Austrian Journal of Statistics
2025, Volume 54, 101–124.
http://www.ajs.or.at/
doi:10.17713/ajs.v54i2.1976

A New Regression Model Based on an Extended Inverse Gaussian Distribution with Application to Soybean Processing Plants in Brazil

Julio Cezar S. Vasconcelos
Universidade Federal de São Paulo

Denize P. dos Santos
Universidade Federal
de Mato Grosso do Sul

Pâmela Rafaela O. B. Cavallari
Universidade Estadual Paulista Júlio
de Mesquita Filho

Edwin M. M. Ortega
Universidade de São Paulo

Roberto Vila
Universidade de Brasília

Gauss M. Cordeiro
Universidade Federal de Pernambuco

Marco Antônio M. Biaggioni
Universidade Estadual Paulista Júlio de Mesquita Filho

Abstract

Grain producers in Brazil often depend on third-party services for the transportation, processing and storage of their production, as, for the most part, they do not have silos on their properties. In this context, efficient logistics is essential to optimize processes and increase reliability between customers and service providers. This study focuses on the logistical analysis of truck traffic at two grain processing plants, examining different receiving protocols to evaluate internal vehicle flow during peak production conditions. The data is analyzed using a multiple regression model with two systematic components based on the proposed New Weibull inverse Gaussian distribution. The research is conducted in grain processing and storage units in the southwest region of São Paulo-SP, belonging to an agro-industrial cooperative. The study monitors all stages of soybean receipt during the peak harvest month, in March 2020. The results indicate the dependence of service times on the sector's logistical variables. This research addresses the pressing need for efficient logistics in the grain industry, especially in soybean processing. By focusing on truck traffic and receiving protocols, the study aims to provide a better understanding to optimize internal logistics processes, thus contributing to improving operational efficiency and customer service in grain processing units.

Keywords: reception/unloading, multiple regression model, service time, simulation study, storage units.

1. Introduction

Soybean growing is among the world's leading agricultural activities. Soybeans are the fourth most consumed grain in the world and the main oilseed. The crop's relevance is due to the many products and byproducts derived from the grains, due to their high level of protein and oil. Their applications range from human food (e.g., production of chocolate, cooking oil, sauces and pasta) to animal feed, plus industrial uses Lu, Carter, and Tegeder (2022); Peoples, Giller, Jensen, and Herridge (2021); Karges, Bellingrath-Kimura, Watson, Stoddard, Halwani, and Reckling (2022).

Brazil has standout presence in the international market for agricultural commodities. But despite the rising production of soybeans in the country, there are many pre and postharvest difficulties facing producers, chiefly regarding transport infrastructure and logistics, price fluctuations and weather conditions, requiring constant efforts by growers to remain competitive in the global market do Nascimento, Pereira, Melo, Freire, de Souza, and Martins (2022); de Holanda, da Silva, de Lavor, and de Sousa (2020).

In Brazil's current conjuncture, logistics has great importance to agribusiness. The logistical hurdles faced by exporters are the main contributor to the so-called "Brazil cost", reflecting the costs of transport of farm products to ports and losses due to postharvest deterioration de Souza (2020). Logistic is the part of the supply chain process related to the planning, implementation, flow and storage of goods.

According to de Oliveira, Marsola, Milanez, and Fatoretto (2022), although Brazil is one of the most important players in the global market for agricultural commodities, and its productivity and output indices have improved substantially in recent years, reducing the logistics costs still pose a major challenge, especially the costs of transport and storage for export.

The application of statistical models such as data envelopment analysis (DEA) Melo, Péra, Júnior, do Nascimento Rebelatto, and Caixeta-Filho (2020) and mixed-integer nonlinear programming (MINLP) model Branco, Branco, de Aguiar, Filho, and Rodrigues (2019), which seek to optimize the performance of long-distance cargo transport through better location of processing and storage units, has been essential to adjust transport infrastructure to the restrictions related to the processing capacity of these units. Toloi, dos Reis, Abraham, Toloi, and Bueno (2021) identified the factors that influence the quality and production of the soybean supply network in the state of Mato Grosso from a set of three decision variables: marketing, logistics and rural production. By means of the analytic hierarchy process (AHP), the authors identified logistics (marketing, transport and storage) as the factor with greatest influence on the quality and production of the supply network, highlighting the lack of adequate and sufficient storage areas, besides the precarious conditions of highways, causing losses of grains during movement between production areas and the final destination.

The impact of truck congestion management at container terminals has also been studied extensively in recent years. Discrete event simulation models Neagoe, Hvolby, Taskhiri, and Turner (2021); Pimpanit and Jarumaneeroj (2021); Rotunno, Zupone, Carnimeo, and Fanti (2023) or that based on non-stationary queuing theory Zhang, Zeng, and Yang (2019) have been used for various purposes such as scheduling appointments for truck arrivals Yi, Scholz-Reiter, Kim, and Kim (2019), to support investments for improving efficiency and turnover of logistic systems Li, Govindan, and Jin (2018) in addition to contributing to the improvement of understanding of the differential impacts of congestion management initiatives on truck and environmental performance in bulk cargo terminals Neagoe *et al.* (2021), Li *et al.* (2018).

Due to the current challenges faced by grain processing cooperatives, among them the long waiting lines of trucks to load and unload grains, this study has the objective of analyzing the logistics of this link in the soybean production chain. To achieve this objective, we initially propose a new distribution, consisting of a combination of the Weibull and inverse Gaussian distributions, called the new *Weibull inverse Gaussian* ("NWIG" for short) distribution. We obtain some of its mathematical properties and then propose a multiple regression model

with two systematic components. We also address goodness of fit and residual analysis for the new multiple regression model. We report the application of this distribution to analyze two soybean storage and processing units, both located in the southwestern region of the state of São Paulo, that have different protocols for receiving grains, to evaluate the efficiency of the internal flow of vehicles during the peak harvest season, and identify the bottlenecks, to enable optimizing their processes. In the protocol used by unit A, each truck enters the premises and waits to be unloaded, while in unit B, before entering the premises, the documents are checked and the truck's data are inserted in a computerized system for determining the order of unloading, thus rationalizing the unloading process.

The rest of the paper presents the following topics. Section 2 introduces the NWIG distribution. Section 3 presents some of its mathematical properties. Section 4 defines the NWIG multiple regression model for censored data, and Section 5 describes some simulations and residual analysis. The usefulness of our results is illustrated by means of soybean data in Section 6. Finally, Section 7 addresses some conclusions.

2. The NWIG distribution

The cumulative distribution function (cdf) of the new-Weibull-G (NW-G) family Cordeiro, Vasconcelos, dos Santos, Ortega, and Sermarini (2022) has the form

$$F(y;\nu,\tau) = \exp\left\{-\nu\left[-\log G(y)\right]^{\tau}\right\}, \qquad y \in \mathbb{R}, \tag{1}$$

where $G(y)$ is any parent cdf, and $\nu > 0$ and $\tau > 0$ are shape parameters.

For the baseline, consider the inverse Gaussian (IG) distribution, which is widely adopted to analyze skewed data, under the reparameterization Johnson, Kotz, and Balakrishnan (1994) in the gamlss package Stasinopoulos and Rigby (2007) in **R**.

The cdf of the IG distribution is (for $y > 0$)

$$\begin{aligned}\Psi(y) = \Psi(y;\mu,\sigma) &= \int_0^y g(t)\,\mathrm{d}t \\ &= \Phi\left(\sqrt{\frac{1}{\sigma^2 y}}\left(\frac{y}{\mu}-1\right)\right) + \exp\left(\frac{2}{\sigma^2\mu}\right)\Phi\left(-\sqrt{\frac{1}{\sigma^2 y}}\left(\frac{y}{\mu}+1\right)\right),\end{aligned} \tag{2}$$

where $\Phi(\cdot)$ is the standard normal cdf, $\mu > 0$ is a location, and $\sigma > 0$ is a scale.

The NWIG cdf (with four positive parameters) is defined by inserting the IG cdf in Equation (1)

$$F(y) = \exp\left\{-\nu[-\log\Psi(y)]^{\tau}\right\}, \quad y > 0, \tag{3}$$

where $\mu > 0$ represents a location parameter, $\sigma > 0$ represents a scale parameter, and the additional parameters ν and τ are introduced, where $\nu > 0$ denotes asymmetry and $\tau > 0$ denotes shape. These parameters contribute to the model's flexibility in capturing asymmetry and shape characteristics within the distribution.

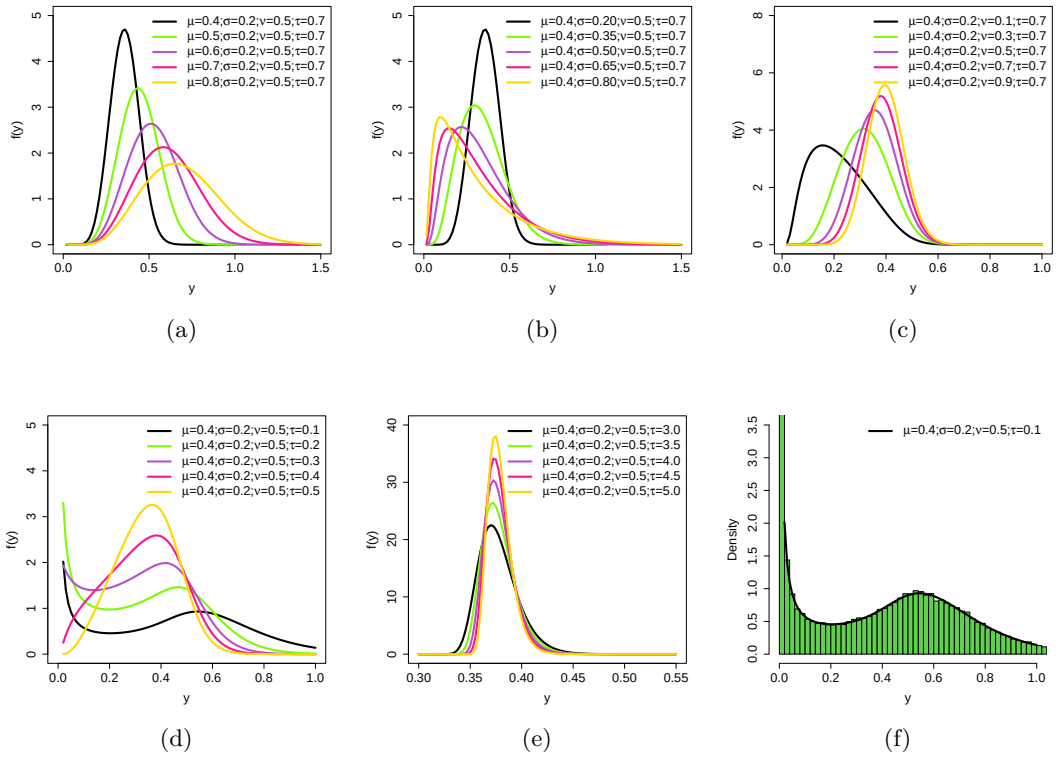
For simplicity, let $Y \sim \mathrm{NWIG}(\mu,\sigma,\nu,\tau)$ have cdf (3). The probability density function (pdf) of Y reduces to

$$f(y) = \frac{\nu\,\tau\left\{\frac{1}{\sqrt{2\pi\,\sigma^2 y^3}}\,\exp\left[-\frac{1}{2\mu^2\sigma^2 y}(y-\mu)^2\right]\right\}}{\Psi(y)}\left[-\log\Psi(y)\right]^{\tau-1}\exp\left\{-\nu\left[-\log\Psi(y)\right]^{\tau}\right\}. \tag{4}$$

Some possible shapes of the density function of Y are displayed in Figures 1(a)-1(e).

The plot comparing the exact NWIG density and the histogram from a simulated data set for some parameter values is reported in Figure 1(f), which indicate that the simulated values are consistent with the NWIG distribution. It can also be seen graphically from figures

1(d) and 1(e) that for values of the parameter $\tau \in (0,1)$ the NWIG distribution is more flexible, meaning it is capable of adapting to a wider range of shapes. In this interval, the distribution can model various forms of asymmetry and kurtosis, including heavier tails and sharper peaks, which allows it to capture more complex data patterns and behaviors. This flexibility is particularly useful when modeling real-world data that may not follow a simple or symmetric distribution. In contrast, for values of the parameter $\tau > 1$ it only produces a unimodal form.



(a) (b) (c)

(d) (e) (f)

Figure 1: Plots of the NWIG density function for one simulated data set. (a) σ=0.2, ν=0.5, τ=0.7 and varying μ; (b) μ=0.4, ν=0.5, τ=0.7 and varying σ; (c) μ=0.4, σ=0.2, τ=0.7 and varying ν; (d) μ=0.4, σ=0.2, ν=0.5 and varying τ in $(0,1)$; (e) $\mu = 0.4$, $\sigma = 0.2$. $\nu = 0.5$ and varying $\tau > 1$; (f) Histogram and NWIG density for simulated data.

The Weibull distribution is a positive continuous probability distribution very well-known in the literature, described in 1939 by the mathematician Waloddi Weibull (Weibull 1939). For the convenience of the reader and for the reading to be self-contained, we define this distribution in the sequence.

A random variable W has a Weibull distribution if its pdf is given by

$$f_W(w) = bkw^{k-1}\exp(-bw^k), \quad w > 0,$$

where $k > 0$ is the shape parameter and $b > 0$ is the scale parameter of the distribution. For simplicity of notation we write $W \sim \text{Weibull}(b,k)$.

3. Some properties

We provide some mathematical properties of $Y \sim \text{NWIG}(\mu,\sigma,\nu,\tau)$ below.

3.1. Modality

Equation (4) can be expressed as

$$f(y) = f_W(-\log \Psi(y)) \left[\frac{\Psi'(y)}{\Psi(y)}\right], \quad W \sim \text{Weibull}(\nu, \tau).$$

By differentiating $f(y)$ with respect to y gives

$$f'(y) = -f'_W(-\log \Psi(y)) \left[\frac{\Psi'(y)}{\Psi(y)}\right]^2 + f_W(-\log \Psi(y)) \frac{\Psi''(y)\Psi(y) - [\Psi'(y)]^2}{\Psi(y)^2}, \tag{5}$$

where

$$\Psi'(y) = g(y) = \frac{\nu\tau}{\sqrt{2\pi\sigma^2 y^3}} \exp\left[-\frac{1}{2\mu^2\sigma^2 y}(y-\mu)^2\right]$$

and

$$\Psi''(y) = \frac{\Psi'(y)}{y} \left[\frac{(y-\mu)^2}{2\mu^2\sigma^2 y} - \frac{y-\mu}{\mu^2\sigma^2} - \frac{3}{2}\right].$$

Equation (5) reduces to

$$f'(y) = f_Y(y) \left[\frac{\Psi'(y)}{\Psi(y)}\right] \left\{\frac{\nu\tau[-\log \Psi(y)]^\tau + 1 - \tau}{[-\log \Psi(y)]^{\tau-1}} + \frac{\Psi(y)}{\Psi'(y)} \frac{1}{y} \left[\frac{(y-\mu)^2}{2\mu^2\sigma^2 y} - \frac{y-\mu}{\mu^2\sigma^2} - \frac{3}{2}\right] - 1\right\}.$$

Then, $f'(y) = 0$ if and only if

$$\frac{\nu\tau[-\log \Psi(y)]^\tau + 1 - \tau}{[-\log \Psi(y)]^{\tau-1}} + \frac{\Psi(y)}{\Psi'(y)} \frac{1}{y} \left[\frac{(y-\mu)^2}{2\mu^2\sigma^2 y} - \frac{y-\mu}{\mu^2\sigma^2} - \frac{3}{2}\right] - 1 = 0,$$

or, equivalently, $\mathbb{R}(y) = \mathbb{S}(y)$, where

$$\begin{aligned} \mathbb{R}(y) &= \left\{\frac{\nu\tau[-\log \Psi(y)]^\tau + 1 - \tau}{[-\log \Psi(y)]^{\tau-1}} - 1\right\} \frac{1}{\Psi(y)}, \\ \mathbb{S}(y) &= -\sqrt{2\pi\sigma^2 y} \exp\left[\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right] \left[\frac{(y-\mu)^2}{2\mu^2\sigma^2 y} - \frac{y-\mu}{\mu^2\sigma^2} - \frac{3}{2}\right]. \end{aligned} \tag{6}$$

Then, all critical points of $f(y)$ are positive roots of $\mathbb{R}(y) = \mathbb{S}(y)$.

It is difficult to determine the number of positive roots of this equation. In Theorem 3, we show theoretically that it has exactly one positive root in a very special case. In general, we believe that the above equation has at most two positive roots.

Before stating and proving Theorem 3, the following three preliminary results are indispensable.

Theorem 1 (Ulrich and Watson (1994))**.** *Let $f(z) = az^4 + bz^3 + cz^2 + dz + e$ be a quartic polynomial with real coefficients and $a > 0$, $e > 0$. Define*

$$\alpha = ba^{-3/4}e^{-1/4}, \quad \beta = ca^{-1/2}e^{-1/2}, \quad \gamma = da^{-1/4}e^{-3/4},$$

$$\Delta = 4\left[\beta^2 - 3\alpha\gamma + 12\right]^3 - \left[72\beta + 9\alpha\beta\gamma - 2\beta^3 - 27\alpha^2 - 27\gamma^2\right]^2,$$

$$\Lambda_1 \equiv (\alpha-\gamma)^2 - 16(\alpha+\beta+\gamma+2), \quad \Lambda_2 \equiv (\alpha-\gamma)^2 - \frac{4(\beta+2)}{\sqrt{\beta-2}}(\alpha+\gamma+4\sqrt{\beta-2}).$$

Then $f(z) \geqslant 0$ for all $z > 0$ if and only if

(1) $\beta < -2$ and $\Delta \leqslant 0$ and $\alpha + \gamma > 0$;

$$\textit{(2)}\ -2 \leqslant \beta \leqslant 6 \textit{ and } \begin{cases} \Delta \leqslant 0 \textit{ and } \alpha + \gamma > 0 \\ \textit{or} \\ \Delta \geqslant 0 \textit{ and } \Lambda_1 \leqslant 0; \end{cases}$$

$$\textit{(3)}\ 6 < \beta \textit{ and } \begin{cases} \Delta \leqslant 0 \textit{ and } \alpha + \gamma > 0 \\ \textit{or} \\ \alpha > 0 \textit{ and } \gamma > 0 \\ \textit{or} \\ \Delta \geqslant 0 \textit{ and } \Lambda_2 \leqslant 0. \end{cases}$$

Proposition 1. *Let $p_4(y) = ay^4 + by^3 + cy^2 + dy + e$ be a polynomial of fourth degree with coefficients*

$$a = 1, \quad b = 6\mu^2\sigma^2, \quad c = \mu^2(3\mu^2\sigma^2 - 2), \quad d = -2\mu^4\sigma^2, \quad e = \mu^4.$$

Then, $p_4(y) \geqslant 0$ for all $y > 0$ provided $\mu\sigma \leqslant 2/\sqrt{3}$.

Proof. Using the notations employed in Theorem 1, some algebraic manipulations lead to

$$\begin{aligned} &-2 < \beta = 3\mu^2\sigma^2 - 2 \leqslant 6, \quad \alpha + \gamma = 4\mu\sigma^2 > 0, \\ &\Delta = -(46656\mu^8\sigma^{16} + 186624\mu^6\sigma^{12} + 435456\mu^4\sigma^8 + 110592\mu^2\sigma^4) < 0, \end{aligned}$$

because $\mu\sigma \leqslant 2/\sqrt{3}$. Since $a = 1 > 0$ and $e = \mu^4 > 0$, the non-negativity of $p_4(y)$ follows from Item (2) of Theorem 1. □

Lemma 2. *The function $y \longmapsto \mathbb{S}(y)$ in (6) is non-decreasing and continuous on $(0, \infty)$ provided $\mu\sigma \leqslant 2/\sqrt{3}$. Further, $\lim_{y\to 0^+} \mathbb{S}(y) = -\infty$ and $\lim_{y\to\infty} \mathbb{S}(y) = \infty$.*

Proof. The continuity of $\mathbb{S}$ and the limits $\lim_{y\to 0^+} \mathbb{S}(y) = -\infty$ and $\lim_{y\to\infty} \mathbb{S}(y) = \infty$ can be observed directly from the definition (6) of $\mathbb{S}$. Differentiating $\mathbb{S}$ with respect to y gives

$$\mathbb{S}'(y) = \frac{\sqrt{\pi/2}}{2\mu^4\sigma^3} \exp\left[\frac{(y-\mu)^2}{2\mu^2\sigma^2 y}\right] \frac{p_4(y)}{y^{5/2}}, \quad y > 0.$$

By using Proposition 1, it follows $\mathbb{S}'(y) \geqslant 0$ for all $y > 0$. This completes the proof. □

Let $\mathcal{A}$ be a set of parameters $(\mu, \sigma, \nu, \tau) \in (0, \infty) \times (0, \infty) \times (0, \infty) \times (0, \infty)$ satisfying the condition

$$\mathbb{R}'(y) < 0, \text{ for all } y > 0, \quad \lim_{y\to 0^+} \mathbb{R}(y) = \infty, \quad \lim_{y\to\infty} \mathbb{R}(y) < 0 \quad \text{and} \quad \lim_{y\to 0^+} f(y) = 0,$$

where $f(y)$ is the pdf in (4).

Notice that $\mathcal{A}$ is non-empty, since $(\mu, \sigma, \nu, 1)$ (with $\nu \geqslant 1$) belongs to $\mathcal{A}$. Indeed, this is a straightforward since (for $\tau = 1$ and $\nu \geqslant 1$)

$$\begin{aligned} &\mathbb{R}(y) = \frac{\nu\tau[-\log\Psi(y)] - 1}{\Psi(y)}, \quad \mathbb{R}'(y) = \frac{\nu \log\Psi(y) - (\nu - 1)}{\Psi^2(y)} \Psi'(y) < 0, \\ &\lim_{y\to 0^+} \mathbb{R}(y) = \infty, \quad \lim_{y\to\infty} \mathbb{R}(y) = -1, \end{aligned}$$

and

$$f(y) = \frac{\nu}{\sqrt{2\pi\sigma^2 y^3}} \exp\left[-\frac{1}{2\mu^2\sigma^2 y}(y-\mu)^2\right] \Psi^{\nu-1}(y), \quad \lim_{y\to 0^+} f(y) = 0.$$

Theorem 3. *If $(\mu, \sigma, \nu, \tau) \in \mathcal{A}$, then the pdf $f(y)$ in (4) is unimodal provided $\mu\sigma \leqslant 2/\sqrt{3}$.*

Proof. By (6) all critical points y of the pdf of Y satisfy $\mathbb{R}(y) = \mathbb{S}(y)$. Note that $y \longmapsto \mathbb{R}(y)$ is a decreasing (and continuous function) because $(\mu, \sigma, \nu, \tau) \in \mathcal{A}$. Since $y \longmapsto \mathbb{S}(y)$ is non-decreasing (see Lemma 2), it is clear that there is a unique point $y_0 \in (0, \infty)$, which is the solution of $\mathbb{R}(y) = \mathbb{S}(y)$, i.e. the pdf $f(y)$ has a single critical point. Since $(\mu, \sigma, \nu, \tau) \in \mathcal{A}$, $\lim_{y\to 0^+} f(y) = 0$. Further, $\lim_{y\to\infty} f(y) = 0$ because $f(y)$ is a pdf. Hence, it follows that $f(y)$ is unimodal with mode y_0. □

3.2. Stochastic representation

First, let

$$T(y; \mu, \sigma) = \log \Psi(y; \mu, \sigma). \tag{7}$$

The inverse function of T, say T^{-1}, is

$$T^{-1}(y; \mu, \sigma) = \Psi^{-1}(\exp(y); \mu, \sigma).$$

Proposition 2. *1. If $W \sim$ Weibull(ν, τ), then $Y = T^{-1}(-W; \mu, \sigma) \sim$ NWIG(μ, σ, ν, τ).*

2. If $Y \sim$ NWIG(μ, σ, ν, τ), then $W = -T(Y; \mu, \sigma) \sim$ Weibull(ν, τ).

Proof. The proof is immediate since

$$\mathbb{P}(Y \leqslant y) = \mathbb{P}(-W \leqslant T(y; \mu, \sigma)) = \mathbb{P}(T^{-1}(-W; \mu, \sigma) \leqslant y), \tag{8}$$

where $W \sim \text{Weibull}(\nu, \tau)$. □

3.3. Scale family

Proposition 3. *If $Y \sim$ NWIG(μ, σ, ν, τ), then $kY \sim$ NWIG$(k\mu, \sigma/\sqrt{k}, \nu, \tau)$ (for $k > 0$).*

Proof. It is a simple task to observe that Ψ in (2) satisfies the following relation:

$$\Psi\left(\frac{y}{k}; \mu, \sigma\right) = \Psi\left(y; k\mu, \frac{\sigma}{\sqrt{k}}\right), \quad k > 0.$$

Taking logarithm to both sides of the identity above and then using the definition (7) of T, we obtain

$$T\left(\frac{y}{k}; \mu, \sigma\right) = T\left(y; k\mu, \frac{\sigma}{\sqrt{k}}\right). \tag{9}$$

By combining (8) with (9), we have (for $k > 0$)

$$\begin{aligned}\mathbb{P}(kY \leqslant y) = \mathbb{P}\left(Y \leqslant \frac{y}{k}\right) &\overset{(8)}{=} \mathbb{P}\left(-W \leqslant T\left(\frac{y}{k}; \mu, \sigma\right)\right) \\ &\overset{(9)}{=} \mathbb{P}\left(-W \leqslant T\left(y; k\mu, \frac{\sigma}{\sqrt{k}}\right)\right) = \mathbb{P}\left(T^{-1}\left(-W; k\mu, \frac{\sigma}{\sqrt{k}}\right) \leqslant y\right),\end{aligned}$$

where $W \sim \text{Weibull}(\nu, \tau)$. In other words, kY and $T^{-1}(-W; k\mu, \sigma/\sqrt{k})$ are equal in distribution. Consequently, by applying Proposition 2 (1) we have $kY \sim \text{NWIG}(k\mu, \sigma/\sqrt{k}, \nu, \tau)$. The proof is complete. □

3.4. Quantile function

Inverting Equation (3), the quantile function (qf) of Y can be expressed as

$$y = Q_Y(u) = Q_{\mathrm{IG}}\left(\exp\left[-\left(-\frac{\log(u)}{\nu}\right)^{1/\tau}\right]\right), \quad u \in (0,1), \tag{10}$$

where $Q_{\mathrm{IG}}(u) = \Psi^{-1}(u)$ is the qf of the IG distribution.

3.5. Existence of moments

The expectation of Y^p (for $p > 0$ and some finite $a > 0$) can be expressed as

$$\begin{aligned}\mathbb{E}(Y^p) &= \int_0^a y^p f(y)\,\mathrm{d}y + \int_a^\infty y^p f(y)\,\mathrm{d}y \\ &\leqslant a^p + \int_a^\infty y^p f(y)\,\mathrm{d}y = a^p + I_a,\end{aligned}$$

where the notation $I_a = \int_a^\infty y^p f(y)\,\mathrm{d}y$ was adopted. Then, Y^p is integrable if $I_a < \infty$. Assuming the condition:

$$\lim_{y\to\infty} \frac{y^p f(y)}{y^\alpha} = 0, \tag{11}$$

from Theorem 2.2 (the Magic Exponent -1) of Geyer (2006), it follows that $I_a < \infty$. So, the following result holds.

Theorem 4. *The positive order moments of Y always exist.*

We now verify the validity of the condition (11).

Proposition 4. *For any $\alpha < -1$, the limit in (11) is valid.*

Proof. Since, for any $\alpha < -1$,

$$\begin{aligned}\lim_{y\to\infty} \frac{y^p f(y)}{y^\alpha} &= \lim_{y\to\infty} f_W(-\log\Psi(y))\left[\frac{y^{p-\alpha}\Psi'(y)}{\Psi(y)}\right], \quad W \sim \mathrm{Weibull}(\nu,\tau), \\ &= \lim_{y\to\infty} f_W(-\log\Psi(y)) \lim_{y\to\infty}\left[y^{p-\alpha}\Psi'(y)\right],\end{aligned}$$

where

$$\Psi'(y) = g(y) = \frac{\nu\tau}{\sqrt{2\pi\sigma^2 y^3}} \exp\left[-\frac{1}{2\mu^2\sigma^2 y}(y-\mu)^2\right],$$

it is enough to verify that

$$\lim_{y\to\infty}\left[y^{p-\alpha}\Psi'(y)\right] = 0. \tag{12}$$

Indeed, by using the known inequality $\exp(-x) \leqslant n!/x^n$, for all $x > 0$ and $n \in \mathbb{N}$, we get

$$y^{p-\alpha}\exp\left[-\frac{1}{2\mu^2\sigma^2 y}(y-\mu)^2\right] \leqslant 2n!\mu^2\sigma^2\,\frac{y^{(p-\alpha)+n}}{(y-\mu)^{2n}}, \quad \forall n \in \mathbb{N}.$$

By considering $n = \lfloor p-\alpha \rfloor + 1$ and then, by taking $x \to \infty$ in the above inequality, it follows from Squeeze theorem

$$\lim_{y\to\infty} y^{p-\alpha}\exp\left[-\frac{1}{2\mu^2\sigma^2 y}(y-\mu)^2\right] = 0.$$

Consequently, (12) is valid, which completes the proof. □

4. The NWIG multiple regression model

The NWIG multiple regression model is defined by the random component in Equation 4 and two systematic components for the parameters μ and σ given by

$$\mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1) \quad \text{and} \quad \sigma_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2), \qquad i = 1, \ldots, n, \tag{13}$$

respectively, where $\boldsymbol{\beta}_1 = (\beta_{11}, \cdots, \beta_{1p})^\top$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \cdots, \beta_{2p})^\top$ are the vectors of regression coefficients, and $\mathbf{x}_i^\top = (x_{i1}, \cdots, x_{ip})$ is a vector of known covariates.

The total log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \nu, \tau)^\top$ from the observations $y_1, \cdots, y_n$ reduces to

$$\begin{aligned} l(\boldsymbol{\theta}) \quad = \quad & n\log(\nu\,\tau) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log(\sigma_i^2 y_i^3) - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{(y_i-\mu_i)^2}{\mu_i^2\sigma_i^2 y_i}\right] - \sum_{i=1}^{n}\log[\Psi(y_i)] \\ & + (\tau-1)\sum_{i=1}^{n}\log\{-\log[\Psi(y_i)]\} - \nu\sum_{n=1}^{n}\{-\log[\Psi(y_i)]\}^\tau. \end{aligned} \tag{14}$$

The MLE $\widehat{\boldsymbol{\theta}}$ in $\boldsymbol{\theta}$ can be calculated by maximizing the log-likelihood (14) numerically using the `gamlss` package (Stasinopoulos and Rigby 2007) available in the **R** software.

What we have done is to implement the NWIG, as described in Section 4.2 in (Stasinopoulos, Rigby, and Akantziliotou 2008). In both regressions we used the function *gamlss*$(\cdot)$, and the estimation algorithm used is RS (Rigby and Stasinopoulos 2005) (the initials RS comes from the authors' names), this algorithm does not use the value expected from cross-derivatives, it maximizes the likelihood on each parameter (μ; σ; ν and τ) in turn until it reaches convergence, in general, the algorithm is stable and fast, so we had no computational problems, such as convergence and time. This fact is empirically proven in Section 5, where some simulations of the proposed regression model are performed in different scenarios. We emphasize that we implemented the NWIG regression model in `gamlss` package Stasinopoulos and Rigby (2007) in **R**. can be obtained from the GitHub link: `https://github.com/DenizePalmito/Implementation-NWIG/blob/main/Implementation_NWIG_in_gamlss.R`

Confidence intervals and hypothesis tests under general regularity conditions can be conducted using the large sample distribution of $\widehat{\boldsymbol{\theta}}$, which is multivariate normal distribution with the covariance matrix given by the inverse of the observed information matrix. More specifically, the asymptotic covariance matrix is given by $\mathbf{I}^{-1}(\boldsymbol{\theta})$ with $\mathbf{I}(\boldsymbol{\theta}) = E[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ such that $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\{\partial^2 l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T\}$.

It is difficult to calculate the expected information matrix $\mathbf{I}(\boldsymbol{\theta})$, but it is possible to use $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, evaluated at the MLE $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, which is consistent. The asymptotic normal approximation for $\widehat{\boldsymbol{\theta}}$ may be expressed as $\widehat{\boldsymbol{\theta}}^T \sim N_{(2p+2)}\{\boldsymbol{\theta}^T, \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\}$.

5. Simulation study

Monte Carlo simulations are replicated 1,500 times for sample sizes $n = 60, 180, 360$ and $1,000$ to evaluate the accuracy of the MLEs. For the NWIG distribution and NWIG multiple regression model, the average estimates (AEs), biases, and means squared errors (MSEs) are calculated.

5.1. NWIG distribution

The steps for the simulations of the NWIG distribution are:

- Generate random numbers $u \sim$U$(0, 1)$;
- Obtain NWIG values from Equation (10);

- Set $\mu = 0.4$, $\sigma = 0.2$, $\nu = 0.1$ and $\tau = 0.7$ (Scenario 1) and $\mu = 0,4$, $\sigma = 0,2$, $\nu = 0,5$ and $\tau = 0,1$ (Scenario 2). These values for the parameters were chosen arbitrarily, and the densities with these values can be seen in Figures 1(c) and 1(d);

- Then, the estimates $\hat{\mu}$, $\hat{\sigma}$, $\hat{\nu}$ and $\hat{\tau}$ are calculated.

The results for simulating the new distribution in Table 1 reveal that the estimates are quite accurate, i.e., the biases and MSEs decrease toward zero as n grows.

Table 1: Findings from the NWIG distribution

		Scenario 1			Scenario 2		
Sample size	Parameter	AEs	Bias	MSEs	AEs	Bias	MSEs
$n = 60$	μ	0.3241	-0.0759	0.0185	0.3856	-0.0144	0.0071
	σ	0.3918	0.1918	0.1155	0.2300	0.0300	0.0104
	ν	0.4740	0.3740	0.3919	0.6616	0.1616	0.2087
	τ	0.6515	-0.0485	0.0350	0.1083	0.0083	0.0124
$n = 180$	μ	0.3512	-0.0488	0.0080	0.3944	-0.0056	0.0025
	σ	0.2917	0.0917	0.0303	0.2132	0.0132	0.0029
	ν	0.2723	0.1723	0.1106	0.5548	0.0548	0.0563
	τ	0.6536	-0.0464	0.0120	0.1077	0.0077	0.0041
$n = 360$	μ	0.3724	-0.0276	0.0037	0.3964	-0.0036	0.0012
	σ	0.2494	0.0494	0.0112	0.2094	0.0094	0.0014
	ν	0.1805	0.0805	0.0295	0.5300	0.0300	0.0256
	τ	0.6758	-0.0242	0.0056	0.1066	0.0066	0.0019
$n = 1.000$	μ	0.3865	-0.0135	0.0012	0.3969	-0.0031	0.0004
	σ	0.2227	0.0227	0.0029	0.2069	0.0069	0.0004
	ν	0.1308	0.0308	0.0049	0.5175	0.0175	0.0084
	τ	0.6896	-0.0104	0.0018	0.1045	0.0045	0.0007

5.2. NWIG multiple regression model

The simulations for the NWIG multiple regression model follow the steps:

- The systematic components are $\mu_i = \exp(\beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \beta_{13}x_{i3} + \beta_{14}x_{i4})$ and $\sigma_i = \exp(\beta_{20} + \beta_{21}x_{i1} + \beta_{22}x_{i2} + \beta_{23}x_{i3} + \beta_{24}x_{i4})$, where β_{10} =1.4321, β_{11} =-1.1963, β_{12} =0.1673, β_{13} =0.9467, β_{14} =0.2317, β_{20} =-2.2894, β_{21} =0.1726, β_{22} =0.8145, β_{23} =0.2153, β_{24} =1.2374, ν =0.5683 and τ =1.2852. It's worth noting that the parameter values were chosen at random, i.e. they are hypothetical values, and the covariates are generated from $x_{i1} \sim$ Bernoulli (0.5), $x_{i2} \sim$ Normal $(0, 1)$, $x_{i3} \sim$ Uniform $(0, 1)$ and $x_{i4} \sim$ Normal $(0, 1)$;

- Then, the response variable $y_i \sim \text{NWIG}(\mu_i, \sigma_i, \nu, \tau)$ is generated from (10);

- For each fitted model, we compute the AEs, biases and MSEs.

Figure 2 shows the step-by-step simulation study. The numbers in Table 2 indicate that the AEs tend to be closer to the true parameter values, and the biases and MSEs decay to zero quickly when n increases, which support the consistency of the estimates.

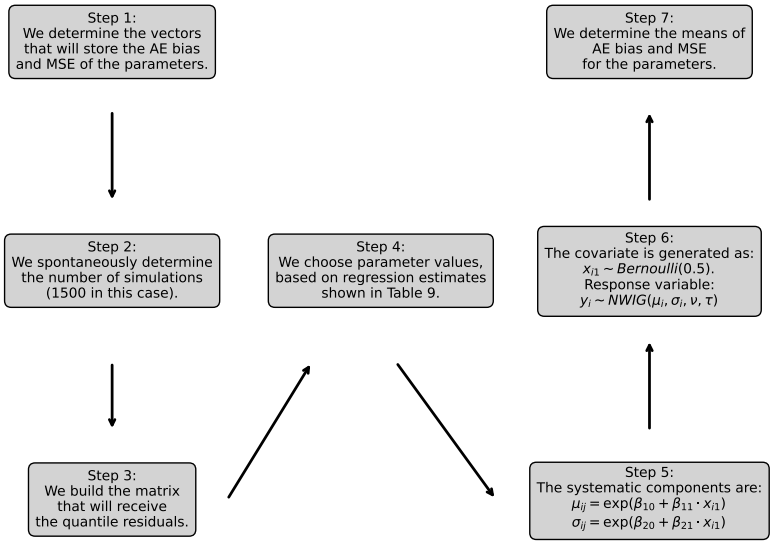


Figure 2: Simulation study flowchart

Table 2: Findings from the NWIG multiple regression model

	$n = 60$			$n = 180$		
Parameter	AEs	Bias	MSEs	AEs	Bias	MSEs
β_{10}	1.4245	-0.0076	0.0025	1.4295	-0.0026	0.0003
β_{11}	-1.1943	0.0020	0.0004	-1.1958	0.0005	<0.0001
β_{12}	0.1656	-0.0017	0.0003	0.1668	-0.0005	<0.0001
β_{13}	0.9495	0.0028	0.0011	0.9468	0.0001	0.0001
β_{14}	0.2294	-0.0023	0.0006	0.2310	-0.0007	0.0001
β_{20}	-2.5200	-0.2306	0.2152	-2.3420	-0.0526	0.0568
β_{21}	0.2369	0.0643	0.0265	0.1789	0.0063	0.0079
β_{22}	0.8427	0.0282	0.0125	0.8223	0.0078	0.0030
β_{23}	0.3678	0.1525	0.0877	0.2548	0.0395	0.0224
β_{24}	1.2871	0.0497	0.0181	1.2509	0.0135	0.0034
ν	0.6319	0.0636	0.1212	0.5977	0.0294	0.0336
τ	1.3395	0.0543	0.0846	1.3101	0.0249	0.0341
	$n = 360$			$n = 1.000$		
Parameter	AEs	Bias	MSEs	AEs	Bias	MSEs
β_{10}	1.4309	-0.0012	0.0001	1.4318	-0.0003	<0.0001
β_{11}	-1.1963	<0.0001	<0.0001	-1.1963	<0.0001	<0.0001
β_{12}	0.1671	-0.0002	<0.0001	0.1673	<0.0001	<0.0001
β_{13}	0.9466	-0.0001	<0.0001	0.9468	0.0001	<0.0001
β_{14}	0.2313	-0.0004	<0.0001	0.2316	-0.0001	<0.0001
β_{20}	-2.3148	-0.0254	0.0256	-2.3040	-0.0146	0.0082
β_{21}	0.1695	-0.0031	0.0044	0.1715	-0.0011	0.0014
β_{22}	0.8205	0.0060	0.0012	0.8181	0.0036	0.0004
β_{23}	0.2278	0.0125	0.0113	0.2168	0.0015	0.0041
β_{24}	1.2459	0.0085	0.0015	1.2414	0.0040	0.0004
ν	0.5801	0.0118	0.0110	0.5688	0.0005	0.0032
τ	1.2898	0.0046	0.0158	1.2802	-0.0050	0.0051

5.3. Residual analysis

The quantile residuals (qrs) $qr_i = qr_i(y_i, \hat{\boldsymbol{\theta}})$ for the NWIG multiple regression model are defined in Dunn and Smyth (1996)

$$qr_i = \Phi^{-1}\left(\exp\left[-\hat{\nu}\left\{-\log\left[\widehat{\Psi}(y_i)\right]\right\}^{\hat{\tau}}\right]\right),$$

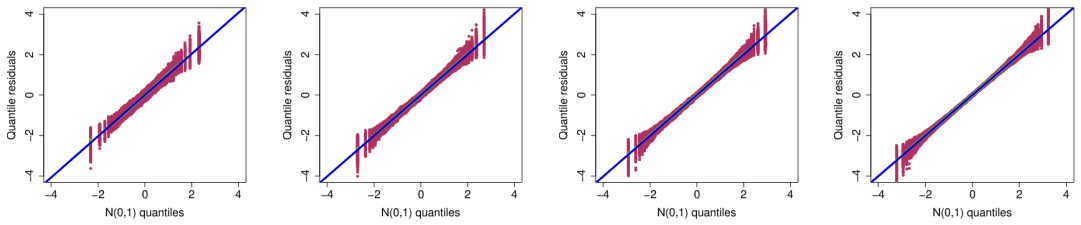
where $\Phi^{-1}(\cdot)$ is the standard normal qf, and

$$\widehat{\Psi}(y_i) = \Phi\left(\sqrt{\frac{1}{\hat{\sigma}_i{}^2 y}}\left(\frac{y_i}{\hat{\mu}_i} - 1\right)\right) + \exp\left(\frac{2}{\hat{\sigma}_i{}^2\hat{\mu}_i}\right)\Phi\left(-\sqrt{\frac{1}{\hat{\sigma}_i{}^2 y_i}}\left(\frac{y_i}{\hat{\mu}_i} + 1\right)\right).$$

Samples of sizes 60, 180, 360 and 1.000 are generated from the configuration described in Section 5.2. The plots in Figure 3 show that the distribution of the qrs approaches the standard normal distribution when n increases.

6. Application to Soybean data

We describe an application to show the utility of the new distribution. The data are obtained from two soybean storage and processing units belonging to an agricultural cooperative located in the interior of São Paulo state, selected in function of their differences in infrastructure, receiving procedures and types of clients (cooperative and non-cooperative members).



(a) (b) (c) (d)

Figure 3: Normal probability plots for the qrs. (a) $n = 60$; (b) $n = 180$; (c) 360; (d) 1.000.

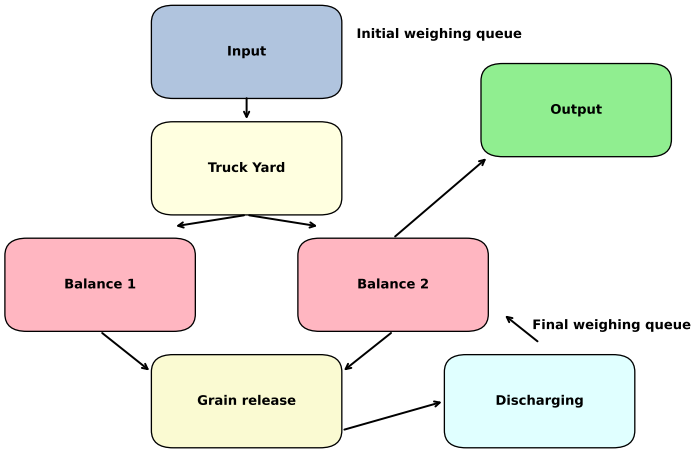
The study is conducted in the field, with the help of the workers at both units, say A and B. These units contain the following structures for the services of processing and storage of soybeans: drier, cleaning machine, silos, scales, hoppers and redlers. The total loiter time is obtained from the records of the entry and exit times of the trucks, where the difference represents the total time spent for receiving or unloading. These times are recorded in the system for the control and security of internal movement. We monitor the services and collected data for one week at each unit, during the peak soybean harvest season, from March 9 to 14, 2020 and March 16 to 21, 2022, respectively, for units A and B. We note all steps involved in receiving the recently harvested soybeans, namely: initial weighing of the truck, waiting time for analysis of a soybean sample, unloading, and final weighing. The movement of soybeans was monitored in loco, with identification of each truck at the moment of arrival of each unit after passing through the front gate, regardless of whether or not the load belonged to a cooperative member.

The variables are defined as follows:

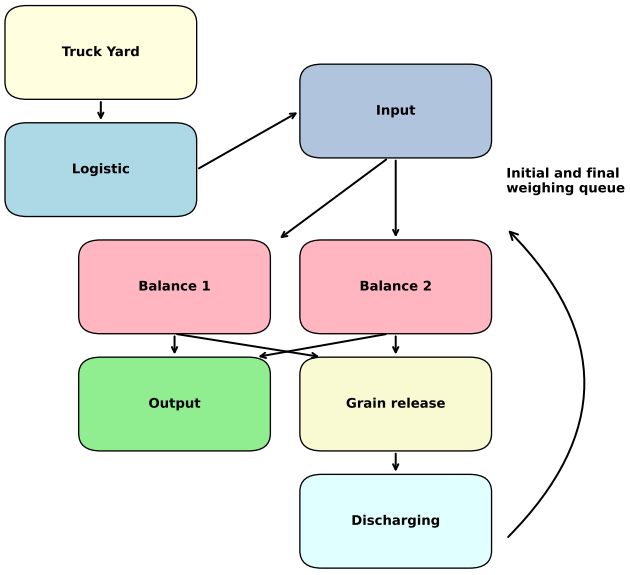
- y_i: total time (in hours) the trucks stay in the soybean processing units, and
- x_{i1}: an explanatory variable that identifies the unit where the truck was serviced (A or B) for $i = 1, \ldots, 340$. Specifically, x_{i1} takes the value A if the truck is from unit A and B if it is from unit B. This variable allows for a clear distinction between the two processing units in the analysis.

6.1. Description of the logistics of receiving soybeans at the processing units

- First, the following data are recorded upon arrival of each truck by the gatekeeper in spreadsheets: time of arrival, time of departure, license plate, driver's name, activity that will be carried out by the cooperative and name of the company that owns the load. The yard of this unit is located within its premises, where the trucks wait in a line for weighing at the scales. An employee is responsible for organizing the order of services and directing the truck to the available scale (number 1 or 2). Upon entry, in the case of unloading/reception, the truck goes to the available scale (1 or 2) for initial weighing, and at the same time a sample of soybeans is collected for analysis of the quality parameters, the documents are checked, and the data are inserted in the system. Then, the driver waits for release of the truck from the scale and takes the load to the hopper area for unloading. After these operations, the truck goes to scale 2 for final weighing, to determine the tare weight for calculating the weight of the load, and a document is issued recording the weight of the soybeans delivered and information on quality parameters. Then, the truck leaves the premises. Figure 4 (a) presents the flow chart based on observation of the processes of unloading (reception) and quality sampling of unit A.



(a)



(b)

Figure 4: Flow chart of the service observed in unit A (a) and unit A (b)

(a) (b) (c)

Figure 5: Characterization of weighing steps in soybean processing units

- Second, the waiting area is located outside the premises of the processing operations, together with the logistics (pre-weighing) sector, with the function of organizing the waiting line for entry. During this wait, the data are inserted in the system (identity the truck and driver and other data from checking the documentation accompanying the load). These facilitate tun the initial weighing operation. During the weighing, a soybean sample is taken to determine the quality parameters, after which the truck is released for unloading in the hopper, and then returns for final weighing to determine the weight of the load, for issuance of the corresponding document, including the grain quality analysis data. Finally, the truck leaves the premises. Figure 4 (b) presents the flow chart based on observation of the processes of unloading (reception) and quality sampling of unit B.

Table 3 addresses the main differences between units A and B.

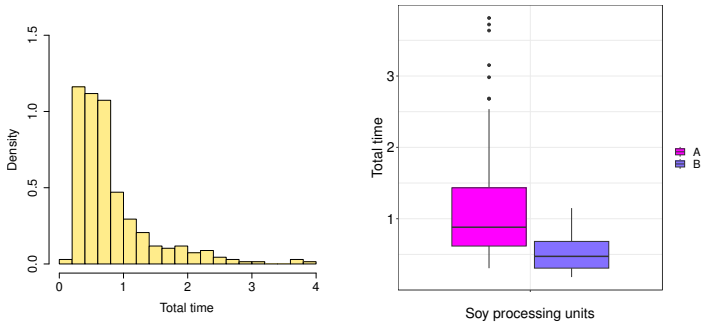
Figure 5 reports some images of the grain processing units, thus showing the steps of weighing (a); movement to the hopper (b); and the truck-trailer rigs and dump trucks unloading in the hopper (c).

Table 3: Differences between protocols implemented in soybean processing units

UNIT A	UNIT B
• It serves farmers from various regions, who generally are not cooperative members, and contract transport services from third parties using truck-trailer rigs. • At the gate, the vehicles' entry data are recorded in the system before they enter the yard, which is located within the premises, serving only for waiting for services. • At unit A, there are no planned logistics, which allows accumulation of loads near the unit. Besides this, the unit's layout allows prioritizing scale 1 for final weighing, thus improving the entry and exit flow of trucks.	• It mainly serves cooperative members and producers from the nearby region. They generally have their own fleet of trucks, mainly dump trucks. • The waiting yard is located before the gate, and has a logistics sector with the function of verifying the documentation and including the truck in the waiting line before it enters the premises, facilitating the weighing operation. • Due to the predominance of dump trucks among those that access its external yard, the loaded trailers from the farms are detached, freeing the driver to return in just the cab to haul new loaded trailers. This does not make much sense to me. A dump truck is generally a single unit, while a truck-trailer rig has a cab that pulls one or two trailers, which can be detached. I don't understand how the predominance of dump trucks would facilitate the service to truck-trailer rigs. Furthermore, if empty trailers are left behind, how are they subsequently used, or returned to the origin? Therefore, logically there would seem to be no difference in the treatment of the types of trucks, except that dump trucks would be easier to unload.

6.2. Descriptive analysis

Figure 6 (a) and Table 4 show data referring to the total time (in hours) the trucks spent in the soy processing units. Note that the data have asymmetry to the right. In addition, there is great variability in the data, however, 50% of the observations indicated that the length of stay of the trucks was less than or equal to 0.649. In Figure 6 (b) indicates that the reception/unloading service time is longer in unit A, this may be related to the fact that in this unit, there is still no logistics sector that inserts the truck driver in the queue and organizes the flow. In this unit, it was observed that some trucks remained for more than three hours at the site (possible atypical points). It is worth mentioning that these values were not removed from the base because they were results from the unit's logistic structure type and it is of practical interest to understand the factors that can motivate the slowness of truck traffic in the unloading process, in the two analyzed units.



(a) (b)

Figure 6: Histogram (a) and Boxplot (b) of the total time of permanence of trucks in soybean processing units

Table 4 summarizes the main descriptive statistics. It can be seen that the mean value is higher than the median, and skewness is positive indicating a positive asymmetric distribution, thus the NWIG distribution is a candidate for modeling this dataset.

Table 4: Descriptive statistics of the total time (in hours) the trucks stay in the soy processing units

n	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Min.	Max.
340	0.8093	0.6497	0.2536	0.6113	0.3737	2.0912	5.2637	0.1825	3.8133

6.3. Marginal analysis

The NWIG distribution is compared with the non-nested distributions: beta Weibull (BW) Cordeiro, Nadarajah, and Ortega (2013) and Kumaraswamy Weibull (KW) Nadarajah, Cordeiro, and Ortega (2012).

The Table 5 provides some useful statistics for assessing the quality of fit, including the A^*(Anderson-Darling) and W^*(Cramér-von Mises) statistics and the KS (Kolmogorov-Smirnov) test. There are strong indications that the NWIG distribution is better suited to model these data.

Table 5: Adequacy measures

Model	A^*	W^*	KS
NWIG	**0.5877**	**0.0931**	**0.0500**
KW	0.9785	0.1372	0.0614
BW	1.0677	0.1414	0.0709
IG	1.0657	0.1456	0.0609

Table 6 reports the MLEs and their standard errors (SEs) (in parentheses), and the Akaike Information Criterion (AIC) for four fitted distributions, which reveal that the new distribution is the best model to the current data.

Table 6: Findings from the marginal analysis

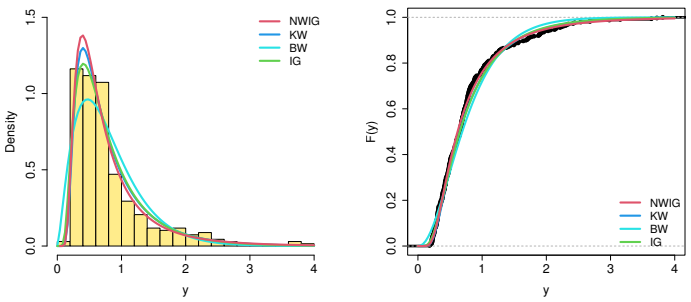
Model	$\log(\mu)$	$\log(\sigma)$	ν	τ	AIC
NWIG	**0.1858**	**-0.1406**	**0.6666**	**1.4477**	**361.2483**
	(0.0544)	**(0.0301)**	**(0.0374)**	**(0.0602)**	
KW	2.5307	25.4640	0.2210	-0.1253	363.3992
	(0.0189)	(3.1360)	(0.0113)	(0.0146)	
BW	-2.7039	2.3541	46.4708	0.0227	423.0491
	(0.0290)	(0.0854)	(0.0141)	(0.0122)	
IG	-0.2115	-0.2440	(-)	(-)	365.6633
	(0.0382)	(0.0384)	(-)	(-)	

The likelihood ratio (LR) test for the extra parameters of the new distribution in Table 7 also supports it.

Table 7: LR tests

Models	Hypotheses	LR statistic	p-value
NWIG vs IG	$H_0 : \nu = 1$ and $\tau = 1$ vs $H_1 : H_0$ is false	8.4149	0.0149

The estimated densities over the histogram, their cdfs and the empirical cdf (black curve) plotted in Figure 7 confirm our findings.



(a) (b)

Figure 7: (a) Fitted densities to the length of stay of trucks inside the cereal processing units. (b) Estimated cdfs and the empirical distribution of the response variable.

6.4. NWIG multiple regression model

Consider the NWIG multiple regression model with the systematic components

$$\mu_i = \exp(\beta_{10} + \beta_{11} x_{i1}); \quad \sigma_i = \exp(\beta_{20} + \beta_{21} x_{i1}), \quad i = 1, \ldots, 340.$$

The values of the AIC for some fitted models to the current data in Table 8 prove that the new multiple regression model is the best one for these data.

Table 8: AIC values

Regression	AIC
NWIG	**217.2485**
IG	222.2446
BW	250.6870
KW	268.2344

The LR test in Table 9 for comparing the NWIG and IG multiple regression models slso supports the wider model.

Table 9: LR test

Regressions	Hypotheses	LR statistic	p-value
NWIG vs IG	$H_0 : \nu = 1$ and $\tau = 1$ vs $H_1 : H_0$ is false	8.9961	0.0111

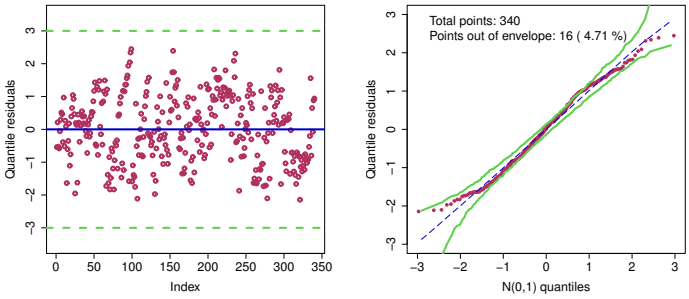
The estimates, SEs, p-values and confidence intervals in Table 10 reveal that, both in μ and σ, the soybean processing units has significant differences, and according to Figure 6, the total time of permanence of the trucks in the units, during the unloading/receiving process, is smaller in unit B and the data is less dispersed, indicating that the trucks' permanence time in this unit are close. Note that in unit A the opposite occurs, this result may be related to the structural conditions of each unit.

Table 10: Findings from the fitted NWIG regression

Parameter	Estimate	SE	p-value	Confidence intervals
$\hat{\beta}_{10}$	1.4321	0.1404	<0.0001	[1.1567; 1.7071]
$\hat{\beta}_{11}$	-1.1963	0.1264	<0.0001	[-1.4438; -0.9486]
$\hat{\beta}_{20}$	-2.2894	0.2388	<0.0001	[-2.5644; -2.0139]
$\hat{\beta}_{21}$	0.1726	0.0534	0.0014	[-0.0750; 0.4201]
ν	0.0071	0.0018	<0.001	[-0.2681; 0.2823]
τ	1.2852	0.1211	<0.001	[1.0098; 1.5603]

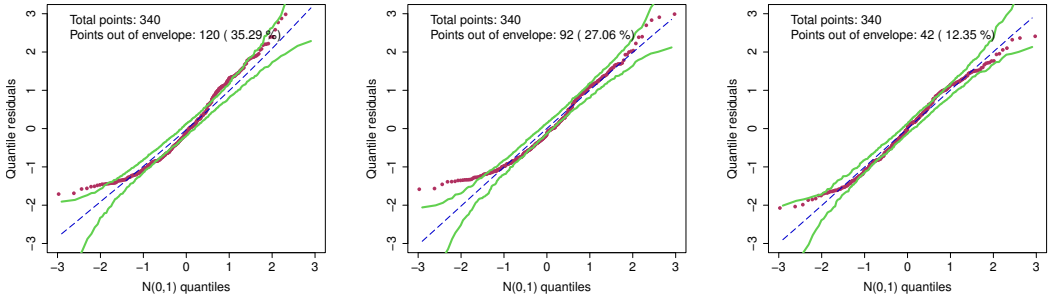
Further, the residuals are randomly distributed around zero for the fitted NWIG regression as shown in Figure 8(a). The normal plot for the residuals with simulated envelope in Figure 8(b) confirms that this regression is adequate for the soybean data.

In Figures 9(a), 9(b) and 9(c) we present the graph of the normal probability plot of the qrs with simulated envelope referring to the regression models BW, KW and IG, respectively. From Figures 8 and 9 we can see that the NWIG regression is the most suitable for modeling this data set. The additional parameters ν and τ of asymmetry and shape, respectively, make the model more flexible and can capture important characteristics of the data. The importance of these parameters for soybean data lies in the model's ability to accurately fit real data, allowing for more reliable decision making. Furthermore, the introduction of these parameters is not limited to soybean data, increasing the generality and usefulness of the proposed distribution in various contexts.



(a) (b)

Figure 8: (a) Qrs for the NWIG regression. (b) Normal probability plot of the qrs with simulated envelope for the NWIG regression.



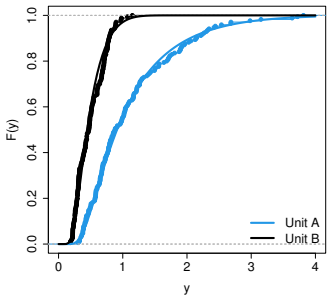
(a) (b) (c)

Figure 9: Normal probability plot of the qrs with simulated envelope. (a)BW regression model; (b) KW regression model; (c) IG regression model.

Figure 10 reports the estimated cdfs from the fitted NWIG multiple regression model, and the empirical cdf. Again, we note that the collecting storage units differ significantly.

The NWIG distribution is particularly relevant to the application of soybean data in this context because soybean production involves complex logistics, including transportation, processing and storage, especially when grain producers in Brazil often rely on third-party services for these activities, due to the lack of silos on their properties. In scenarios like this, where multiple processes are involved and several factors can affect the efficiency of operations, having a distribution like NWIG, which can capture asymmetry and heavy tails, is essential. This flexibility allows the model to better represent data complexities, such as variable service times and other logistical variables, which are essential for optimizing process management and improving reliability between customers and service providers in the grain processing industry. Therefore, the NWIG distribution is especially suitable for analyzing the efficiency of internal vehicle flow in grain processing units, as it can provide a better understanding of how service times depend on logistical variables in this sector.

In general, we note some important differences between units A and B. For example, unit B mainly serves cooperative members and farmers from the nearby region, who generally have their own truck fleet. Besides this, the waiting yard is located before the gate and has a “pre-weighing” or “logistics” sector with the function of checking the documentation accompanying the load and inserting the truck in the waiting line before it enters the premises, thus facilitating the weighing operation. In turn, unit A serves producers from a wider range of regions, who generally are not cooperative members and who mainly contract the services of third parties to haul soybeans from the farm to the unit. The gatekeeper records the entry and exit of vehicles and the yard is located inside the premises, serving only for waiting.



(a)

Figure 10: Estimated cdfs from the fitted regression for units A and B

Therefore, we can conclude that unit B has a better logistics structure, so that the steps of receiving, initial weighing, unloading, final weighing and release generally take less time. Good logistics planning allows optimizing the management of the storage and processing within the units, improving the trust between the client and service provider. The initial computerized logistics sector where documents are checked and trucks are inserted in the waiting line to unload in unit B generates better agility in the internal movement of trucks.

Filassi and de Oliveira (2021), Toloi *et al.* (2021) and Oliveira, Filassi, Lopes, and Marsola (2020) mentioned that the Brazilian logistics sector has not achieved significant advances in recent years, causing fragility both due to the lack of infrastructure for transport of products and inability to adequately store these products. Da da Motta Moreira, Coelho, de Jesus, Torres, and Ribeiro (2021) also described the logistical challenges facing supply chains to overcome limitations and other problems. Fliehr, Zimmer, and Smith (2019) noted that the lack of grain storage capacity, the excessive dependence on trucks, the poor highway conditions and inefficient terminal operations impede the adequate flow of grains from farms to consumers. The authors identified that these drawbacks reduce the prices received by farmers, decreasing their profitability and thus reducing national output of agricultural commodities. Seeking a solution for the storage sector and other logistical bottlenecks in the country, Filippi and Guarnieri (2019) stated that the so-called "Rural Storage Condominium" arrangement is a workable alternative to reduce the storage deficit and overcome other bottlenecks, as well as having advantages such as better strategic marketing of output, reduction of storage and logistics costs, and higher profits. In turn, de de Oliveira and Alvim (2017) reported that transport and storage logistics, required to adapt rapidly to meet the increasing demand for differentiated and segregated products.

7. Concluding remarks

We proposed a new Weibull inverse Gaussian (NWIG) regression to explain the total time that the trucks spend at the cereal processing units. We determined the maximum likelihood estimates of the regression coefficients, and investigate their accuracy by Monte Carlo simulations. We defined the quantile residuals, and showed that its empirical distribution is close to the standard normal distribution.

The new regression is very useful to analyze data of this nature and may provide more realistic fits than other regressions, since a great variability was observed due to the service processes, the computerized logistics sector implemented in unit B generated greater agility in internal traffic of the trucks, having, however, smaller impacts on the unloading operation. The results indicated that unit B presents a logical planning with shorter permanence time of the trucks, thus indicating greater agility in the internal traffic. In unit A, which serves almost all bulk trucks, it would be ideal to install a tipper in the hopper, which would result in improved unloading time and implementation the logistics sector to reduce weighing

time. In this way, it is possible that logistics can optimize the management of processes that occur inside collecting storage units, increasing the reliability between the customer and the provider of the service. Based on the application, it was empirically shown that the proposed regression based on the NWIG distribution is superior in relation to the BW, KW and IG regression models. However, this fact cannot be generalized because for other data sets this superiority may not happen. It is concluded that readers can use this new regression for other datasets in the presence of strong asymmetry. Future research using the new NWIG distribution can be done towards studying other estimation methods such as Bayesian and Jackniffe Methodology. Regarding the NWIG regression models, diagnostic analysis can be proposed to study the robustness of the regression model in relation to possible influential points and propose regression models with random effect and semiparametrics based on the NWIG distribution. Finally, the proposed regression model can be extended to the bivariate case.

Acknowledgment

This work was supported by CNPq and CAPES, Brazil.

References

Branco JEH, Branco DH, de Aguiar EM, Filho JVC, Rodrigues L (2019). "Study of Optimal Locations for New Sugarcane Mills in Brazil: Application of a MINLP Network Equilibrium Model." *Biomass and Bioenergy*, **127**, 105249.

Cordeiro GM, Nadarajah S, Ortega EMM (2013). "General Results for the Beta Weibull Distribution." *Journal of Statistical Computation and Simulation*, **83**, 1082–1114.

Cordeiro GM, Vasconcelos JCS, dos Santos DP, Ortega EM, Sermarini RA (2022). "Three Mixed-Effects Regression Models Using an Extended Weibull with Applications on Games in Differential and Integral Calculus." *Brazilian Journal of Probability and Statistics*, **36**, 751–770.

da Motta Moreira SG, Coelho AW, de Jesus ER, Torres EG, Ribeiro TF (2021). "Proposta para Reduzir o Tempo de Descarregamento das Cargas Manuais." *Inova+ Cadernos da Graduação da Faculdade da Indústria*, **2**.

de Holanda GG, da Silva AF, de Lavor NB, de Sousa FNT (2020). "Custos Logísticos do Transporte no Modal Rodoviário: Desafios para a Competitividade das Empresas/Logistic Costs of Transportation on the Road Mode: Challenges for Competitiveness of Companies." *Id on Line Revista Multidisciplinar e de Psicologia*, **14**, 570–585.

de Oliveira ALR, Alvim AM (2017). "The Supply Chain of Brazilian Maize and Soybeans: The Effects of Segregation on Logistics and Competitiveness." *International Food and Agribusiness Management Review*.

de Oliveira ALR, Marsola KB, Milanez AP, Fatoretto SLR (2022). "Performance Evaluation of Agricultural Commodity Logistics from a Sustainability Perspective." *Case Studies on Transport Policy*, **10**, 674–685.

de Souza ALA (2020). "Escoamento de Commodities Agrícolas Brasileiras." *Agricultura 4.0*.

do Nascimento VVF, Pereira GFX, Melo CCV, Freire AI, de Souza FBM, Martins AD (2022). "Análise Logística na Produção de Grãos no Brasil." *Research, Society and Development*, **11**, e47911730597–e47911730597.

Dunn PK, Smyth GK (1996). "Randomized Quantile Residuals." *Journal of Computational and Graphical Statistics*, **5**, 236–244.

Filassi M, de Oliveira ALR (2021). "Competitiveness Drivers for Soybean Exportation and the Fundamental Role of the Supply Chain." *Revista de Economia e Sociologia Rural*, **60**.

Filippi ACG, Guarnieri P (2019). "New Forms of Rural Organization: The Rural Warehouse Condominiums." *Revista de Economia e Sociologia Rural*, **57**, 270–287.

Fliehr O, Zimmer Y, Smith LH (2019). "Impacts of Transportation and Logistics on Brazilian Soybean Prices and Exports." *Transportation Journal*, **58**, 65–77.

Geyer CJ (2006). "Stat 5101 Notes: Expectation." Avaliable at https://www.stat.umn.edu/geyer/old06/5101/notes/n1.pdf.

Johnson NL, Kotz S, Balakrishnan N (1994). *Continuous Univariate Distributions*, volume 1. 2nd edition. Wiley, New York.

Karges K, Bellingrath-Kimura SD, Watson CA, Stoddard FL, Halwani M, Reckling M (2022). "Agro-Economic Prospects for Expanding Soybean Production beyond Its Current Northerly Limit in Europe." *European Journal of Agronomy*, **133**, 126415.

Li NC, Govindan GK, Jin Z (2018). "Disruption Management for Truck Appointment System at a Container Terminal: A Green Initiative." *Transportation Research Part D: Transport and Environment*, **61**, 261–273.

Lu MZ, Carter AM, Tegeder M (2022). "Altering Ureide Transport in Nodulated Soybean Results in Whole-Plant Adjustments of Metabolism, Assimilate Partitioning, and Sink Strength." *Journal of Plant Physiology*, **269**.

Melo IC, Péra TG, Júnior PNA, do Nascimento Rebelatto DA, Caixeta-Filho JV (2020). "Framework for Logistics Performance Index Construction Using DEA: An Application for Soybean Haulage in Brazil." *Transportation Research Procedia*, **48**, 3090–3106.

Nadarajah S, Cordeiro GM, Ortega EMM (2012). "General Results for the Kumaraswamy-G Distribution." *Journal of Statistical Computation and Simulation*, **82**, 951–979.

Neagoe M, Hvolby HH, Taskhiri MS, Turner P (2021). "Using Discrete-Event Simulation to Compare Congestion Management Initiatives at a Port Terminal." *Simulation Modelling Practice and Theory*, **112**, 102362.

Oliveira ALRD, Filassi M, Lopes BFR, Marsola KB (2020). "Logistical Transportation Routes Optimization for Brazilian Soybean: An Application of the Origin-Destination Matrix." *Ciência Rural*, **51**.

Peoples MB, Giller KE, Jensen ES, Herridge DF (2021). "Quantifying Country-to-Global Scale Nitrogen Fixation for Grain Legumes: I. Reliance on Nitrogen Fixation of Soybean, Groundnut and Pulses." *Plant and Soil*, pp. 1–14.

Pimpanit P, Jarumaneeroj P (2021). "A Discrete Event Simulation Model for Evaluating Inland Terminal's Efficiency: A Case Study of Ladkrabang Inland Container Depot." In *2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA)*, pp. 627–631. IEEE.

Rigby RA, Stasinopoulos DM (2005). "Generalized Additive Models for Location, Scale and Shape." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507–554.

Rotunno G, Zupone GL, Carnimeo L, Fanti MP (2023). "Discrete Event Simulation as a Decision Tool: A Cost Benefit Analysis Case Study." *Journal of Simulation*, pp. 1–17.

Stasinopoulos DM, Rigby RA (2007). "Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R." *Journal of Statistical Software*, **23**, 1–46.

Stasinopoulos M, Rigby B, Akantziliotou C (2008). "Instructions on How to Use the GAMLSS Package in R."

Toloi RC, dos Reis JGM, Abraham ER, Toloi MNV, Bueno RE (2021). "Fatores de Decisão e Qualidade na Rede de Suprimentos da Soja de Mato Grosso." *Agrarian*, **12**, 248–260.

Ulrich G, Watson LT (1994). "Positivity Conditions for Quartic Polynomials." *SIAM Journal on Scientific Computing*, **15**, 528–544.

Weibull W (1939). "The Statistical Theory of the Strength of Materials." *Ingeniors Vetenskaps Academy Handlingar (151). Stockholm: Generalstabens Litografiska Anstalts Förlag*, pp. 1–45.

Yi S, Scholz-Reiter B, Kim T, Kim KH (2019). "Scheduling Appointments for Container Truck Arrivals Considering Their Effects on Congestion." *Flexible Services and Manufacturing Journal*, **31**, 730–762.

Zhang X, Zeng Q, Yang Z (2019). "Optimization of Truck Appointments in Container Terminals." *Maritime Economics & Logistics*, **21**, 125–145.

Affiliation:

Julio Cezar S. Vasconcelos
Instituto de Ciência e Tecnologia
Universidade Federal de São Paulo
Av. Cesare Monsueto Giulio Lattes, São José dos Campos-SP, 12247-014, Brazil
E-mail: juliocezarvasconcelos@hotmail.com