Robustness of Dynamic Score-Driven Models Utilizing the Generalized Gamma Distribution

Petra Tomanová[1]

Abstract. Time series processes might exhibit a complex dynamic structure that cannot be captured by static linear models. However, once model parameters are allowed to be time-varying with possibly complex non-linear updating mechanisms, a huge number of model specification choices arise. A modern framework of generalized autoregressive score (GAS), which lies on the idea that at each time the local fit of the model is improved by the scaled score of the log conditional observation density, helps to overcome this issue. Despite reduced ambiguity of model specification, some arbitrary aspects remain. We investigate the impact of the choice of parametrization, link and scaling functions on the model performance and thus we assess the robustness of the GAS methodology. In this paper, we restrict ourselves to the non-negative time series processes for which we utilize a flexible generalized gamma distribution. Using two data sets of retail and financial durations, we demonstrate that these somewhat arbitrary aspects do not dramatically impact the performance of the GAS models unless numerical issues are encountered.

Keywords: generalized autoregressive score model, generalized gamma distribution, link function, parametrization, scaling function

JEL Classification: C22, C46
AMS Classification: 91G70

1 Introduction

In order to capture the dynamic behavior of time series processes, it might be inevitable to allow model parameters to vary over time. Generally, one of two possible directions might be taken: (i) model parameters are viewed as stochastic processes with their own source of error and thus, they are not perfectly predictable based on past observations; (ii) model parameters are functions of lagged dependent and exogenous variables. We focus on the latter which brings benefits such as perfect predictability based on past information, straightforward evaluation of likelihood, and possible natural extensions to other more complicated dynamics. To avoid the ambiguous choice of updating mechanism for time-varying parameters, we utilize a modern framework of the *generalized autoregressive score (GAS)* proposed by Creal et al. [5].

The GAS models assume that the time-varying parameter of any underlying probability distribution follows a recursion consisting of the autoregressive term and the scaled score of the logarithmic observation density. In spite of the fact that the updating mechanism follows a defined rule, four decisions have to be made – to choose (i) a proper underlying distribution, (ii) its parametrization, (iii) link function, and (iv) scaling function of the score. The underlying distribution should be sufficiently flexible to correctly model the time series process. Its selection is rather natural, however, the "right" choice of its parametrization might be ambiguous. It is commonly driven by criteria such as favourability of the parameter interpretation and ability to ensure that the parameter values satisfy possible constraints such as positiveness. Parametrization alongside with link and scaling function is rather arbitrary with an unclear impact on model performance. In this paper, we investigate this issue for non-negative time series processes. We utilize a flexible generalized gamma distribution which contains several common distributions such as gamma, Weibull, and exponential as special cases.

Our robustness analysis lies in a simulation study for which the time series processes reflect real data behavior. The first dataset is taken from Tomanová and Holý [9] where authors analyze arrivals in queueing systems. Second, we use transactions of a major company traded in the Nordic stock market publicly available in R package `ACDm` [1]. Both datasets are related to the duration analysis, however, the first one concerns the retail data and the second one the financial market data. For a survey of financial duration analysis, see Pacurar [6] and Bhogal and Thekke Variyam [2].

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czechia, petra.tomanova@vse.cz

The aim of this paper is to analyze whether the arbitrary choices in GAS methodology substantially affect the model performance. In Section 2, the generalized autoregressive score framework is recalled. Then, two specifications of generalized gamma distribution, their scores and link functions are discussed in Section 3. Two simulation studies based on two datasets are performed in Section 4 and 5. Section 6 concludes.

2 Generalized Autoregressive Score Framework

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a real-valued stochastic sequence of observations with conditional probability density

$$p_y\left(y_t|h(f_t(\theta));g\right),$$

for all $t\in\mathbb{Z}$, where $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ represents a scalar time-varying parameter that depends on a vector of time-invariant parameters $\theta\in\Theta$, $h:\mathbb{R}\to\mathbb{R}$ is a link function, and $g\in\mathcal{G}$ is a vector of time-invariant parameters that indexes the conditional density p_y. Under the *generalized autoregressive score (GAS)* framework of [5], the dynamic process for the time-varying parameter $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ is given by the autoregressive updating equation

$$\begin{aligned} f_{t+1}(\theta) &= c + bf_t(\theta) + as_t(f_t(\theta);g),\\ s_t(f_t(\theta);g) &= S(f_t(\theta);g)\nabla_t(f_t(\theta);g),\\ \nabla_t(f_t(\theta);g) &= \frac{\partial \log p_y(y_t|f;g)}{\partial f}\Big|_{f=f_t(\theta)}, \end{aligned}$$

where c, b, and a are time-invariant parameters that together with g form the parameter vector θ belonging to the parameter space Θ, and $S(f_t(\theta);g)$ is a univariate scaling factor for the score $\nabla_t(f_t(\theta);g)$ of the conditional observation density p_y. Scaling function is typically set to the k-th power of the inverse of the Fisher information matrix

$$\begin{aligned} S(f_t(\theta);g) &= (\mathcal{I}_t(f_t(\theta);g))^{-k},\\ \mathcal{I}_t(f_t(\theta);g) &= \mathrm{E}_{t-1}[\nabla_t(f_t(\theta);g)\nabla_t(f_t(\theta);g)'], \end{aligned}$$

where typically $k\in\{0,0.5,1\}$, see [4] and [5] for further details and the explanation.

For the purpose of this paper, let the probability density function be defined by the observation equation

$$y_t = h(f_t(\theta))u_t \quad \forall t\in\mathbb{Z},$$

where for all $g\in\mathcal{G}$, $\{u_t\}_{t\in\mathbb{Z}}$ is an independently identically distributed sequence with u_t independent of f_t for every t and $u_t\sim p_{u,g}(u_t)$. Then a probability measure for $\{y_t\}$ is defined when a point (h,p_u,S,θ) is chosen,

$$(h,p_u,S,\theta)\in\mathcal{H}\times\mathcal{P}_u\times\mathcal{S}\times\Theta,$$

where $\mathcal{H}$ denotes the space of link functions, $\mathcal{P}_u$ the space of families of densities p_u for u_t, and $\mathcal{S}$ the space of scaling functions [4]. In this paper, we restrict ourselves to non-negative $\{y_t\}_{t\in\mathbb{Z}}$. As a density p_u, we utilize a flexible generalized gamma distribution and investigate the impact of a link function h and a scaling function S to the model performance while estimating the parameter vector θ by the *maximum likelihood method*.

3 Generalized Gamma Distribution

The *generalized gamma distribution* is a continuous probability distribution for non-negative variables proposed by Stacy [8]. It is a three-parameter generalization of the two-parameter gamma distribution and contains several common distributions such as the exponential and the Weibull distribution as special cases. The distribution has the scale parameter $\beta>0$ and the shape parameters $\psi>0$ and $\varphi>0$. The conditional probability density function is

$$p_y(y|\beta,\psi,\varphi) = \frac{1}{\Gamma(\psi)}\frac{\varphi}{\beta}\left(\frac{y}{\beta}\right)^{\psi\varphi-1}e^{-\left(\frac{y}{\beta}\right)^{\varphi}}$$

for $y\in\mathbb{R}^+$, where $\Gamma(\cdot)$ is the gamma function. Special cases include the gamma distribution for $\varphi=1$, the Weibull distribution for $\psi=1$ and the exponential distribution for $\psi=1$ and $\varphi=1$. The expected value and variance is

$$
\begin{aligned}
\mathrm{E}[Y] &= \beta \frac{\Gamma\left(\psi + \varphi^{-1}\right)}{\Gamma\left(\psi\right)}, \\
\mathrm{var}[Y] &= \beta^2 \frac{\Gamma\left(\psi + 2\varphi^{-1}\right)}{\Gamma\left(\psi\right)} - \left(\beta \frac{\Gamma\left(\psi + \varphi^{-1}\right)}{\Gamma\left(\psi\right)}\right)^2.
\end{aligned}
$$

For $y \in \mathbb{R}^+$, the score and the Fisher information for the parameter β is

$$
\begin{aligned}
\nabla_\beta(y, \beta, \psi, \varphi) &= \frac{\partial \log p_y(y|\beta, \psi, \varphi)}{\partial \beta} = \frac{\varphi}{\beta}\left(y^\varphi \beta^{-\varphi} - \psi\right), \\
\mathcal{I}_\beta(\beta, \psi, \varphi) &= \mathrm{E}\left[\nabla_\beta(y, \beta, \psi, \varphi)^2 \middle| \beta, \psi, \varphi\right] = \frac{\psi\varphi^2}{\beta^2}.
\end{aligned}
$$

As a scaling function, we utilize the inverse of the Fisher information ($k = 1$), the square root of the inverse of the Fisher information ($k = 0.5$), and the unit scaling ($k = 0$). Each choice for the scaling function gives rise to a new GAS model.

We analyze the impact of the most common link function $h(\cdot)$ – logarithmic transformation – which ensures positiveness of the time-varying parameter. The link function $h(\beta)$ affects the score and the Fisher information

$$
\begin{aligned}
\nabla_{h(\beta)}(y, h(\beta), \psi, \varphi) &= \dot{h}^{-1}(\beta)\nabla_\beta(y, \beta, \psi, \varphi), \\
\mathcal{I}_{h(\beta)}(h(\beta), \psi, \varphi) &= \dot{h}'^{-1}(\beta)\mathcal{I}_\beta(\beta, \psi, \varphi)\dot{h}^{-1}(\beta),
\end{aligned}
$$

where $\dot{h} = \partial h(\beta)/\partial \beta'$. When $h(\beta) = \log \beta$ then $\dot{h} = 1/\beta$. Thus, the score and the Fisher information for the parameter $\beta^* = \log \beta$ is

$$
\begin{aligned}
\nabla_{\beta^*}(y, \beta^*, \psi, \varphi) &= \frac{\partial \log p_y(y|\beta^*, \psi, \varphi)}{\partial \beta^*} = \varphi\left(y^\varphi e^{-\varphi\beta^*} - \psi\right), \\
\mathcal{I}_{\beta^*}(\beta^*, \psi, \varphi) &= \mathrm{E}\left[\nabla_{\beta^*}(y, \beta^*, \psi, \varphi)^2 \middle| \beta^*, \psi, \varphi\right] = \psi\varphi^2,
\end{aligned}
$$

for $y \in \mathbb{R}^+$. Note that the Fisher information for β^* is not dependent on β^* itself. Thus, the scaling function choice for this model has no impact on model performance.

An alternative parametrization of the generalized gamma distribution can be utilized. Prentice [7] reparameterized and extended the distribution of the logarithm of a generalized gamma variate. In this paper, we consider the parametrization that is related to the common one as

$$
\begin{aligned}
\beta &= \exp(\mu)\lambda^{\frac{2\sigma}{\lambda}}, & \psi &= \frac{1}{\lambda^2}, & \varphi &= \frac{\lambda}{\sigma}, \\
\mu &= \log\beta + \frac{1}{\varphi}\log\psi, & \sigma &= \frac{1}{\varphi\sqrt{\psi}} & \lambda &= \frac{1}{\sqrt{\psi}},
\end{aligned}
$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda > 0$. The score for the parameter μ is

$$
\nabla_\mu(y, \mu, \sigma, \lambda) = \frac{\partial \log p_y(y|\mu, \sigma, \lambda)}{\partial \mu} = \frac{y^{\frac{\lambda}{\sigma}} \exp\{-\mu\}^{\frac{\lambda}{\sigma}} - 1}{\lambda\sigma}.
$$

To summarize, we consider three common scaling functions (the inverse of the Fisher information, the square root of the inverse of the Fisher information, and the unit scaling), two different link functions (level and logarithmic transformation), and two different parametrizations of the distribution function (common and the alternative one). The analyses are performed on two different data set – retail data and financial market data.

4 Empirical Example of Retail Data

The first data set is taken from Tomanová and Holý [9]. The data are originally obtained from the database of an online bookshop with one brick-and-mortar location in Prague, Czechia. It contains information about order arrivals (number of ordered books) such as time and volume. The aim is to model times between arrivals (durations) for the purpose of process simulation and optimization. The durations are diurnally and seasonally adjusted, see [9] for further details.

The data set covers the period of June 8 – December 20, 2018, resulting in 28 full weeks and 5,753 observations. For adjusted data the unconditional mean is 1.00, the unconditional variance is 1.07 and the distribution is illustrated in Figure 1. The generalized gamma distribution is clearly favored over the exponential distribution for the raw data, however, the difference is not so striking for the adjusted data. However, the conditional generalized gamma distribution provides a considerably better fit even for the adjusted data, see [9] for further details. The estimated GAS model with various specifications is presented in Table 1. The maximum likelihood is obtained for the unit scaling with common parametrization and no transformation (level). However, the other specifications have almost the same log-likelihood.

Furthermore, we conduct a simulation study in order to investigate the effects of model misspecification. We simulate 1,000 observations using the models given by Table 1 and estimate models with various parametrizations, link and scaling functions. We repeat the simulation 20,000 times and report the mean differences in log-likelihoods between the estimated models and the true model in Table 2. All differences between the model performances are negligible. Naturally, the largest differences in log-likelihoods are in the case of the true inverse scaling when assuming unit scaling and vice versa. The square-root scaling offers a middle way. The logarithmic model and the model with the alternative parametrization (which also contains the logarithm of β) behave almost identically.

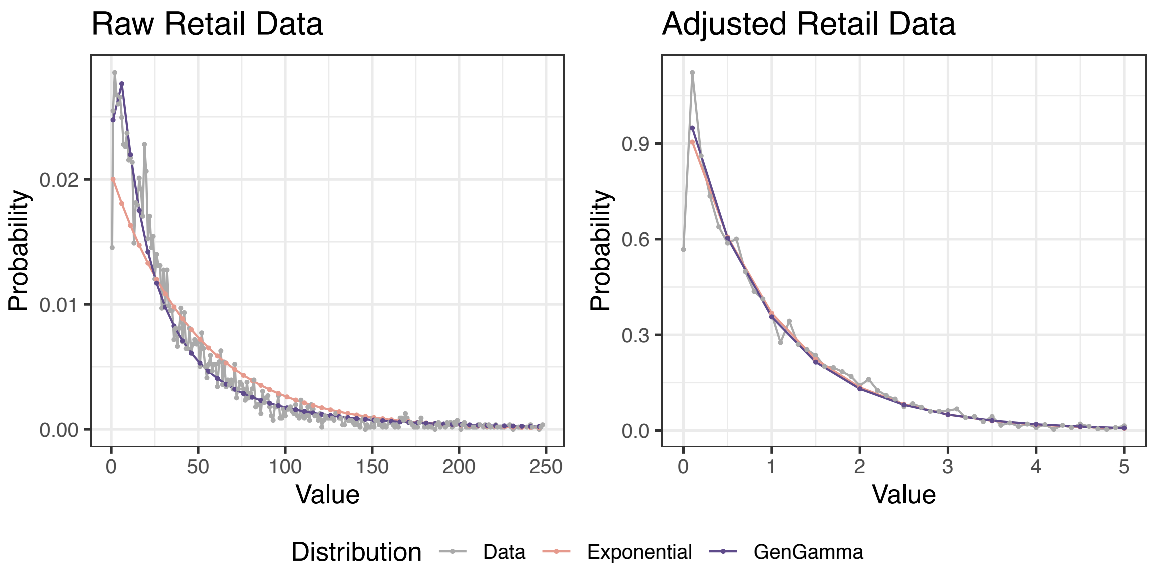

Figure 1 The unconditional probability density function for the retail data.

Model	c	b	a	ψ	φ	Log-Lik	AIC
Level/Unit	0.28	0.65	0.05	1.17	0.89	−5721.49	11452.97
Level/Sqrt	0.26	0.68	0.07	1.16	0.89	−5721.73	11453.46
Level/Inverse	0.23	0.72	0.07	1.15	0.90	−5722.17	11454.33
Log/Unit	−0.06	0.72	0.07	1.15	0.90	−5723.31	11456.62
Alt/Unit	−0.01	0.72	0.07	0.93	1.04	−5723.31	11456.62

Table 1 The estimated coefficients with the log-likelihoods and the AIC for the retail data.

	Estimated Model				
True Model	level/Unit	level/Sqrt	level/Inv	log/Unit	Alt/Unit
level/Unit	0.00	−0.04	−0.19	−0.13	−0.13
level/Sqrt	−0.04	0.00	−0.06	−0.04	−0.04
level/Inv	−0.15	−0.02	0.00	−0.02	−0.02
log/Unit	−0.09	−0.01	−0.02	0.00	0.00
Alt/Unitt	−0.10	−0.02	−0.03	−0.00	0.00

Table 2 The differences in the log-likelihoods for various models based on the retail data.

5 Empirical Example of Financial Market Data

The second data set includes transactions of a major company traded in the Nordic stock market which is publicly available in R package `ACDm` [1]. The data set covers two weeks of intraday transactions recorded at 1-second precision and includes the transaction date, timestamp, price and volume. The total number of transactions is 99,330 from which the price durations are calculated. Price duration is measured as a time until the price process changes by a given level. Our data set can be replicated by using function `ACDm::computeDurations(transData, type = 'price', priceDiff = .01)`. Then the price durations are adjusted by cubic splines with a vector of nodes: `c(seq(600, 1105, 120), 1105)`.

The number of price durations is 2,054, the unconditional mean is 0.99, the unconditional variance is 2.16 and the distribution is illustrated in Figure 2. Again the generalized gamma distribution provides a better fit than the exponential distribution. Table 3 contains the estimated coefficients with log-likelihoods and AIC values for various model specifications. The best models are based on (i) logarithmic transformation, and (ii) alternative parametrization, both with unit scaling. The models based on common parametrization without logarithmic link function perform distinctively worse. The model with unit scaling produces the worst results due to problems with numerical optimization. The source of the problem might be the constrain $\beta > 0$ which is not naturally satisfied as in the case of logarithmic transformation of β.

We conduct a simulation study with 1,000 replications and report the differences in log-likelihoods in Table 4. The results show that the model with logarithmic transformation and the model with alternative parametrization exhibit very similar results in line with the previous study. The model utilizing the common parametrization with inverse scaling performs comparably with those two. The performance of models with unit and square root scaling is very poor due to numerical issues. Thus, when the time-varying parameter has to be positive, the logarithmic transformation should be preferred even if the true model is different since the numerical problems might produce much more severe drops in performance than the model misspecification.

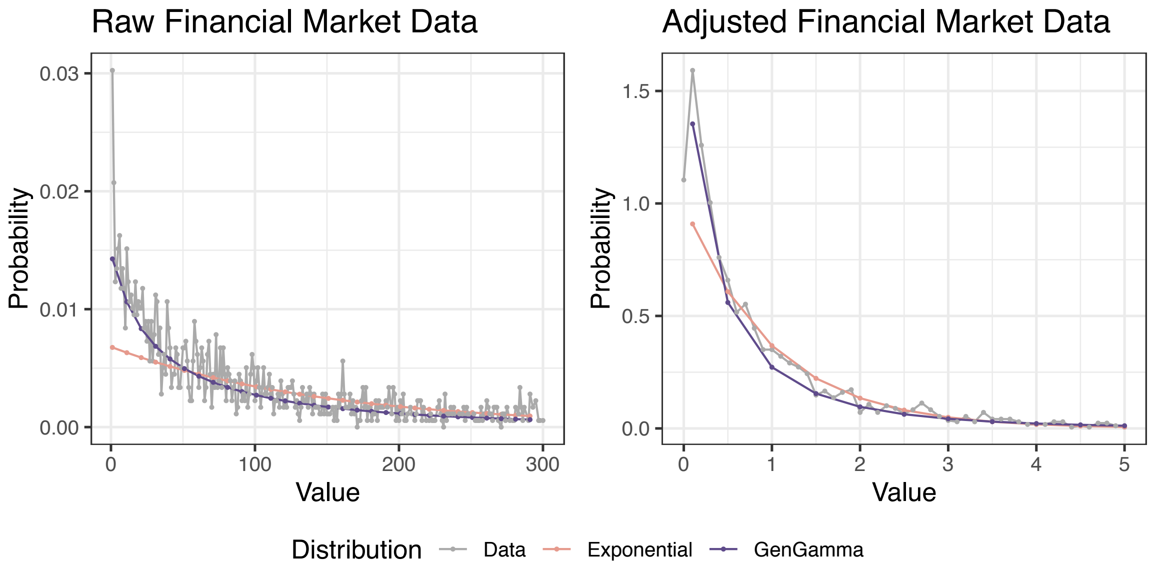

Figure 2 The unconditional probability density function for the financial market data.

Model	c	b	a	ψ	φ	Log-Lik	AIC
Level/Unit	0.01	0.78	0.00	3.14	0.43	−1801.49	3612.98
Level/Sqrt	0.01	0.86	0.01	2.93	0.47	−1724.94	3459.88
Level/Inverse	0.00	0.95	0.21	3.22	0.44	−1722.06	3454.11
Log/Unit	−0.33	0.91	0.32	3.75	0.41	−1718.38	3446.75
Alt/Unit	−0.05	0.91	0.32	0.52	1.26	−1718.38	3446.75

Table 3 The estimated coefficients with the log-likelihoods and the AIC for the financial market data.

	Estimated Model				
True Model	level/Unit	level/Sqrt	level/Inv	log/Unit	Alt/Unit
level/Unit	0.00	1.49	1.36	1.42	1.42
level/Sqrt	−23.48	0.00	−3.19	−2.24	−2.24
level/Inv	−115.06	−30.27	0.00	−1.86	−1.45
log/Unit	−93.31	−16.72	−2.32	0.00	0.00
Alt/Unitt	−92.75	−17.88	−1.88	−0.00	0.00

Table 4 The differences in the log-likelihoods for various models based on the financial market data.

6 Discussion and Conclusion

We investigate the impact of parametrization, link function, and scaling function choices on the GAS model performance, and thus, we address the robustness of the GAS methodology. Our results show that the differences in performance are negligible unless the numerical issues are encountered. This might be easily avoided by utilizing the logarithmic link function or suitable parametrization without causing substantial drops in model performance since our results show that the impact of parametrization or link function is negligible. Thus, the practitioners can select these aspects of GAS models purely based on convenience without a worries that the choice might lead to an unwanted drop in model performance.

Our findings are in line with the paper [3] where authors analyzed financial durations based on GAS models and found that there are no significant differences between the three considered scaling functions. However, they only considered the logarithmic link function and one parametrization. Moreover, the authors used a discrete approach – specifically they utilized the zero-inflated negative binomial distribution. A more extensive study is still missing in the literature, thus in the future work we will analyze other model specifications as well and explore this topic more thoroughly.

Acknowledgements

The work on this paper was supported by the grant No. F4/53/2019 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Belfrage, M. (2016). Package 'ACDm': Tools for Autoregressive Conditional Duration Models. https://CRAN.R-project.org/package=ACDm.

[2] Bhogal, S. K., & Thekke Variyam, R. (2019). Conditional Duration Models for High-Frequency Data: A Review on Recent Developments. *Journal of Economic Surveys*, 33(1), 252–273.

[3] Blasques, F., Holý, V., & Tomanová, P. (2020). Zero-Inflated Autoregressive Conditional Duration Model for Discrete Trade Durations with Excessive Zeros. https://arxiv.org/pdf/1812.07318.pdf.

[4] Blasques, F., Koopman, S. J., & Lucas, A. (2014). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electronic Journal of Statistics*, 8(1), 1088–1112.

[5] Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, 28(5), 777–795.

[6] Pacurar, M. (2008). Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature. *Journal of Economic Surveys*, 22(4), 711–751.

[7] Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, 61(3), 539-544.

[8] Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*, 33(3), 1187–1192.

[9] Tomanová, P., & Holý, V. (2020). Clustering of Arrivals in Queueing Systems: Autoregressive Conditional Duration Approach. https://arxiv.org/pdf/2003.14382.pdf.