Statistica Sinica **32** (2022), 1099-1120
doi:https://doi.org/10.5705/ss.202020.0353

SOFTPLUS INGARCH MODELS

Christian H. Weiß[1], Fukang Zhu[2] and Aisouda Hoshiyar[1]

[1]*Helmut Schmidt University and* [2]*Jilin University*

Abstract: Numerous models have been proposed for count time series, including the integer-valued autoregressive moving average (ARMA) and integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) models. However, while both models lead to an ARMA-like autocorrelation function (ACF), the attainable range of ACF values is much more restricted, and negative ACF values are usually not possible. The existing log-linear INGARCH model allows for negative ACF values, but the linear conditional mean and the ARMA-like autocorrelation structure are lost. To resolve this dilemma, a novel family of INGARCH models is proposed that uses the softplus function as a response function. The softplus function is approximately linear, but avoids the drawback of not being differentiable in zero. The stochastic properties of the novel model are derived. The proposed model exhibits an approximately linear structure, confirmed using extensive simulations, which makes its model parameters easier to interpret than those of a log-linear INGARCH model. The asymptotics of the maximum likelihood estimators for the parameters are established, and their finite-sample performance is analyzed using simulations. The usefulness of the proposed model is demonstrated by applying it to three real-data examples.

Key words and phrases: Count time series, INGARCH models, maximum likelihood estimation, negative autocorrelation, softplus link

1. Introduction

Numerous models have been proposed for count time series, that is, quantitative time series, where the range consists of nonnegative integers from the set $\mathbb{N}_0 = \{0, 1, \ldots\}$; recent surveys are provided by Weiß (2018, 2021). Many count time series models are inspired by the traditional autoregressive moving average (ARMA) models for real-valued time series. Some of these adapt the ARMA recursion to the integer case by using so-called “thinning operations”; the resulting models are commonly referred to as integer-valued ARMA (INARMA) models. Others use a regression approach to ensure a linear conditional mean. However, despite their close relation to ARMA models, these models are often referred to as integer-valued generalized autoregressive conditional heteroskedasticity (IN-

Corresponding author: Fukang Zhu, School of Mathematics, Jilin University, Changchun 130012, China. E-mail: zfk8010@163.com.

GARCH) models; see also the discussion on p. 74 of Weiß (2018). Although both INARMA and INGARCH models lead to an ARMA-like autocorrelation structure (i. e., their autocorrelation function (ACF) satisfies a set of Yule–Walker equations), the attainable range of ACF values is often much more restricted than that of the ordinary ARMA models, because negative ACF values are usually not possible. The latter is due to parameter constraints, which, in turn, result from the constraint of nonnegative outcomes (counts) for the process. If negative ACF values are required, conditional regression models with a log link might be used, but then the linear conditional mean and, thus, the ARMA-like ACF are lost.

To resolve this dilemma, we propose a novel family of conditional regression models for stationary count processes $(X_t)_{\mathbb{Z}}$ that use the softplus function to link the conditional mean $M_t = E(X_t \mid X_{t-1}, \ldots)$ to a linear expression in past observations X_{t-k} and past conditional means M_{t-l}. The softplus function was proposed by Dugas et al. (2000), and is defined as $s(x) = \ln(1 + \exp x)$, for all $x \in \mathbb{R}$. It has been used in a regression context by Zhang and Zhou (2017), Zhao et al. (2018), and Wiemann and Kneib (2019). Its increasing popularity is due to the following properties:

- s is a truly positive, continuous, and differentiable function on whole $\mathbb{R}$;
- except for a region around zero, it closely approximates the rectified linear unit function, $\mathrm{ReLU}(x) = \max\{0, x\}$.

In contrast to the ReLU function, the softplus function $s(x)$ avoids the drawback of not being differentiable in zero, while being approximately linear for $x > 0$. These properties are illustrated in Figure 1 (a), where $s(x)$ is compared to $\mathrm{ReLU}(x)$, as well as to the common response functions (inverse link functions) $\mathrm{logit}^{-1}(x) = \big(1 + \exp(-x)\big)^{-1}$ and $\exp(x)$. Note that $s'(x) = \mathrm{logit}^{-1}(x)$.

The softplus function can be generalized by introducing an additional adjustment parameter $c > 0$, defining $s_c(x) = c \ln\big(1 + \exp(x/c)\big)$ (see Mei and Eisner (2017)), which controls the deviation between $s_c(x)$ and $\mathrm{ReLU}(x)$. We have $s_1(x) = s(x)$ (so $c = 1$ is the default choice) and $\lim_{c\to 0} s_c(x) = \mathrm{ReLU}(x)$, as illustrated by Figure 1 (b). Furthermore, it holds that

$$s_c'(x) = \frac{\exp(x/c)}{1 + \exp(x/c)}, \qquad s_c''(x) = \frac{1}{c}\,\frac{\exp(x/c)}{\big(1 + \exp(x/c)\big)^2}. \tag{1.1}$$

In Section 2, we briefly survey the INGARCH models with their linear conditional mean. In Section 3, we propose a new type of INGARCH model that

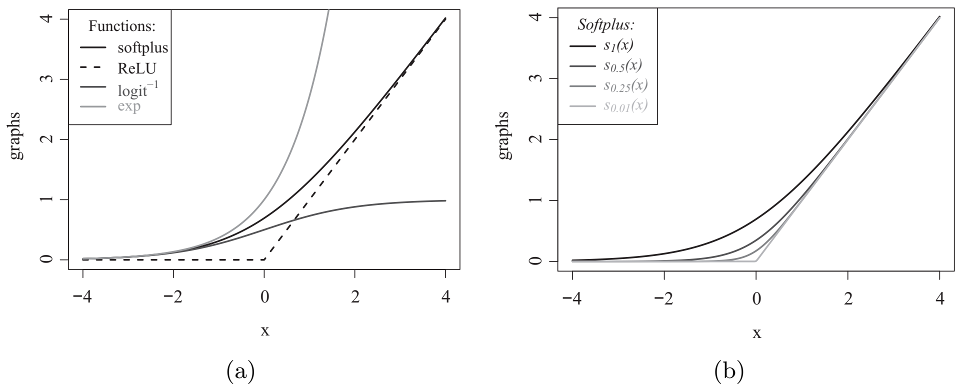

Figure 1. Plots of response functions (against x): softplus $s(x)$ vs. ReLU(x), logit$^{-1}(x)$, and $\exp(x)$ in (a); different softplus functions $s_c(x)$ in (b).

uses the softplus function as a response function. The stochastic properties are derived, and it is shown that the softplus INGARCH model exhibits an approximately linear structure. In Section 4, we derive the asymptotics of the maximum likelihood estimators for the proposed model's parameters, and we analyze their finite-sample performance using simulations. Section 5 demonstrates the usefulness of the novel softplus INGARCH model by applying it to three real-data examples. Finally, Section 6 concludes the paper and outlines potential topics for future research.

2. INGARCH Models

The INGARCH(p, q) model with $p \geq 1$ and $q \geq 0$ requires the conditional mean $M_t = E(X_t | X_{t-1}, \ldots)$ to be a linear expression in the last p observations and the last q conditional means ("feedback terms"); that is,

$$M_t = a_0 + \sum_{i=1}^{p} a_i\, X_{t-i} + \sum_{j=1}^{q} b_j\, M_{t-j}. \tag{2.1}$$

Because the mean of a count random variable is a positive real number, the constraints $a_0 > 0$ and $a_1, \ldots, a_p, b_1, \ldots, b_q \geq 0$ have to hold. The INGARCH model is fully specified once the type of the conditional distribution of X_t, given $X_{t-1}, \ldots$ has been fixed.

The default choice is a conditional Poisson distribution; that is, X_t, conditioned on $X_{t-1}, \ldots$, is Poisson distributed according to Poi(M_t). This model has been discussed by several authors, including Ferland, Latour and Oraichi

(2006), Fokianos, Rahbek and Tjøstheim (2009) and Weiß (2009). Provided that $a_\bullet + b_\bullet := \sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1$, the (Poisson) INGARCH process exists, and is strictly stationary with finite first- and second-order moments (Ferland, Latour and Oraichi (2006)). For $p = q = 1$, all moments exist (Ferland, Latour and Oraichi (2006)), and the mixing properties have been established by Neumann (2011). Because of the linear conditional mean, the unconditional mean is equal to

$$\mu = \frac{a_0}{1 - \sum_{i=1}^{p} a_i - \sum_{j=1}^{q} b_j}, \tag{2.2}$$

and the variance and autocovariances can be computed by solving a set of Yule–Walker equations (Weiß (2009)):

$$\begin{aligned}
\gamma(0) &= \mu + \gamma_M(0), \quad \gamma_M(0) = \sum_{i=1}^{p} a_i \gamma(i) + \sum_{j=1}^{q} b_j \gamma_M(j), \\
\gamma(k) &= \sum_{i=1}^{p} a_i \gamma(|k-i|) + \sum_{j=1}^{\min\{k-1,q\}} b_j \gamma(k-j) + \sum_{j=k}^{q} b_j \gamma_M(j-k), \qquad (2.3) \\
\gamma_M(k) &= \sum_{i=1}^{\min\{k,p\}} a_i \gamma_M(|k-i|) + \sum_{i=k+1}^{p} a_i \gamma(i-k) + \sum_{j=1}^{q} b_j \gamma_M(|k-j|),
\end{aligned}$$

for $k \geq 1$, where $\gamma(h) := Cov(X_t, X_{t-h})$ and $\gamma_M(h) := Cov(M_t, M_{t-h})$. Despite the conditional Poisson distribution being equidispersed (variance equals the mean), the unconditional distribution exhibits overdispersion; that is, the dispersion ratio $\sigma^2/\mu > 1$. In the purely autoregressive case of an INARCH(p) model (i. e., if $q = 0$), the Yule–Walker equations (2.3) imply that the ACF $\rho(h) := Corr(X_t, X_{t-h})$ satisfies

$$\rho(k) = \sum_{i=1}^{p} a_i \, \rho(|k-i|). \tag{2.4}$$

Thus, except for the restriction of nonnegative coefficients a_i, Equation (2.4) is identical to the Yule–Walker equations of an ordinary AR(p) model. Consequently, the model order of an INARCH model can be identified by using the partial ACF (PACF) $\rho_{\mathrm{p}}(h)$. Note that the name "INGARCH" for the models defined by (2.1) is a bit misleading; see also the discussion on p. 74 in Weiß (2018). In contrast to the ordinary GARCH models, the INGARCH models are conditionally linear, which also leads to the Yule–Walker type equations (2.3) and (2.4) for the ACF.

Example 1. For the special case of an INGARCH$(1,1)$ model, the mean is given by $\mu = a_0/(1-a_1-b_1)$; see (2.2). Furthermore, (2.3) implies the variance $\sigma^2 = \gamma(0)$ is equal to

$$\sigma^2 = \frac{1-(a_1+b_1)^2+a_1^2}{1-(a_1+b_1)^2}\cdot\mu,$$

and the ACF $\rho(h)$ is given by

$$\rho(k) = (a_1+b_1)^{k-1}\,\frac{a_1\left(1-b_1(a_1+b_1)\right)}{1-(a_1+b_1)^2+a_1^2} \quad \text{for } k\geq 1;$$

see Weiß (2009). If $b_1 = 0$ (thus excluding the feedback term M_{t-1} from the model), we have the INARCH(1) model, where $\mu = a_0/(1-a_1)$, $\sigma^2 = \mu/(1-a_1^2)$, and $\rho(k) = a_1^k$. Thus, $\rho_{\mathrm{p}}(k) = 0$, for $k > 1$.

In addition to the basic Poisson INGARCH model, several extensions have been developed in the literature, where another type of conditional distribution is used for X_t given $X_{t-1},\ldots$ (see Weiß (2018)). The considered distributions not only have a mean parameter (used to connect to M_t), but also have parameters that allow us to control, for instance, the extent of overdispersion or zero inflation. For example, Zhu (2010) and Xu et al. (2012) defined two types of negative-binomial (NB) INGARCH models, and Zhu (2012a) defined a generalized-Poisson INGARCH model, all of which fall within the compound-Poisson INGARCH family proposed by Gonçalves, Mendes-Lopes and Silva (2015). Other examples are the zero-inflated Poisson INGARCH model developed by Zhu (2012b), the COM-Poisson INGARCH model of Zhu (2012c), the INGARCH model based on the one-parameter exponential family by Davis and Liu (2016), and the mixed Poisson INGARCH model of Silva and Barreto-Souza (2019).

Example 2. As an illustration of possible extensions, consider the NB-INGARCH model proposed by Zhu (2010). It is again defined by Equation (2.1) for the conditional mean M_t, but it has the additional parameter $N > 0$ to control the extent of (conditional) overdispersion. More precisely, the conditional distribution of $X_t\big|X_{t-1},\ldots$ is the NB-distribution with parameters N and $\pi_t = 1/(1+M_t/N)$, where the limit $N\to\infty$ leads to the Poisson INGARCH model. Thus, the aforementioned mean and ACF properties remain as they are, but the conditional variance is inflated by the factor $1+M_t/N$. Considering the special cases of Example 1, Zhu's NB-INGARCH$(1,1)$ model differs only in terms of the unconditional variance, which is given by $\sigma^2 = (1-(a_1+b_1)^2+a_1^2)/(1-(a_1+b_1)^2-a_1^2/N)\cdot \mu\left(1+(\mu/N)\right)$. Thus, Zhu's NB-INARCH(1) model (where $b_1 = 0$) has the variance $\sigma^2 = \mu(1+\mu/N)/(1-a_1^2-a_1^2/N)$; the remaining properties are as in

Example 1.

The Yule–Walker equations in (2.3) are analogous to those of the ordinary ARMA process, and in the purely autoregressive case ($q = 0$), they are actually identical. Nevertheless, the attainable range of ACF values is much more limited than for traditional ARMA processes because of the parameter constraints $a_0 > 0$ and $a_1, \ldots, a_p, b_1, \ldots, b_q \geq 0$. These, in turn, are required to ensure that M_t always takes a positive value. To overcome this limitation, one may use an additional link function, such as the logarithmic link. Such a log-linear INGARCH model is suggested by Fokianos and Tjøstheim (2011), who define $\ln M_t$ as a linear function in $\ln(X_{t-1}+1), \ldots, \ln M_{t-1}, \ldots$, where the linear coefficients $a_0, a_1, \ldots, b_1, \ldots$ can now also take negative values. The corresponding response function is the exponential function (see Figure 1 (a)), and it follows that the conditional mean of such a model is multiplicative:

$$M_t = e^{a_0} \cdot (X_{t-1}+1)^{a_1} \cdots M_{t-1}^{b_1} \cdots .$$

However, in real count time series, one commonly observes an additive structure. Furthermore, analytic expressions for the mean, variance, and ACF of a log-linear INGARCH model are not available, which complicates the application of the model in practice.

3. Softplus INGARCH Models

3.1. Definition and properties

To overcome the limitations of the ordinary INGARCH model, while (approximately) preserving its additive structure, we propose the novel *softplus Poisson INGARCH model*

$$X_t|\mathcal{F}_{t-1} : \mathrm{Poi}(M_t), \tag{3.1}$$

where $\mathcal{F}_t$ is the σ-field generated by $\{(X_t, M_t), (X_{t-1}, M_{t-1}), \ldots\}$. The model relies on using the softplus function $s_c(x) = c \ln\big(1 + \exp(x/c)\big)$ as a response function (see Figure 1), and it defines the conditional mean $M_t = E(X_t|\mathcal{F}_{t-1})$ recursively using the equation

$$M_t = s_c\left(\alpha_0 + \sum_{i=1}^{p} \alpha_i\, X_{t-i} + \sum_{j=1}^{q} \beta_j\, M_{t-j}\right) \quad \text{with } c > 0, \tag{3.2}$$

where $\alpha_0, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \in \mathbb{R}$. The default choice for c is $c = 1$.

Remark 1. Taking the limit $c \to 0$, the softplus function $s_c(x)$ becomes the function $\text{ReLU}(x) = \max\{0, x\}$ (see Section 1), and the softplus equation (3.2) then becomes $M_t = \max\{0, \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j M_{t-j}\}$. This can be understood as a type of dynamic censored regression model (tobit model), as discussed in de Jong and Herrera (2011). However, compared to (3.2), this ReLU INGARCH model has some drawbacks. First, the ReLU response function is not differentiable in zero. Second, the conditional mean M_t might become zero such that we have a degenerate conditional count distribution with all probability mass in the value zero. This may cause, for example, problems in a likelihood computation (if the tth observation x_t is positive, but $M_t = 0$ is computed). The latter might be circumvented by using $\max\{\delta, \cdot\}$, with some $\delta > 0$, for model construction, but the choice of δ is arbitrary, and the function is still not differentiable in whole $\mathbb{R}$. On the other hand, if the observed counts are rather large (as in the data examples discussed in Sections 5.2 and 5.3), then the softplus function is virtually linear, such that the softplus and the ReLU model are not distinguishable in practice. Moreover, for low counts, such as in Section 5.1, the difference between $s_c(x)$ and $\text{ReLU}(x)$ is often negligible.

Let $\boldsymbol{Z}_t = (X_t, \ldots, X_{t-p+1}, M_t, \ldots, M_{t-q+1})$. Then, $\{\boldsymbol{Z}_t\}_t$ is a Markov process. The following theorem discusses the existence and uniqueness of a stationary distribution, as well as the absolute regularity of the softplus INGARCH process.

Theorem 1. *Consider the softplus INGARCH process defined by* (3.1)*. If* $\sum_{i=1}^{p} \max\{0, \alpha_i\} + \sum_{j=1}^{q} \max\{0, \beta_j\} < 1$ *and* $\sum_{j=1}^{q} |\beta_j| < 1$*, then (i) the Markov process* $\{\boldsymbol{Z}_t\}_t$ *has a unique stationary distribution; (ii) a stationary version of the process* $\{X_t\}_t$ *is absolutely regular with* β*-mixing coefficients bounded by* $C\rho^{\sqrt{n}}$*, for some constant* $C \in (0, \infty)$ *and some* $\rho \in (0, 1)$*; and (iii) a stationary version of the process* $\{(X_t, M_t)\}_t$ *is ergodic.*

The proofs of all theorems are provided in the Supplementary Material S4.

Note that $\sum_{i=1}^{p} |\alpha_i| + \sum_{j=1}^{q} |\beta_j| < 1$ is a sufficient (but not necessary) condition for that in Theorem 1, that is, $\sum_{i=1}^{p} \max\{0, \alpha_i\} + \sum_{j=1}^{q} \max\{0, \beta_j\} < 1$ and $\sum_{j=1}^{q} |\beta_j| < 1$.

Remark 2. The proof of Theorem 1 relies on the results derived by Doukhan and Neumann (2019). Although these authors focus on the case of a conditional Poisson distribution, like we do in Equation (3.1), they point out that the involved stability properties also hold for mixed Poisson and compound Poisson distributions (Doukhan and Neumann (2019, p.96)). This opens the opportunity to define non-Poisson extensions of the softplus INGARCH model, analogous to the extensions of the ordinary INGARCH model discussed in Section 2. For example, in Example 2, one may define a softplus NB-INGARCH model in the spirit of Zhu (2010) as $X_t|\mathcal{F}_{t-1} : \text{NB}\big(N, 1/(1+M_t/N)\big)$, with the conditional mean M_t still satisfying the softplus equation (3.2). A detailed analysis of such extensions is planned for future research, but further illustration is presented in a data example in Section 5.3.

Henceforth, we focus on the special case $p = q = 1$, for simplicity. The following theorem states that all moments of model (3.1) are finite, which is analogous to the ordinary INGARCH model in Ferland, Latour and Oraichi (2006), and is crucial in deriving the large-sample properties.

Theorem 2. *Consider the softplus INGARCH process defined by* (3.1) *with* $p = q = 1$. *Then, the moments are all finite if* $|\alpha_1| + |\beta_1| < 1$.

The result in Theorem 2 can be extended to the case $p > 1$ and $q = 0$ using arguments similar to those in Zhu and Wang (2011) and Doukhan, Fokianos and Tjøstheim (2012).

3.2. Approximate moment calculation

While Theorem 2 ensures the existence of moments, it is not possible to find exact closed-form formulae for them. However, because the softplus function closely approximates the piecewise linear ReLu function, we can use the linear INGARCH model's moment formulae (2.2) and (2.3) as approximations of the softplus INGARCH's true moment properties. More precisely, we derive the formulae for the mean, variance, and ACF from the linear INGARCH's equations (2.2) and (2.3), and substitute the parameters a_i and b_j with the softplus INGARCH's parameters α_i and β_j (including if some of them are negative). Certainly, the quality of such an approximate moment calculation is not clear in advance. Therefore, we conducted a numerical study with diverse model parametrizations; see Tables S1–S3 in the Supplementary Material S1, where we also considered the boundary case $c \to 0$, that is, the ReLU INGARCH model discussed in Remark 1. We computed the true moment values (labeled as "sp" in Tables S1–S3 in the Supplementary Material S1) and the approximate "linear" moment values from (2.2) and (2.3) (labeled as "lin"). More precisely, because we focus on INARCH(1) and INGARCH(1, 1) processes, we used the closed-form formulae provided by Example 1 for the approximate moments. For the true softplus (and ReLU) INGARCH's moments, analytic formulae are not available. Therefore, these values were approximated by computing the respective sample moments of a "very long" simulated time series (we used length 10^6). The obtained results are summarized in Tables S1–S3 in the Supplementary Material

S1.

Table S1 considers the case of a softplus INARCH(1) model. Comparing the true moment properties ("sp") with the linearly approximated ones ("lin") for the default choice $c = 1$, we generally observe a rather good agreement, including when α_1 is negative. Thus, the softplus INARCH(1) model behaves very similarly to a truly linear model. There are only a few exceptions that require some further discussion. For model #13, the true and the approximate mean deviate notably from each other. This can be explained by having a very small intercept value α_0 in the region where the softplus function deviates from linearity; recall Figure 1 (b). Therefore, in this case, rather than use the default choice $c = 1$, we recommend using a somewhat smaller value. Table S1 also shows the results for $c = 0.5$, $c = 0.25$, and $c \to 0$. For model #13, we see a clear improvement of the approximate linearity with decreasing c. The same phenomenon, but in a much milder form, occurs for model #7. Notable deviations are also observed for model #16 (mainly in $\rho_{\mathrm{p}}(1)$ and σ^2/μ), where we have a low mean and a strong degree of negative autocorrelation. Improvement is again achieved by reducing the value of c. However, even in the boundary case $c \to 0$ (ReLU INARCH(1) model), there are still some deviations. This is because the ReLU function is not strictly linear (as assumed by the "lin" calculations), but only piecewise linear. Thus, for a low mean and a strong negative autocorrelation, it is impossible to perfectly mimic linearity.

Tables S2–S3 in the Supplementary Material S1 refer to softplus INGARCH($1, 1$) models. Because such models are mainly applied to counts with a very slowly decaying ACF, we chose the model parameters such that $|\alpha_1|+|\beta_1|$ is large, namely $|\alpha_1|+|\beta_1| = 0.70$ (strong dependence; see Table S2) and $|\alpha_1|+|\beta_1| = 0.95$ (extreme dependence; see Table S3). Nevertheless, we still observe rather good agreement in most cases. One exception is model #13 with $c = 1$ in Tables S2–S3 (and clearly mitigated in models #1 and #16), where we again have a very low intercept value α_0, such that c should be chosen to be less than one, say, $c = 0.25$. Furthermore, models #22 and #23 with $c = 1$ (and clearly mitigated in model #10 in Table S3) have to be mentioned, with deviations mainly in the ACF and the dispersion ratio. Here, both dependence parameters are negative (thus, strong or even extreme negative dependence), and we observe a clear improvement with decreasing c. In particular, there is hardly any difference between the cases $c = 0.25$ and $c \to 0$; that is, the softplus function with $c = 0.25$ is sufficiently close to the ReLU function. However, at least for models #22 and #23 in Tables S2–S3, we never get perfect agreement with the linear approximations. The reason is the same as for the INARCH(1) model #16 discussed before: neither the softplus nor the ReLU function are strictly linear, which is problematic for a low mean and strong negative dependence.

In summary, provided that the marginal mean (especially the intercept α_0) is not too small and that the extent of negative dependence is not extreme (in these cases, the parameter c should be chosen as less than one), the softplus INGARCH model's mean, variance, and (P)ACF are well approximated by formulae (2.2) and (2.3). However, in contrast to the ordinary INGARCH model, negative ACF values are possible, and these values are most often well approximated by (2.2) and (2.3). Because of this approximately linear behavior of the softplus INGARCH model, its model parameters are easier to interpret than those of a log-linear INGARCH model. In addition, the approximate linearity can be used to compute approximate moment estimates for α_i and β_j, which, in turn, can be used as starting values for a numerical computation of the maximum likelihood (ML) estimates.

4. ML Estimation

In this section, we discuss the ML estimator (MLE) for the softplus INGARCH$(1,1)$ model parameters. For the model's identification, c in the softplus function should be specified before estimating the parameter (default choice $c = 1$).

4.1. Asymptotic properties

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^\top = (\alpha_0, \alpha_1, \beta_1)^\top$ be the parameter of interest. Its parameter space is $\boldsymbol{\Theta}$ and its true value is $\boldsymbol{\theta}^0$. For $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, define the stationary and ergodic process $M_t = M_t(\boldsymbol{\theta}) = s_c(\lambda_t)$, where $\lambda_t = \lambda_t(\boldsymbol{\theta}) = \alpha_0 + \alpha_1 X_{t-1} + \beta_1\, M_{t-1}(\boldsymbol{\theta})$. Then, the log-likelihood function is given by the following, up to a constant:

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} \Big(X_t\, \ln M_t(\boldsymbol{\theta}) - M_t(\boldsymbol{\theta}) \Big). \tag{4.1}$$

To compute $L_n(\boldsymbol{\theta})$ in practice, the initial value $M_0(\boldsymbol{\theta})$ has to be specified; a possible solution is to choose $M_0(\boldsymbol{\theta}) = \alpha_0$.

The score function is defined by

$$S_n(\boldsymbol{\theta}) = \frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \left(\frac{X_t}{M_t(\boldsymbol{\theta})} - 1 \right) \frac{\partial M_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

where the components of $\partial M_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ are given by

$$\begin{aligned}
\frac{\partial M_t(\boldsymbol{\theta})}{\partial \alpha_0} &= \frac{\exp(\lambda_t(\boldsymbol{\theta})/c)}{1+\exp(\lambda_t(\boldsymbol{\theta})/c)}\left(1+\beta_1\frac{\partial M_{t-1}(\boldsymbol{\theta})}{\partial \alpha_0}\right),\\
\frac{\partial M_t(\boldsymbol{\theta})}{\partial \alpha_1} &= \frac{\exp(\lambda_t(\boldsymbol{\theta})/c)}{1+\exp(\lambda_t(\boldsymbol{\theta})/c)}\left(X_{t-1}+\beta_1\frac{\partial M_{t-1}(\boldsymbol{\theta})}{\partial \alpha_1}\right),\\
\frac{\partial M_t(\boldsymbol{\theta})}{\partial \beta_1} &= \frac{\exp(\lambda_t(\boldsymbol{\theta})/c)}{1+\exp(\lambda_t(\boldsymbol{\theta})/c)}\left(M_{t-1}(\boldsymbol{\theta})+\beta_1\frac{\partial M_{t-1}(\boldsymbol{\theta})}{\partial \beta_1}\right).
\end{aligned}$$

The Hessian matrix is obtained by further differentiation of the score equations, that is,

$$\begin{aligned}
H_n(\boldsymbol{\theta}) &= -\sum_{t=1}^{n}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\\
&= \sum_{t=1}^{n}\frac{X_t}{M_t^2(\boldsymbol{\theta})}\frac{\partial M_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial M_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top} - \sum_{t=1}^{n}\left(\frac{X_t}{M_t(\boldsymbol{\theta})}-1\right)\frac{\partial^2 M_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},
\end{aligned}$$

where the expression for $\partial^2 M_t(\boldsymbol{\theta})/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)$ is given in the proof (Supplementary Material S4) of Theorem 3. According to Ferland, Latour and Oraichi (2006) and Ahmad and Francq (2016), we have the information matrix equality $\mathbf{I} = \mathbf{J}$, where

$$\begin{aligned}
\mathbf{I} &= E\left(\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right) = E\left(\frac{1}{M_t(\boldsymbol{\theta})}\frac{\partial M_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial M_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\right), \qquad (4.2)\\
\mathbf{J} &= -E\left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right).
\end{aligned}$$

Define $\widetilde{M}_t$ as a proxy for M_t as $\widetilde{M}_t = \widetilde{M}_t(\boldsymbol{\theta}) = s_c(\widetilde{\lambda}_t)$, for $t \geq 1$, with unknown initial values X_0 and $\widetilde{M}_0$. The initial values can be fixed values, values depending on $\boldsymbol{\theta}$, or values depending on the observations. The MLE is defined as any measurable solution of

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}}\ \widetilde{L}_n(\boldsymbol{\theta}), \quad \widetilde{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n}\tilde{l}_t(\boldsymbol{\theta}), \qquad (4.3)$$

where $\tilde{l}_t(\boldsymbol{\theta}) = X_t \ln \widetilde{M}_t - \widetilde{M}_t$. To show the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, the following assumptions are made.

Assumption 1. $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ *and* $\boldsymbol{\Theta}$ *is compact.*

Assumption 2. M_t *and* $\widetilde{M}_t$ *take values on* $(\underline{\omega}, +\infty)$*, for some* $\underline{\omega} > 0$.

The lower bound $\underline{\omega}$ in Assumption 2 is only needed for technical reasons in the proof (see the Supplementary Material S4) of the following theorem. In practice, we can select it to be very close to zero, for example, $\underline{\omega} = 0.00001$.

Theorem 3. *Consider model* (3.1) *with* $p = q = 1$*, and suppose that Assumptions* 1 *and* 2 *hold. Then, the MLE* $\hat{\boldsymbol{\theta}}_n$ *defined by* (4.3) *is strongly consistent. In addition, if* $\boldsymbol{\theta}^0$ *lies in the interior of* $\boldsymbol{\Theta}$*, then as* $n \to \infty$,

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0\right) \xrightarrow{d} \mathrm{N}(0, \mathbf{I}^{-1}),$$

where the matrix $\mathbf{I}$ *is given in* (4.2).

Remark 3. The asymptotic covariance matrix $\mathbf{I}^{-1}$ can be consistently estimated using the robust sandwich matrix $\hat{\mathbf{J}}^{-1}\,\hat{\mathbf{I}}\,\hat{\mathbf{J}}^{-1}$, where

$$\begin{aligned}
\hat{\mathbf{I}} =& \frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{\widetilde{M}_t(\hat{\boldsymbol{\theta}}_n)} - 1\right)^2 \frac{\partial \widetilde{M}_t(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\,\frac{\partial \widetilde{M}_t(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}^\top},\\
\hat{\mathbf{J}} =& \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\widetilde{M}_t(\hat{\boldsymbol{\theta}}_n)}\,\frac{\partial \widetilde{M}_t(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\,\frac{\partial \widetilde{M}_t(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}^\top}.
\end{aligned}$$

In practice, the ML estimates are computed by numerically maximizing the log-likelihood function (4.1). As recommended in Section 3.2, one may use the approximate moment estimates as the initial values for the numerical optimization routine.

Remark 4. The asymptotic theory for the MLE can be extended to the case $p > 1$ and $q > 1$ by employing the techniques of Cui and Wu (2016).

4.2. Simulation study

To analyze the finite-sample performance of the ML estimation, we conducted a simulation study with diverse model parametrizations and with sample sizes $n = 100, 250, 500$. For each scenario, the number of replications was 10^4. In addition to the actual ML estimates, we computed the approximate standard errors (s. e.) from the inverse Hessian of the maximized log-likelihood function. As a result, we checked both the performance of the estimates and that of the approximate s. e. The full simulation results are presented in Tables S4–S8 in the Supplementary Material S2.

Table S4 presents the results for a softplus INARCH(1) model, where the parameter values are chosen such that the marginal mean is approximately equal to 2, 5, or 15 (see also the data examples in Sections 5.1 and 5.3). Furthermore, we are concerned with a medium level of autocorrelation (either positive or negative). Because the softplus INARCH(1) process is just a Markov chain (i. e., its memory is only of length one), we also included the very low sample size $n = 50$.

Table S4 shows that both the bias and the s. e. quickly decrease if the sample size n increases. For negative α_1, the estimates are nearly unbiased already for $n = 50$, and, in general, exhibit slightly less bias and s. e. than for positive α_1. In the latter case, the bias appears negligible if the sample size n becomes larger than 100. Table S4 also considers the effect of the softplus parameter c, where the default choice $c = 1$ is compared to $c = 0.5$. Except that the s. e. values are slightly smaller for $c = 0.5$ and $\mu = 2, 5$, the effect of c is, in general, rather small (for $\mu = 15$, we have very large counts such that both $s_1(x)$ and $s_{0.5}(x)$ are virtually identical). Furthermore, in all cases, the mean of the approximate s. e. is very close to the simulated value of the s. e. Thus, the approximate s. e. performs quite well in practice.

Tables S5–S6 in the Supplementary Material S2 refer to the softplus INGARCH$(1, 1)$ model with either $|\alpha_1| + |\beta_1| = 0.70$ (strong dependence; see Table S5) or $|\alpha_1| + |\beta_1| = 0.95$ (extreme dependence; see Table S6). Except for the low-mean case $\mu = 2$, the parameter c again has little effect on the estimation performance. Compared to the INARCH(1) case, we now have a more complex dependence structure (controlled by the parameter β_1). Because of this (and because of the additional parameter to be estimated), the bias and s. e. are, in general, much larger this time. Nevertheless, they clearly improve with increasing n, as before. It can be seen that the estimation performance is worse if $|\beta_1|$ is large than if $|\alpha_1|$ is large. In this case, the current observation is determined mainly by the unobservable past mean (feedback term), whereas it is connected to the last observation if $|\alpha_1|$ is large. Actually, the worst case in Tables S5–S6 is model #1, where both α_1 and β_1 are positive and β_1 is largest. The estimation performance is particularly bad for model #1 in Table S6, where $\alpha_1 + \beta_1 = 0.95$ is close to one ("unit-root problem"). Comparing the simulated s. e. with the mean of the approximate s. e., there are large discrepancies for models #1 and #4 with $n = 100$, whereas these values approach each other for $n \geq 250$. Therefore, we recommend collecting $n \geq 250$ data values in the case of strongly dependent data to ensure a reasonable estimation performance.

Finally, we analyzed the effect of a misspecified c on the estimation performance. The results for some softplus INARCH(1) and INGARCH$(1, 1)$ scenarios are summarized in Tables S7 and S8, respectively. There, a model with $c = 1$ was fitted to the data, although the true DGP has $c = 2$ (so fitted c to small) or $c = 0.5$ (so fitted c to large). While there is only little effect on the s. e. of the estimators, the effect on the bias is a bit more pronounced, especially for $c = 2$ where the nonlinearity is stronger than that assumed by the model. Thus, to avoid an inappropriate choice for c, it is recommended to accompany any model

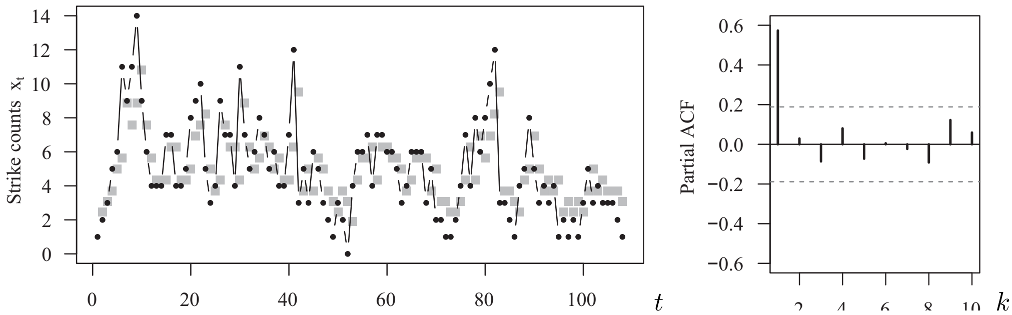

Figure 2. Plot of strike counts $x_1, \ldots, x_{108}$ (the grey dots refer to the conditional means of the fitted softplus-INARCH(1) model) and their sample PACF $\hat{\rho}_{\mathrm{p}}(k)$; see Section 5.1.

fitting with a careful model selection and adequacy checks, as in the subsequent data examples.

5. Real-Data Examples

We now present several data examples to demonstrate the usefulness of the novel softplus INGARCH model.

5.1. Strike count data

In our first example, we use a count time series that has been successfully modeled using an ordinary INGARCH model in the past. More precisely, we analyze strike count data (originally published by the U. S. Bureau of Labor Statistics, `http://www.bls.gov/wsp/`), where Weiß (2010) showed that the ordinary Poisson INARCH(1) model constitutes an excellent fit; see also the discussion in Weiß (2018). The data consist of $n = 108$ monthly counts of "work stoppages" (strikes and lock-outs by $\geq 1{,}000$ workers); see the plot in Figure 2. The data have an AR(1)-like sample (P)ACF with $\hat{\rho}(1) \approx 0.573$. The sample mean is ≈ 4.944, and the dispersion ratio ≈ 1.587 shows a notable degree of overdispersion. Because the INARCH(1) model's estimated parameters in Table 1 are positive, the model is identical to the ReLU model discussed in Remark 1.

From our analysis of Table S1 in the Supplementary Material S1, recall Section 3.2, we know that a softplus INARCH(1) model with a mean close to five and a lag-1 ACF close to 0.5 behaves very similarly to a truly linear model. Hence, this model appears to be a reasonable alternative to the ordinary INARCH(1) model. Therefore, we also fitted the softplus INARCH(1) model to the data, using the (conditional) ML approach for the parameter estimation. The results are summarized in Table 1. It can be seen that the parameter estimates and the

Table 1. ML estimation for strike count data: estimates and approximate standard errors, maximized log-likelihood.

Model	c	$\hat{\alpha}_0$ or $\hat{a}_0$	s. e.	$\hat{\alpha}_1$ or $\hat{a}_1$	s. e.	max. L
softplus-INARCH(1)	1	1.728	(0.416)	0.650	(0.085)	-230.16
	0.75	1.778	(0.401)	0.642	(0.083)	-230.13
	0.5	1.804	(0.390)	0.638	(0.081)	-230.14
INARCH(1)	—	1.811	(0.386)	0.636	(0.081)	-230.15

corresponding approximate s. e. of the softplus INARCH(1) model are very close to those of the ordinary INARCH(1) model, which is not surprising given that the softplus INARCH(1) model is nearly linear. Although not necessary from a practical point of view, we also experimented with values $c < 1$. It can be seen that the tabulated values approach those of the ordinary INARCH(1) model as $c \to 0$. If we consider the maximized log-likelihood (column "max. L") as the criterion for model selection (because all candidate models in Table 1 have the same number of parameters, the model selection based on "max. L" leads to an identical decision to that when common information criteria are used), there is a tiny preference for the softplus INARCH(1) model with $c = 0.75$. However, all models in Table 1 perform nearly equally well. If computing the standardized Pearson residuals to check the model adequacy (Weiß (2018, Sec. 2.4)), then we always obtain a mean value of about 0.002 (very close to the target value zero), a variance of about 0.986 (close to the target value one), and no ACF values that are significantly different from zero (the conditional means, as used to compute the Pearson residuals, are plotted in Figure 2 as grey dots). Furthermore, the acceptance envelope for the sample PACF in the Supplementary Material S3.1 confirms the model adequacy. In conclusion, the softplus INARCH(1) model can be used as a substitute for the (truly linear) ordinary INARCH(1) model without concern. However, in contrast to the ordinary INARCH(1) model, the softplus INARCH(1) model also allows for negative parameter values; that is, it has much more comprehensive modeling abilities than the ordinary INARCH(1) model. This advantage is crucial in the following examples.

5.2. Chemical process data

In the previous example, the ACF consists of only positive values, such that the ordinary INARCH(1) model could be applied to these data. Our second data example is chosen such that negative ACF values are also observed. We consider the series of $n = 70$ consecutive yields from a batch chemical process, printed as "Time series 4.1" in Appendix A.3 of O'Donovan (1983); see the plot in Figure 3.

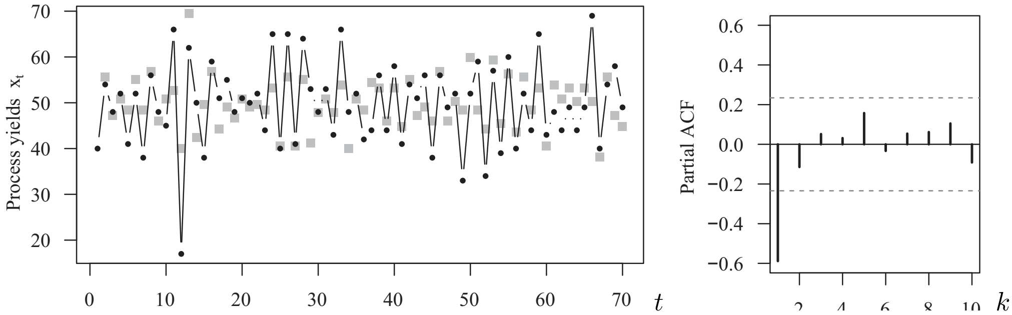

Figure 3. Plot of process yields $x_1, \ldots, x_{70}$ (the grey dots refer to the conditional means of the fitted softplus-INARCH(1) model) and their sample PACF $\hat{\rho}_{\mathrm{p}}(k)$; see Section 5.2.

Table 2. ML estimation for process yield data: estimates and approximate standard errors, maximized log-likelihood.

Model	c	$\hat{\alpha}_0$ or $\hat{a}_0$	s. e.	$\hat{\alpha}_1$ or $\hat{a}_1$	s. e.	max. L
softplus-INARCH(1)	1	79.783	(4.820)	-0.603	(0.094)	-238.0
log-INARCH(1)	—	5.681	(0.296)	-0.455	(0.076)	-242.0

For such batch data, negative ACF values are commonly observed, because a high-yielding batch often causes residues that reduce the yield of the subsequent batch (and vice versa). Given the negative ACF values, ordinary INGARCH models cannot be used for the data. Because the sample PACF in Figure 3 is significant only at lag one, we again decide on an AR(1)-like autocorrelation structure and, thus, consider the softplus INARCH(1) model (for completeness, we also fitted the log-linear INARCH(1) model). The sample mean, taking the value ≈ 49.69, is very large. Therefore, there is no reason to deviate from the default choice $c = 1$, because the softplus function $s_1(x)$ is virtually linear for such large values of x (recall Figure 1). In fact, the obtained ML estimates in Table 2 (s. e. in parentheses) are identical to those when using the ReLU INARCH(1) model instead (recall Remark 1).

The softplus INARCH(1) estimate $\hat{\alpha}_1 \approx -0.603$ is very close to the actually observed ACF value $\hat{\rho}(1) \approx -0.588$, so the fitted model mimics the autocorrelation structure rather well (see also the acceptance envelope for the PACF in the Supplementary Material S3.1). Furthermore, the fitted model's dispersion ratio is ≈ 1.571 (recall Example 1), and is thus close to the sample value of ≈ 1.706. Therefore, the fitted softplus INARCH(1) model deals with these overdispersed data well. The model adequacy is confirmed by an analysis of the Pearson residuals, leading to the mean ≈ 0.000 being close to zero, the variance ≈ 1.142

Table 3. ML estimation for crash count data: estimates and approximate standard errors, maximized log-likelihood.

Model	c	$\hat{\alpha}_0$ or $\hat{a}_0$	s. e.	$\hat{\alpha}_1$ or $\hat{a}_1$	s. e.	$\hat{\beta}_1$ or $\hat{b}_1$	s. e.	$\hat{N}$	s. e.	max. L
softplus NB-INGARCH(1, 1)	1	15.411	(2.311)	0.253	(0.053)	-0.455	(0.163)	19.935	(3.895)	-1,065.6
log-linear NB-INGARCH(1, 1)	–	2.739	(0.521)	0.263	(0.052)	-0.340	(0.200)	20.282	(4.005)	-1,064.4

being close to one, and no significant ACF values. Finally, note that the fitted log-linear INARCH(1) model clearly performs worse in terms of the maximized log-likelihood. In addition, its Pearson residuals show stronger deviations from the respective target values (mean ≈ -0.001, variance ≈ 1.242). In particular, its parameter values in Table 2 do not have as simple an interpretation as those of the softplus INARCH(1) model do, because the latter directly express essential moment properties.

5.3. Crash count data

Here, we examine data on daily crash counts on the major roads of Utrecht in 2001 (length $n = 365$), as discussed in Zhu and Wang (2015). The data have a sample mean ≈ 12.82 and a dispersion ratio ≈ 1.712; their plot is provided in Figure 4. The first two data examples have an AR(1)-like sample (P)ACF, such that INARCH(1)-type models are sufficient to describe the data. In contrast, this data example exhibits a more complex autocorrelation structure. Therefore, despite not being consistent with Figures 2 and 3, this time, we show the ordinary sample ACF in Figure 4. The plotted ACF values are only moderate, but are slowly decaying. For this reason, Zhu and Wang (2015) fitted a Poisson INGARCH$(1, 1)$ model to the data, including the additional feedback term M_{t-1} to increase the process memory. However, the estimate of the parameter b_1 fell on the lower bound of the bounding box, which was chosen as 0.001. Therefore, Zhu and Wang (2015) also tried a log-linear Poisson INGARCH$(1, 1)$ model (i. e., with a log-link instead of a linear one), which allows a negative estimate for b_1 at the price of a non-linear conditional mean. However, the failure of the ordinary Poisson INGARCH$(1, 1)$ model does not necessarily imply that a linear model is not appropriate for the data; it would just have been necessary to allow some model parameters to potentially become negative. To solve this dilemma, we now fit a softplus INGARCH$(1, 1)$ model to the data to obtain a negative estimate for β_1.

We started by fitting the different types of Poisson INGARCH$(1, 1)$ model

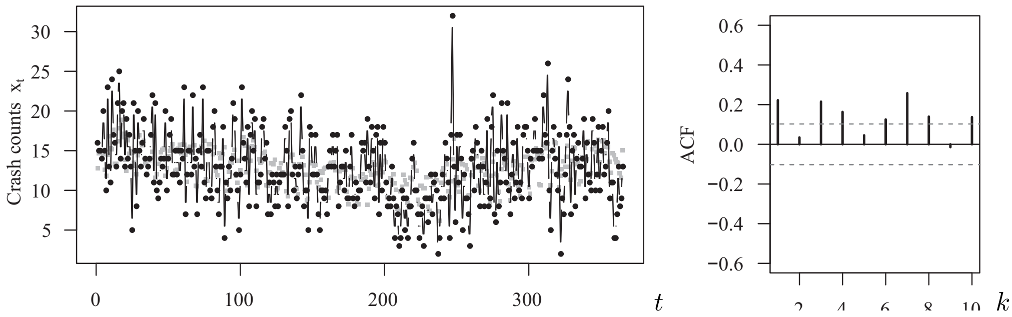

Figure 4. Plot of crash counts $x_1, \ldots, x_{365}$ (the grey dots refer to the conditional means of the fitted softplus NB-INARCH(1) model) and their sample ACF $\hat{\rho}(k)$; see Section 5.3.

to the data; see the Supplementary Material S3.2 for detailed results. However, neither the log-linear nor the softplus Poisson INGARCH$(1, 1)$ models turned out to be appropriate for the data, because neither captures the large extent of overdispersion (dispersion ratio ≈ 1.712). Therefore, we repeated the same analysis, but using a conditional NB-distribution instead of the Poisson one; see Remark 2 for details. The results of the ML estimation are summarized in Table 3. It can be seen that the softplus NB-INGARCH$(1, 1)$ model indeed has a significantly negative estimate for β_1 (whereas the estimate $\hat{b}_1$ of the log-linear model is not significant). Because of the large mean, further decreasing the value of c below the default choice $c = 1$ has no effect on the estimates. As a result, the softplus model cannot be distinguished from a ReLU model, according to Remark 1, for these data. An analysis of the standardized Pearson residuals leads to the means ≈ 0.000 (both models) and variances ≈ 0.990 (softplus) and ≈ 0.992 (log-linear), all being close to the respective target values of zero and one. Thus, the conditional dispersion structure is well captured by both models. However, the residuals' ACF shows several significant values for both models; that is, neither model is able to explain the serial dependence structure. Thus, it is not appropriate to model the slowly decaying autocorrelation structure by including the feedback term M_{t-1} in the INGARCH-type models.

Therefore, we next followed the strategy outlined in Zhu, Shi and Liu (2015), and included appropriate covariates in the model instead of using a feedback term. To explain the actual dependence structure, we used the log-linear and softplus INARCH(1) models, with (standardized) daily temperature and an indicator for weekdays as the covariates. The corresponding linear coefficients are denoted as γ_1 and γ_2, respectively. First, we fitted the model using a conditional Poisson distribution (see the Supplementary Material S3.2), but it turned out that these

Table 4. ML estimation for crash count data, with daily temperature and weekdays indicator as covariates: estimates and approximate standard errors, maximized log-likelihood.

Model	c	$\hat{\alpha}_0$ or $\hat{a}_0$	s. e.	$\hat{\alpha}_1$ or $\hat{a}_1$	s. e.	$\hat{\gamma}_1$	s. e.	$\hat{\gamma}_2$	s. e.	$\hat{N}$	s. e.	max. L
softplus NB-INARCH(1)	1	8.112	(0.715)	0.182	(0.049)	-1.218	(0.227)	3.343	(0.448)	34.705	(9.745)	-1,035.1
log-linear NB-INARCH(1)	–	1.904	(0.123)	0.170	(0.046)	-0.092	(0.018)	0.278	(0.040)	34.503	(9.663)	-1,036.1

models cannot fully capture the observed dispersion structure. Thus, we again used a conditional NB-distribution in the manner proposed by Zhu (2010); recall Remark 2. The results are summarized in Table 4. Compared to Table 3, the maximized log-likelihood values have improved considerably. The respective Pearson residuals have means ≈ 0.000 (both models) and variances ≈ 1.006 (softplus) and ≈ 1.007 (log-linear), all being close to the respective target values of zero and one. Furthermore, their ACF has a slightly significant value only at lag 3 (≈ 0.161 for softplus, ≈ 0.160 for log-linear, where the approximate s. e. $n^{-1/2} \approx 0.052$); that is, both models do rather well in explaining the actual serial dependence structure. Therefore, we conclude that both types of NB-INARCH(1) regression models are adequate for the crash count data, but the softplus model has an advantage in terms of the maximized log-likelihood value (see Table 4).

6. Conclusion

We have proposed a novel INGARCH model based on the softplus function, which has a flexible range of ACF values. The new model exhibits an approximately linear structure, which makes its model parameters easier to interpret than those of a log-linear INGARCH model. The MLE is used to estimate the model parameters, and its large-sample properties are derived. Extensive simulation studies and three real-data examples show the usefulness of the proposed model.

Some suggestions for future research are given as follows. First, following on from the discussion in Remark 2 and Section 5.3, our novel softplus INGARCH model should be extended to non-Poisson conditional distributions, analogously to the INGARCH extensions by Zhu (2010), Xu et al. (2012), Gonçalves, Mendes-Lopes and Silva (2015), and others, addressed in Section 2. From Doukhan and Neumann (2019), we know that the stability properties in Theorem 1 also hold for mixed Poisson and compound Poisson distributions, which removes theoretical barriers to establishing the consistency and asymptotic normality of the estimators for the unknown parameters. Second, diagnostic tests for uncovering deviations from a conditional Poisson distribution should be developed, similarly to, for example, the tests in Weiß, Gonçalves and Mendes Lopes (2017). Third, the log-linear INGARCH model has been applied in many fields. For example, Chen et al. (2018) investigated the causal relationship between human influenza cases and air pollution, and Hall, Raskutti and Willett (2019) used it to determine the impact of a network structure on a time series evolution. This suggests that related problems should also be studied using the novel softplus INGARCH model.

Supplementary Material

The online Supplementary Material provides detailed results for the approximate moment calculations of Section 3.2 (S1), detailed results for the simulation study of Section 4.2 (S2), additional results for the real-data examples discussed in Section 5 (S3), and proofs for Theorems 1–3 (S4).

Acknowledgments

The authors thank the associate editor and the two referees for their useful comments on an earlier draft of this article. Zhu's work was supported by the National Natural Science Foundation of China (Nos. 11871027, 11731015) and Natural Science Foundation of Jilin Province (No. 20210101143JC).

References

Ahmad, A. and Francq, C. (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis* **37**, 291–314.

Chen, C. W. S., Hsieh, Y., Su, H. and Wu, J. J. (2018). Causality test of ambient fine particles and human influenza in Taiwan: Age group-specific disparity and geographic heterogeneity. *Environment International* **111**, 354–361.

Cui, Y. and Wu, R. (2016). On conditional maximum likelihood estimation for INGARCH(p,q) models. *Statistics and Probability Letters* **118**, 1–7.

Davis, R. A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statistica Sinica* **26**, 1673–1707.

de Jong, R. and Herrera, A. M. (2011). Dynamic censored regression and the open market desk reaction function. *Journal of Business & Economic Statistics* **29**, 228–237.

Doukhan, P., Fokianos, K. and Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statistics and Probability Letters* **82**, 942–948. Correction, ibid, **83**, 1926–1927.

Doukhan, P. and Neumann, M. H. (2019). Absolute regularity of semi-contractive GARCH-type processes. *Journal of Applied Probability* **56**, 91–115.

Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C. and Garcia, R. (2000). Incorporating second-order functional knowledge for better option pricing. In *Proceedings of the 13th International*

Conference on Neural Information Processing Systems (NIPS'00) (Edited by Leen et al.), 451–457. MIT Press, Cambridge.

Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH processes. *Journal of Time Series Analysis* **27**, 923–942.

Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association* **104**, 1430–1439.

Fokianos, K. and Tjøstheim, D. (2011). Log-linear Poisson autoregression. *Journal of Multivariate Analysis* **102**, 563–578.

Gonçalves, E., Mendes-Lopes, N. and Silva, F. (2015). Infinitely divisible distributions in integer-valued GARCH models. *Journal of Time Series Analysis* **36**, 503–527.

Hall, E. C., Raskutti, G. and Willett, R. M. (2019). Learning high-dimensional generalized linear autoregressive models. *IEEE Transactions on Information Theory* **65**, 2401–2422.

Mei, H. and Eisner, J. (2017). The neural Hawkes process: A neurally self-modulating multivariate point process. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* (Edited by von Luxburg et al.), Curran Associates Inc., 6757–6767.

Neumann, M. H. (2011). Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **17**, 1268–1284.

O'Donovan, T. M. (1983) *Short Term Forecasting: An Introduction to the Box–Jenkins Approach.* John Wiley & Sons, Chichester.

Silva, R. B. and Barreto-Souza, W. (2019). Flexible and robust mixed Poisson INGARCH models. *Journal of Time Series Analysis* **40**, 788–814.

Weiß, C. H. (2009). Modelling time series of counts with overdispersion. *Statistical Methods and Applications* **18**, 507–519.

Weiß, C. H. (2010). The INARCH(1) model for overdispersed time series of counts. *Communications in Statistics–Simulation and Computation* **39**, 1269–1291.

Weiß, C. H. (2018). *An Introduction to Discrete-valued Time Series.* John Wiley & Sons, Chichester.

Weiß, C. H. (2021). Stationary count time series models. *WIREs Computational Statistics* **13**, e1502.

Weiß, C. H., Gonçalves, E. and Mendes Lopes, N. (2017). Testing the compounding structure of the CP-INARCH model. *Metrika* **80**, 571–603.

Wiemann, P. and Kneib, T. (2019). Using the softplus function to construct alternative link functions in generalized linear models and beyond. Working paper. University of Göttingen.

Xu, H.-Y., Xie, M., Goh, T. N. and Fu, X. (2012). A model for integer-valued time series with conditional overdispersion. *Computational Statistics and Data Analysis* **56**, 4229–4242.

Zhang, Q. and Zhou, M. (2017). Permuted and augmented stick-breaking Bayesian multinomial regression. *Journal of Machine Learning Research* **18**, 7479–7511.

Zhao, H., Liu, F., Li, L. and Luo, C. (2018). A novel softplus linear unit for deep convolutional neural networks. *Applied Intelligence* **48**, 1707–1720.

Zhu, F. (2010). A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis* **32**, 54–67.

Zhu, F. (2012a). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications* **389**, 58–71.

Zhu, F. (2012b). Zero-inflated Poisson and negative binomial integer-valued GARCH models. *Journal of Statistical Planning and Inference* **142**, 826–839.

Zhu, F. (2012c). Modeling time series of counts with COM-Poisson INGARCH models. *Mathematical and Computer Modelling* **56**, 191–203.

Zhu, F., Shi, L. and Liu, S. (2015). Influence diagnostics in log-linear integer-valued GARCH models. *AStA Advances in Statistical Analysis* **99**, 311–335.

Zhu, F. and Wang, D. (2011). Estimation and testing for a Poisson autoregressive model. *Metrika* **73**, 211–230.

Zhu, F. and Wang, D. (2015). Empirical likelihood for linear and log-linear INGARCH models. *Journal of the Korean Statistical Society* **44**, 150–160.

Christian H. Weiß

Department of Mathematics and Statistics, Helmut Schmidt University, 22043 Hamburg, Germany.

E-mail: weissc@hsu-hh.de

Fukang Zhu

School of Mathematics, Jilin University, 2699 Qianjin Street, Changchun 130012, China.

E-mail: zfk8010@163.com

Aisouda Hoshiyar

Department of Mathematics and Statistics, Helmut Schmidt University, 22043 Hamburg, Germany.

E-mail: aisouda.hoshiyar@hsu-hh.de

(Received February 2020; accepted October 2020)