-1)} + A_{3,4} \cdot \mu_{(k-1,k-1,k-1)}
\]
\[
+ A_{2,4} \cdot \mu_{(k-1,k-1)} + A_{1,4} \cdot \mu_{(k-1)} + A_{0,4} \cdot \mu
\]
\[
= \cdots = \alpha^{4k} \cdot \mu_{(0,0,0,0)} + A_{3,4} \cdot \sum_{j=0}^{k-1} \alpha^j \mu_{(k-1-j,k-1-j,k-1-j)}
\]
\[
+ A_{2,4} \cdot \sum_{j=0}^{k-1} \alpha^j \mu_{(k-1-j,k-1-j)} + A_{1,4} \cdot \sum_{j=0}^{k-1} \alpha^j \mu_{(k-1-j)} + \mu \cdot A_{0,4} \cdot \sum_{j=0}^{k-1} \alpha^j.
\]

Parts (i)-(iii) of Proposition 5 imply that
\[
\alpha^j \mu_{(k-1-j,k-1-j,k-1-j)} = e \cdot \alpha^j + \alpha^{j-1} \left( \mu_{(k,k)} - e \right),
\]
\[
\alpha^j \mu_{(k-1-j,k-1-j)} = f \cdot \alpha^j + \alpha^{j-1} \left( \mu_{(k,k)} - f \right),
\]
\[
\alpha^j \mu_{(k-1-j)} = \mu^2 \alpha^j + \alpha^{j-1} \left( \mu_k - \mu^2 \right),
\]
with \( e := \mu^4 + 3\mu^3 + \mu^2 \) and \( f := \mu^3 + \mu^2 \). This yields
\[
\mu_{(k,k,k,k)} = \alpha^{4k} \cdot \mu_{(0,0,0,0)} + \left( e \cdot A_{3,4} + f \cdot A_{2,4} + \mu^2 \cdot A_{1,4} + \mu \cdot A_{0,4} \right) \sum_{j=0}^{k-1} \alpha^j
\]
\[
+ \frac{\left( A_{3,4} \cdot \left( \mu_{(k,k,k)} - e \right) + A_{2,4} \cdot \left( \mu_{(k,k)} - f \right) + A_{1,4} \cdot \left( \mu_k - \mu^2 \right) \right)}{\alpha} \sum_{j=0}^{k-1} \alpha^j
\]
\[
= \alpha^{4k} \cdot \mu_{(0,0,0,0)} + \frac{1 - \alpha^{4k}}{1 - \alpha^4} \left( e \cdot A_{3,4} + f \cdot A_{2,4} + \mu^2 \cdot A_{1,4} + \mu \cdot A_{0,4} \right)
\]
\[
+ \frac{1 - \alpha^{3k}}{\alpha(1 - \alpha^2)} \left( A_{3,4} \cdot \left( \mu_{(k,k,k)} - e \right) + A_{2,4} \cdot \left( \mu_{(k,k)} - f \right) + A_{1,4} \cdot \left( \mu_k - \mu^2 \right) \right).
\]

Case 5: \( n = m = l = k = 0 \). Because of the Poisson property, we have
\[
\mu_{(0,0,0,0)} = \mathbb{E} [X_t] = \mu^5 + 10\mu^4 + 25\mu^3 + 15\mu^2 + \mu.
\]
Accordingly, the above formula for $\mu(k,k,k,k)$ simplifies to

$$\mu(k,k,k,k) = \mu \left(1 + 14\mu + 18\mu^2 + 4\mu^3\right) a^k + \mu^2 \left(1 + 7\mu + 6\mu^2 + \mu^3\right).$$

Inserting into the formula for $\mu(k,l,l,l)$, we obtain

$$\mu(k,l,l,l) = \mu^5 + \mu^4 \left(3 + a^k + 3a^l + 3a^{l-k}\right) + \mu^3 \left(1 + 3a^k + 9a^l + 6a^{l-k} + 6a^{2l-k}\right) + \mu^2 \left(a^k + 7a^l + a^{l-k} + 6a^{2l-k}\right) + \mu a^l.$$

Plugging-in into $\mu(k,l,m,m)$ leads to

$$\mu(k,l,m,m) = \mu^5 + \mu^4 \left(1 + a^k + a^{l-k} + 2a^{m-k} + a^l + 2a^{m-l} + 2a^m\right) + \mu^3 \left(a^k + a^{l-k} + 2a^{k-l+m} + 4a^{l-k+m} + 2a^{2m-k-l} + 3a^{m-k}\right) + 2a^{2m-k} + 2a^l + a^{m-l} + 2a^{2m-l} + 5a^m) + \mu^2 \left(a^{k-l+m}\right) + 2a^{l-m-k} + a^{m-k} + 4a^{2m-k} + a^l + 2a^{2m-l} + 4a^m) + \mu a^m.$$

Thus, it follows for all $n \geq m \geq l \geq k \geq 0$ that

$$\mu(k,l,m,n) = \mu^5 + \mu^4 \left(a^{l-k} + a^{m-k} + \alpha^{m-n} + \alpha^{m-l} + \alpha^{n-m} + \alpha^m + \alpha^n + \alpha^k + \alpha^l\right) + \mu^3 \left(a^{k-l+m} + 2a^{l+m-k} + a^{l+n-k-m} + 2a^{m+l-n-k-l} + a^{k-l+n}\right) + 2a^{l+m-n} + a^{m-n} + 2a^m + 3a^n) + \mu^2 \left(a^{k-l+n} + 2a^{l+n-k} + 4a^{m+n-k} + \alpha^{n-k} + a^{l+m-n} + 2a^{m+n-l} + \alpha^m + 3\alpha^n\right) + \mu a^n.$$

This completes the proof of part (iv) of Proposition 5.

### A.5 Proof of Theorem 1

For a Poisson INAR(1) process, we have $\sigma^2 = \mu$. Thus, the constants $a, b, A, B$ defined in Proposition 2 simplify to

$$a = -2 \mu (1 - \alpha), \quad b = \mu^2 (1 - \alpha), \quad A = (1 - \alpha) \mu(1 + 2\mu), \quad B = -(1 - \alpha) \mu^2(1 + \mu).$$

Furthermore, $(X_t)_{\mathbb{Z}}$ is time-reversible such that $\mu(0,1) - \mu(1,1) = 0$. It also follows from Proposition 5 that

$$a \mu(k,k) - a \mu(0,k,k) + \mu(1,k+1,k+1) = -b \left((1 + 2\mu) a^k + \mu(\mu + 1)\right).$$

So altogether, Proposition 4 implies that

$$\mathbb{V}[(\hat{\alpha} - \alpha)X_T + (\hat{\beta} - \beta)] \approx \frac{1}{T} A^2 \frac{1 + \alpha}{\mu} + \frac{2}{1 - \alpha} \frac{1 - \alpha}{T} a B + O(T^{-2})$$

$$+ \frac{b}{\mu} (A \mu - 4 B) - \frac{1}{T} \frac{2b^2}{\mu^2} \left(1 + 2\mu\right) \frac{1 - \alpha^T}{1 - \alpha} - \frac{2b^2}{\mu} (\mu + 1)$$
\[ + \frac{1}{\mu^2} \frac{1}{T^2} \sum_{s,t=1}^{T} \left( 2aA \cdot E[X_s X_{sT}] + a^2 E[X_s X_{sT}^2] \right) \]
\[ + \alpha^2 \mu^2 E[X_s^2 X_{sT}^2] - 2a^2 \mu \cdot E[X_s^2 X_{sT} X_{sT}] + \alpha^2 E[X_s^2 X_{sT}^2] \]
\[ + \mu^2 E[X_{sT-1} X_{sT-1} X_{sT}] - 2 \mu \cdot E[X_{sT-1} X_{sT-1} X_{sT} X_{sT}] + E[X_{sT-1} X_{sT-1} X_{sT} X_{sT}] \]
\[ + 2 \mu A \alpha E[X_s X_{sT}] - (2A + 2b) \alpha E[X_s^2 X_{sT}] - 2 \alpha \alpha E[X_s^2 X_{sT}^2] \]
\[ - 2 \mu A E[X_s X_{sT-1}] + (2A + 2b) E[X_{sT-1} X_{sT}] + 2 \alpha E[X_s X_{sT-1} X_{sT}] \]
\[ - 2 \alpha^2 E[X_s^2 X_{sT-1}] + 4a \mu \cdot E[X_s^2 X_{sT-1} X_{sT}] - 2 \alpha E[X_s^2 X_{sT-1} X_{sT}] \] .

The terms in the first two lines can be further summarized to
\[
\frac{1}{T} \frac{A^2}{\mu} \frac{1 + \alpha}{1 - \alpha} + \frac{1}{T} \frac{1 - \alpha^T}{1 - \alpha} (2a B + b a (1 + 2\mu)) + \frac{b}{\mu} (A \mu - 4 B + a \mu (\mu + 1))
\]
\[= \frac{1}{T} \frac{A^2}{\mu} \frac{1 + \alpha}{1 - \alpha} - \frac{1}{T} \frac{1 - \alpha^T}{1 - \alpha} \frac{ab}{\mu} + \frac{b}{\mu} (A \mu - 2B). \]

Hence, we obtain
\[
\forall \left((\hat{\alpha} - \alpha) X_T + (\hat{\beta} - \beta)\right) \approx \frac{1}{T} \frac{A^2}{\mu} \frac{1 + \alpha}{1 - \alpha} - \frac{1}{T} \frac{1 - \alpha^T}{1 - \alpha} \frac{ab}{\mu} + O(T^{-2})
\]
\[+ \frac{b}{\mu} (A \mu - 2B) + \frac{1}{\mu^2} \frac{1}{T^2} \sum_{s,t=1}^{T} (f(s, t) + g(s, t)), \]

where
\[ f(s, t) = 2aA \cdot E[X_s X_{sT}] + a^2 E[X_s X_{sT}^2] + \alpha^2 \mu^2 E[X_s^2 X_{sT}^2] \]
\[ - 2a^2 \mu \cdot E[X_s^2 X_{sT} X_{sT}] + \alpha^2 \mu^2 E[X_s^2 X_{sT}^2] + \mu^2 E[X_{sT-1} X_{sT-1} X_{sT}] \]
\[ - 2 \mu \cdot E[X_{sT-1} X_{sT-1} X_{sT}] + E[X_{sT-1} X_{sT-1} X_{sT} X_{sT}] \]

comprises the terms being symmetric in \(s\) and \(t\), and where
\[ g(s, t) = 2 \mu A \alpha E[X_s X_{sT}] - (2A + 2b) \alpha E[X_s^2 X_{sT}] - 2 \alpha \alpha E[X_s^2 X_{sT}^2] \]
\[ - 2 \mu A E[X_s X_{sT-1}] + (2A + 2b) E[X_s X_{sT-1} X_{sT}] + 2 \alpha E[X_s X_{sT-1} X_{sT}] \]
\[ - 2 \alpha^2 E[X_s^2 X_{sT-1}] + 4a \mu \cdot E[X_s^2 X_{sT-1} X_{sT}] - 2 \alpha E[X_s^2 X_{sT-1} X_{sT}] \]

comprises the asymmetric terms. These sums are computed as
\[ \sum_{s,t=1}^{T} f(s, t) = \sum_{t=1}^{T} f(t, t) + 2 \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} f(s, t) \quad (A2) \]

and
\[ \sum_{s,t=1}^{T} g(s, t) = \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} g(s, t) + \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} g(s, t), \quad (A3) \]
respectively. We obtain the following expressions:

\[
\sum_{t=1}^{T} f(t, t) = \sum_{t=1}^{T} \left( \alpha^2 \mu^2 \mathbb{E}[X_t^4] + \mu^2 \mathbb{E}[X_t^2 X_{t-1}^2] + 2\alpha A \cdot \mathbb{E}[X_t^2 X_{T+t}] + \alpha^2 \mathbb{E}[X_t^2 X_T^2] \\
-2\alpha^2 \mu \cdot \mathbb{E}[X_t^4 X_T] + \alpha^2 \mathbb{E}[X_t^2 X_{T+t}] - 2\mu \cdot \mathbb{E}[X_t^2 X_{T+t} X_T] + \mathbb{E}[X_t^2 X_{T+t}^2 X_T] \right)
= T \alpha^2 \mu^2 \mu_{(0,0,0)} + T \mu^2 \mu_{(0,1,1)} + \sum_{k=0}^{T-1} (2\alpha A \cdot \mu_{(0,k)} + \alpha^2 \mu_{(0,k,k)})
-2\alpha^2 \mu \mu_{(0,0,0,k)} + \alpha^2 \mu_{(0,0,0,k,k)} - 2\mu \cdot \mu_{(0,1,1,k+1)} + \mu_{(0,1,1,k+1,k+1)} ,
\]

and analogously,

\[
\sum_{t=1}^{T} \sum_{s=t+1}^{T} f(s, t) = \sum_{t=1}^{T} \sum_{s=t+1}^{T} \left( 2\alpha A \mu_{(t-s,T-s)} + \alpha^2 \mu_{(t-s,T-s,T-s)} \right)
+ \alpha^2 \mu^2 \mu_{(0,t-s,T-s)} - 2\alpha^2 \mu \mu_{(0,t-s,T-s,t)} + \alpha^2 \mu_{(0,t-s-T,s,T-s)} + 2\mu \mu_{(1,t-s,T-s+1,T-s+1)} + \mu_{(1,t-s,T-s+1,T-s+1,T-s+1)} \right) \right)
= \sum_{l=1}^{T-1} \sum_{k=1}^{T-l} \left( 2\alpha A \mu_{(k,l)} + \alpha^2 \mu_{(k,l,l)} + \alpha^2 \mu_{(0,k,k,k)} - 2\alpha \mu_{(0,k,k,l,l)} \right)
+ \alpha^2 \mu_{(0,k,k,l,l,l)} + \mu_{(1,k,k+1,l,l)} - 2\mu \mu_{(1,k,k+1,l+1,l+1)} + \mu_{(1,k,k+1,l+1,l+1,l+1)} ,
\]

where we replaced \( k = t - s \) and \( l = T - s \). Similarly,

\[
\sum_{t=1}^{T} \sum_{s=t}^{T} g(s, t) = \sum_{t=1}^{T} \sum_{s=t}^{T} \left( 2 \alpha A \mu \mu_{(t-s,T-s)} - 2(A+2b) \alpha \mu_{(t-s,T-s,T-s)} \right)
- 2 \alpha \mu \mu_{(t-s,T-s,T-s,T-s)} - 2 \alpha A \mu_{(1,t-s+1)} + 2(A+2b) \mu_{(1,t-s+1,T-s+1)} + \alpha^2 \mu_{(1,t-s+1,l,l)} + 4\alpha \mu_{(1,t-s+1,l+1,l+1)}
+ 2 \alpha \mu_{(1,t-s+1,l+1,s+1,T-s+1)} - 2 \alpha \mu_{(1,t-s+1,l+1,s+1,T-s+1)} \right)
= \sum_{l=1}^{T-1} \sum_{k=0}^{l} \left( 2 \alpha A \mu \mu_{(k,k)} - 2(A+2b) \alpha \mu_{(k,k,l,l)} - 2 \alpha \mu_{(k,k,l,l,l)} \right)
+ 2 \alpha A \mu_{(1,k+1)} + 2(A+2b) \mu_{(1,k+1,l+1)} + 2 \alpha \mu_{(1,k+1,l+1,l+1)}
+ 2 \alpha \mu_{(1,k+1,l+1,l+1,l+1)} + 4\alpha \mu \mu_{(1,k+1,l+1,l+1,l+1)} - 2 \alpha \mu_{(1,k+1,l+1,l+1,l+1,l+1)} ,
\]

again using \( k = t - s \) and \( l = T - s \). Finally,

\[
\sum_{t=1}^{T} \sum_{s=t+1}^{T} g(s, t) = \sum_{t=1}^{T} \sum_{s=t+1}^{T} \left( 2 \alpha A \mu \mu_{(0,s-t)} - 2(A+2b) \alpha \mu_{(0,s-t,T-t)} \right)
- 2 \alpha \mu \mu_{(0,s-t,T-t,T-t)} - 2 \alpha A \mu_{(s-t-1,s-t)} + 2(A+2b) \mu_{(s-t-1,s-t,T-t)} + 2 \alpha \mu_{(s-t-1,s-t,T-t-T-t)} + 2 \alpha \mu_{(s-t-1,s-t-T-t,T-t)}
+ 4\alpha \mu \mu_{(0,s-t-1,s-t-1,s-t-1,T-t-T-t)} - 2 \alpha \mu_{(0,s-t-1,s-t-1,s-t,T-t,T-t)} ,
\]
where this time, we used the transformations \( k = s - t \) and \( l = T - t \). Finally, we insert the moment formulas according to Proposition 5 and compute the sums. Approximating \( \alpha^T \approx 0 \), tedious algebra leads to the expression given in Theorem 1. Hence, the proof is complete.