Generalized Linear Autoregressive Models

Caden Hewlett

2025-07-09

Introduction and Motivation

Currently, in our DKR formulation we have our primary DKR term (often denoted as $\tilde{y}_t$), which is given by our familiar equation

$$\tilde{y}_t = \sum_{i=1}^{k} \left(\beta_{0i} + \beta_{1i} \left(\mathbf{z} * \kappa_i\right)[t]\right) \cdot \left(\mathbf{x} * \kappa_i\right)[t] \tag{1}$$

However, we acknowledge that the *true* stream flow y_t has potentially autocorrelated errors

$$y_t = \tilde{y}_t + \varepsilon_t, \text{ where } \varepsilon_t = \sum_{j=1}^{p} \phi_j \varepsilon_{t-j} + u_t$$

Where u_t are the uncorrelated errors (which are Gaussian, Laplace, etc.)

In our model implementation, we estimate the AR order in a rather ad-hoc procedure, allowing our "true" fitted value $\hat{y}_t$ to be written as:

$$\hat{y}_t = \tilde{y}_t + \sum_{j=1}^{\hat{p}} \hat{\phi}_j \big(y_{t-j} - \tilde{y}_{t-j}\big)$$

And thus, from the theory perspective, the uncorrelated error term u_t is given by

$$u_t = (y_t - \hat{y}_t) - \sum_{j=1}^{p} \phi_j \big(y_{t-j} - \tilde{y}_{t-j}\big)$$

This becomes a key factor in the derivation of our log-likelihood equation and consequently our optimization procedure. There have been some notable issues in our present optimization procedure, including the shape of our residuals.

However, we can leverage the assumption of a Gamma-distributed response variable and (relatively new) time series literature to construct a better model and set of likelihood equations, while maintaining our DKR model as the primary predictor of the response variable in a regression-like context.

GARMA: Model Derivation

The derivation of the PDF begins with the exponential family conditional density:

$$f(y_t \mid \underline{\mathbf{H}}_t) = \exp\left\{\frac{y_t\vartheta_t - b(\vartheta_t)}{\varphi} + d(y_t, \varphi)\right\} \tag{1}$$

The above is from the 2003 paper by Benjamin et. al, where $\underline{\mathbf{H}}_t = \{\mathbf{x}_t, \ldots \mathbf{x}_1, y_{t-1}, \ldots y_1, \mu_t, \ldots \mu_1\}$.

In the above, ϑ_t and φ are the canonical and scale parameters, respectively. Both $b(\cdot)$ and $d(\cdot)$ are distribution-specific functions.

The conditional mean and variance of y_t are therefore $\mu_t = \mathbb{E}[y_t \mid \underline{\mathbf{H}}_t] = b'(\vartheta_t)$ and $\text{var}(y_t \mid \underline{\mathbf{H}}_t) = b'(\vartheta_t)$.

Because it is an adjustment of a standard GLM component, μ_t is related to the predictor, η_t by a link function $g(\cdot)$.

The general model for μ_t is given by $g(\mu_t) = \eta_t = \mathbf{x}^\top\boldsymbol{\beta} + \tau_t$, where τ_t is an additional component that allows ARMA terms to be included additively in the predictor.

The general form of τ_t is given by

$$\tau_t = \sum_{j=1}^{p} \phi_j \mathcal{A}(y_{t-j}, \mathbf{x}_{t-j}, \boldsymbol{\beta}) + \sum_{j=1}^{q} \theta_j \mathcal{M}(y_{t-j}, \mu_{t-j}) \tag{2}$$

Where $\mathcal{A}$ and $\mathcal{M}$ are functions representing the AR and MA terms respectively, with parameters $\boldsymbol{\phi}^\top = \langle\phi_1, \ldots, \phi_p\rangle$ and $\boldsymbol{\theta}^\top = \langle\theta_1, \ldots, \theta_q\rangle$,

The moving average error terms, $\mathcal{M}$ are (in this case) residuals on the predictor scale $(g(y_t) - \eta_t)$ but any residual metric (Pearson, deviance, original scale $y_t - \mu_t$, etc.)

However, the above is too general for a practical application. The authors use the predictor scale residuals to define $\mathcal{M}$ and regression differencing to define $\mathcal{A}$, i.e.

$$g(\mu_t) = \eta_t = \mathbf{x}_t^\top\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j\big(g(y_{t-j}) - \mathbf{x}_{t-j}^\top\boldsymbol{\beta}\big) + \sum_{j=1}^{q} \theta_j\big(g(y_{t-j}) - \eta_j\big) \tag{3}$$

In some cases - and likely ours, too, as we have fits near zero - the authors recommend implementing a threshold parameter $y_{t-j}^\star = \max\{y_{t-j}, c\}, c \in (0, 1)$ to prevent zero-values, especially when $g(\cdot)$ is an inverse or logarithmic function.

Some Results

In the paper, the authors prove that the marginal mean of y_t of the GARMA model defined by (1) and (3) with the identity link function $g(\cdot)$ is $\mathbf{x}_t^\top\boldsymbol{\beta}$, provided that $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is invertible (roots beyond the unit circle in $\mathbb{C}$).

$$\mathbb{E}(y_t) = \mathbf{x}_t^\top\boldsymbol{\beta}$$

Other properties (like closed-form variance) also hold in this case.

Gamma-GARMA

For a Gamma-distributed GARMA(p, q) model, we can allow $g(u) = 1/u$ (the canonical link), $g(u) = \log(u)$ (since $y_t > 0$), or the identity link.

Using Section 5 (Illustrative Example in the Paper), our goal is to write Equation (1) in terms of a single parameter (either scale or shape) as well as the conditional mean μ_t.

Let us now suppose that the conditional distribution of the response is Gamma-distributed with shape α and scale β.

The PDF of $y_t \mid \underline{\mathbf{H}}_t$ is then given by

$$f_{Y_t}(y_t \mid \underline{\mathbf{H}}_t) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y_t^{\alpha-1} e^{-y_t/\beta}$$

Directly, $\mathbb{E}\big[y_t \mid \underline{\mathbf{H}}_t\big] = \mu_t = \alpha\beta$. It follows that $\beta = \mu_t/\alpha$. With this in mind, we can reparamaterize the above in terms of the conditional mean μ_t.

$$\begin{aligned} f_{Y_t}(y_t \mid \underline{\mathbf{H}}_t) &= \frac{1}{\Gamma(\alpha)} \cdot \left(\frac{1}{\left(\frac{\mu_t}{\alpha}\right)}\right)^\alpha \cdot y_t^{\alpha-1} \cdot \exp\left(-\frac{y_t}{\left(\frac{\mu_t}{\alpha}\right)}\right) \\ &= \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu_t}\right)^\alpha y_t^{\alpha-1} \exp\left(-\frac{\alpha y_t}{\mu_t}\right) \end{aligned} \tag{4}$$

Where, in the above, $\mu_t = g^{-1}(\eta_t)$, where η_t is defined in Equation 3.

Connection to DKR

Since we have already been applying AR-correction procedures standard to regression models, such as the Cochrane-Orcutt procedure, modelling our streamflow Y_t as a Gamma-distributed variable under a GARMA process would be rather direct; all we need to do is replace the classical regression parameter $\mathbf{x}_t^\top \boldsymbol{\beta}$ with our DKR equation $\tilde{y}_t$

Specifically, we can write

$$\eta_t = g(\mu_t) = \tilde{y}_t + \tau_t$$

Or, in the expanded form, recalling Equation 3 and Equation 1

$$g(\mu_t) = \underbrace{\sum_{i=1}^{k} \left(\beta_{0i} + \beta_{1i}\left(\mathbf{z} * \kappa_i\right)[t]\right) \cdot \left(\mathbf{x} * \kappa_i\right)[t]}_{\text{DKR Component}} + \underbrace{\sum_{j=1}^{p} \phi_j\big(g(y_{t-j}) - \tilde{y}_{t-j}\big) + \sum_{j=1}^{q} \theta_j\big(g(y_{t-j}) - \eta_{t-j}\big)}_{\text{ARMA Component}}$$

Where μ_t is the Scale parameter in the conditional distribution of y_t with density derived in the previous section as:

$$f_{Y_t}(y_t \mid \underline{\mathbf{H}}_t) = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu_t}\right)^\alpha y_t^{\alpha-1} \exp\left(-\frac{\alpha y_t}{\mu_t}\right)$$

Maximum Likelihood Estimation

The GARMA model-fitting procedure used by the authors performs maximum likelihood estimation using iteratively reweighted least squares (IRLS), which is based on the standard GLM procedure.

This involves using the Fisher scoring algorithm. This involves a step involves taking the first derivative of $\mu_t = g^{-1}(\eta_t) = g^{-1}(\mathbf{x}_t^\top \boldsymbol{\beta} + \tau_t)$. This is not normally an issue, since the link functions $g(\cdot)$ are smooth, $\mathbf{x}_t^\top \boldsymbol{\beta}$

is differentiable and τ_t is a fixed constant conditional on $\underline{\mathbf{H}}_t$. However, in our case we are replacing $\mathbf{x}_t^\top \boldsymbol{\beta}$ with $f_{\text{DKR}}(\mathbf{x}_t, \mathbf{z}_t)$ which we know is not differentiable.

Our parameters in this case are fairly similar to our current parameters. We have hyper-parameters k, p and q for the number of kernels, AR order and MA order, respectively.

For a fixed hyper-parameter set, we let $\mathbf{P} = [\beta_{0,1:k}, \beta_{1,1:k}, \sigma_{1:k}, \delta_{1:k}]_{k\times 4}$ be the $k \times 4$ matrix containing the DKR model parameters. By constrcution and our current convention, each row of $\mathbf{P}$ contains the model parameters for a single kernel of a k-kernel model. We further let $\boldsymbol{\phi} = \langle \phi_1, \dots \phi_p \rangle$ for fixed order p and $\boldsymbol{\theta} = \langle \theta_1, \dots \theta_q \rangle$ be the MA coefficients for fixed order q. In addition to these parameters, we must estimate α, the shape parameter of the conditional gamma distribution of the response variable. Alternatively, we could fix $\alpha = 1$ to investigate an exponentially-distributed streamflow.

Consequently, we let $\boldsymbol{\Xi} = \{\mathbf{P}, \boldsymbol{\phi}, \boldsymbol{\theta}, \alpha\}$ be the complete parameter set of the model. Due to the ARMA coefficients, we consider the log-likelihood of the data $\{y_{\ell+1}, \dots y_n\}$, where $\ell = \max\{p, q\}$ for a trial pair p, q. Alternatively, we could define $\ell = \max\{p, q, \mathrm{L}(\kappa_1), \mathrm{L}(\kappa_2), \dots, \mathrm{L}(\kappa_k)\}$ where the function $\mathrm{L} : (\delta, \sigma) \mapsto \mathbb{N}$ is the kernel length function.

Regardless, the log-likelihood in this case is given by

$$\begin{aligned}
\ell(\boldsymbol{\Xi}) &= \sum_{t=\ell+1}^{n} \log f(y_t \mid \underline{\mathbf{H}}_t) \\
&= \sum_{t=\ell+1}^{n} \log\Big(\frac{1}{\Gamma(\alpha)}\Big(\frac{\alpha}{\mu_t}\Big)^{\alpha} y_t^{\alpha-1} \exp\Big(-\frac{\alpha y_t}{\mu_t}\Big)\Big) \\
&= \sum_{t=\ell+1}^{n} \left\{\alpha\Big[\log(\alpha) + \log(y_t) - \log(\mu_t)\Big] - \log(y_t) - \log\big(\Gamma(\alpha)\big) - \frac{\alpha y_t}{\mu_t}\right\}
\end{aligned}$$

Which we can use in our optimization procedure.

Additional Notes: Justification of Mean Paramaterization

In Section 5 of the GARMA Paper in which a concrete example is given, the authors allow a negative binomial conditional distribution on the response. They denote the conditional probability mass function in this case as

$$f_{Y_t}(y_t \mid \underline{\mathbf{H}}_t) = \frac{\Gamma(y_t + k)}{\Gamma(k)\Gamma(y_t + 1)} \left\{\frac{\mu_t}{\mu_t + k}\right\}^{y_t} \left\{\frac{k}{\mu_t + k}\right\}^{k}$$

Where k is a dispersion parameter and μ_t is the conditional mean of the response.

I found online (source) that the mean paramaterization of the negative binomial distribution is given by:

$$\mathbb{P}(X = x) = \frac{\Gamma(r + x)}{x!\,\Gamma(r)} \left(\frac{m}{m + r}\right)^{x} \left(\frac{r}{m + r}\right)^{r}$$

The second and third terms match our paramaterization above, which is a good sign.

Recall that $\Gamma(n) = (n-1)!$ for positive integers. Thus, $\Gamma(y_t + 1) = (y_t + 1 - 1)! = y_t!$ which matches the $x!$ in the cited mean paramaterization of the negative binomial distribution.

Thus, to derive the GARMA conditional PDF for our case, we must paramaterize the Gamma distribution in terms of its mean.