An Extreme Value Regression Approach to the Prediction of Shrimp Size Distribution

Rafael Perez Abreu Carrion  
*Centro de Investigación en Matemáticas*

Hector de la Torre  
*Centro de Investigación en Matemáticas*

Samantha Estrada  
*The University of Texas at Tyler*

---

Recommended Citation

An Extreme Value Regression Approach to the Prediction of Shrimp Size Distribution

Rafael Perez Abreu Carrion
Centro de Investigación en Matemáticas
Hector de la Torre
Centro de Investigación en Matemáticas
Samantha Estrada
The University of Texas at Tyler

A statistical non-normal regression model was developed to characterize the growth of white shrimp (Litopenaeus vannamei) and the shrimp’s weight distributions throughout the fattening process. Empirical data was collected from submersible cages in the Gulf of California, Mexico. In this article, the authors demonstrate the efficiency in predicting the shrimp size distribution by using the extreme value regression methodology. The extreme value regression has been used in quality control engineering, reliability and survival analysis; however, it has yet to be applied in aquaculture setting. Findings suggest that the extreme value regression was the best model to fit the weight of shrimp. The extreme value regression model can be used to predict not only the average weight but also the shrimp size distribution and the percentiles as a function of the number of days that the shrimp stay in the farm.

Keywords: Shrimp, Non-Normal Regression, Extreme Value Regression, Submersibles Sea Cages, growth, Litopenaeus vannamei.

Introduction

In terms of economic value, shrimp farming is one of the important fishery resources in Mexico (Pérez-Castañeda et al., 2015). In Mexico, shrimp farming is valued at 11,479 million pesos (MXN; roughly, 575, 600 USD; Comision Nacional de Acuacultura y Pesca, 2018). Thus, shrimp aquaculture production has grown in recent years (Pérez-Castañeda et al., 2015). Mathematical models in aquaculture have become an essential tool for estimating and computing the growth of aquaculture species (Araneda, Hernández, Gasca-Leyva, and Vela, 2013; Franco, Ferreira, & Nobre, 2006). The growth modeling methodology allows researchers to accumulate observations throughout a culture period. Utilizing this empirical data, researchers are able to analyze growth variables to maximize production, aiding farmers in decision-making, and calculating cost-benefit analyses (Araneda et al., 2013; Yu, Leung, and Bienfang, 2006). Due to the economic importance of shrimp, not only in Mexico but worldwide, research has focused on developing these growth models for shrimp in recent years (Franco et al., 2006).
Functional Forms Used to Model the Growth of Shrimp

A prediction of shrimp growth is fundamental to make an adequate decision for farmers and businesses who are rearing shrimp (Tian et al., 1993). In the aquaculture literature, there exist several functional forms to use as a growth model for shrimp and other species; For example, there are studies focusing on temperature and feed, or the effect of salinity and temperature on shrimp growth (Ponce-Palafox, Martinez-Palacios, & Ross, 1997; Wyban, Walsh & Godin, 1995) as well as studies using classical statistical methods to model shrimp growth (Carvajal & Nebot, 1998). Typically, quantitative methods model the average weight through statistical regression techniques using several functional forms. For example, Esmaeili and Tarazkar (2011) compared linear and non-linear regression models to estimate growth focusing on artificial neural networks (ANNs) to predict shrimp growth. Similarly, Yu et al. compared eight regression functional forms to model the shrimp average weight: Linear, Polynomial, Log reciprocal, von Bertalanffy, Gompertz, Logistic, Exponential, and ANN (Yu et al., 2006). The authors concluded that the ANN method delivered the most accurate prediction out of the eight models. Additionally, in aquaculture research, Katsanevakis performed a similar model selection study for modeling fish growth (2006).

Shrimp weight-marketing price depends on the shrimp size, not only average weight prediction is required, but also the distribution of the weight and size (Instituto Nacional de Pesca, 2018). In this sense, few researchers have focused on both predicting shrimp average growth and modeling the size distribution. However, research has shown that size heterogeneity is produced due to size-dependent variables that affect the growth and level of overcapacity which can increase, decrease or stabilize heterogeneity (Peacor, Bence & Pfister, 2007). For example, Araneda et al., (2013) studied modeling the shrimp growth considering the size variability of white shrimp in freshwater. The authors state: “to the extent of the authors’ knowledge, there are at present no studies at the intensive level where heterogeneity is studied when modeling growth in aquaculture” (p. 1, Araneda et al., 2013). The authors named the size distribution as "size heterogeneity," and presented three functional forms to model average with size heterogeneity: Gompertz, von Bertalanffy and Pütter. The researchers concluded that utilizing size heterogeneity produced better statistical results than when it is not considered as part of the model.

Non-normal Regression Models

Because the distribution of shrimp weight does not follow a normal distribution, it is necessary to try modelling the shrimp size with other alternative distributions. The logarithm and the Box-Cox transformations do not work well with these data. Another approach would be to use non-normal regression models such as the extreme value regression. Using these kinds of models is common in quality control engineering and survival analysis (Crowder, Kimber, Sweeting & Smith, 1994; Meeker & Escobar, 2014). In fact, researchers have applied extreme value regression to model multiple levels of stress in life-testing (Ng, Balakrishnan, Chan, 2007), predicting bank failures (Calabrese & Giudici, 2015), and predicting a snow
AN EXTREME VALUE REGRESSION APPROACH TO THE PREDICTION OF SHRIMP SIZE DISTRIBUTION

avalanche runout (McClung & Mears, 1991). Non-normal regression models include the lognormal regression, Weibull regression, and extreme value regression. These models are useful to model the scale parameter of non-normal distributions as a function of one or more variables. Typically, the dependent variable is time and the independent variable can be the load or the temperature. In these methodologies the focus is set into the percentile’s statistics more than in the mean of the distribution.

The goal of these models is to have a statistical equation to estimate the shrimp size and weight as a function of time. Time, which is the dependent variable is measured in days. Similarly, as linear regression, the weight distribution mean is modeled as a linear or transformation function of the independent variables. These regression methods have been used to solve practical problems in lifetime. For example, Meeker and Escobar (2014) explain how to model the time that a computer program spends executing (in a multitask system) as a function of the load (i.e. the number of users). In Meeker and Escobar’s example, the regression model is not a normal regression model. In the same manner, for this study, the shrimp weight depends on the number of days in the farm.

The initial step is to identify what kind of distribution follows the variable of interest (or response variable) for each level of the explanatory variable. Then, the location parameter and scale (or shape) parameter are fitted separately for each level. Finally, a joint model is estimated in which the scaled parameter σ is limited to be equal for all levels of the explanatory variable and at the same time the location parameters are estimated. Next, the assumptions of the model must be assessed. Similarly, to a normal linear regression model, these assumptions are: y = weights are independent and follow a non-normal distribution with location parameter \( \mu(x) = \beta_0 + \beta_1 x \) and constant scale (or shape) parameter \( \sigma \). Note that only parameter \( \mu \) depends on \( x \). If variable Y follows a non-normal distribution, the residuals \( \varepsilon = Y - \hat{Y} \) also follow a non-normal distribution. Then, the analyses of residuals can be performed in order to evaluate these assumptions (Crowder et al., 1994; Meeker & Escobar, 2014). For example, if the assumption is that values follow extreme value distribution (also known as Gumbel distribution), it can be evaluated by graphical or analytical methods. For these graphical methods, an extreme value probability plot is useful. If residuals follow an extreme value distribution, a straight line of residuals plot should be observed. Analytical methods of goodness-of-fit tests such as Kolmogorov-Smirnov, Cramer-von Mises or Anderson-Darling are valuable to this process.

**Purpose and Rationale**

In this study, farmers breed white shrimp (Litopenaeus vannamei) inside of submersible sea cages (SC) in the Gulf of California, Mexico. White shrimp post-larvae (PL) were seeded at the beginning of June 2010 (time = 0) and were harvested on October 25 (time = 135 days). Scuba divers feed the shrimp inside sea cages each day. In order to monitor the shrimp growth in this period of fattening, a sample was taken regularly. For each sample, biometric measurements were performed to obtain individual information about the weight, size, and other shrimp quality characteristics. However, in this study it was not enough to know the average weights
of shrimp, because the shrimp market prices depend on the shrimp size, it is also important to know the statistical distribution of shrimp size.

A balanced harvest point is desirable between feeding shrimp and the final harvest to commercialize the shrimp. For farmers it is important to know the shrimp length or size as the final product, because these factors determine the project feasibility, marketing conditions, their final price in market, and return on investment of this new form of shrimp production. Farmers lacked information on two main issues: (a) to know the optimal date to harvest shrimp, and (b) to know if they need to continue feeding shrimp with the possibility that it may stop growing after a few days. The uncertainty of this information can cause cost-effectiveness problems for farmers. To resolve this issue, farmers can predict shrimp growth utilizing a statistical model. Thus, the goal was modeling the size through non-linear models to these data; specifically, to model the distribution of the weight of shrimp at different times before and after the harvest. In this work, the authors propose to apply a non-normal regression model such as the extreme value distribution to a shrimp growth model. Past research has focused on linear, non-linear regression models, artificial neural networks (ANNs) but no research has explored the extreme value distribution (Esmaeili & Tarazkar, 2011; Yu et al., 2006). Thus, this study will be addressing a gap in the literature focusing on the performance of the extreme value distribution in modeling shrimp growth.

**Methodology**

**Sampling**

The samples were taken during the fattening period and the sample size was different for each sample due to the sampling method. To select the sample, scuba divers dropped a small net from the top of the submersible sea cages when the net was extended over the cage’s bottom, then the scuba divers pulled the net selecting the shrimp. The sampling days were selected according to the weather conditions and resources available. Note, the sampling time intervals were not equal.

**Variables**

The variables considered in this work are the following:

The independent variable or explanatory variable is the number of days since larva was seeded.

The dependent variable or response variable is the size of shrimp.

A simple view of this information can be seen in the scatter plot Days versus Weight in Figure 1.
The general idea of the regression model is to describe the shrimp size distribution as a function of the variable day(s). \( \Pr(Y \leq y; x) = F(y; x) = F(y) \), where \( Y \) is the shrimp weight at time \( x \) (in days). The mean of the distribution is \( \mu = \beta_0 + \beta_1 x \) with only one explainer variable.

The simplest distribution model for normal, logistic, and extreme value distributions is:

\[
\Pr(Y \leq y) = F(y; \mu, \sigma) = F(y; \beta_0, \beta_1, \sigma) = \Phi \left( \frac{y - \mu}{\sigma} \right)
\]

(1)

Where \( \Phi \) is the accumulate quantile of the normal distribution, \( \mu = \beta_0 + \beta_1 x \) and \( \sigma \) does not depend on \( x \) the explanatory variable. The quantile function for this model is:

\[
y_p = \mu + \Phi^{-1}(p)\sigma = \beta_0 + \beta_1 x + \Phi^{-1}(p)\sigma.
\]

(2)

The localization and scale parameters estimation of this regression model can be done in different manners. The most common is the maximum likelihood estimation (MLE), which consists of finding estimates for \( \beta_0, \beta_1 \) and \( \sigma \) such that the sample likelihood function is maximized. The likelihood function is given by:

\[
L(\beta_0, \beta_1, \sigma; x) = \prod_{i=1}^{n} L(\beta_0, \beta_1, \sigma; data_i) = \prod_{i=1}^{n} \left[ \frac{1}{\sigma} \phi \left( \frac{y_i - \mu_i}{\sigma} \right) \right]^{\delta_i} \left[ 1 - \Phi \left( \frac{y_i - \mu_i}{\sigma} \right) \right]^{1-\delta_i}
\]

(3)

Numerical methods are used to maximize this function, in which the first derivative with respect to \( \beta_0, \beta_1, \sigma \) of the likelihood function \( L(\beta_0, \beta_1, \sigma; x) \) are set equal to zero and then the equations are solved for \( \beta_0, \beta_1, \sigma \). This procedure was executed in the Statistical Analysis Software (SAS) via PROC RELIABILITY and PROC LIFEREG. The simplest distribution model for Log-Localization-Scale (including log-normal, log-logistic and Weibull) is similar as previous, the only difference is that...
we use \( \log(y) \) instead of \( y \).

\[
\Pr(Y \leq y) = F(y; \mu, \sigma) = \Phi \left( \frac{\log(y) - \mu}{\sigma} \right)
\]

(4)

\[
L(\beta_0, \beta_1, \sigma; x) = \prod_{i=1}^{n} L(\beta_0, \beta_1, \sigma; \text{data}_i) = \prod_{i=1}^{n} \left[ \frac{1}{\sigma} \phi \left( \frac{\log(y_i) - \mu_i}{\sigma} \right) \right]^{\delta_i} \left[ 1 - \Phi \left( \log(y_i) - \mu_i \right) \right]^{1-\delta_i}
\]

(5)

Where \( \mu = \beta_0 + \beta_1 \cdot x \) and \( \sigma \) does not depend on \( x \) the independent variable. The quantile function for this model is:

\[
y_p = \mu + \Phi^{-1}(p)\sigma = \beta_0 + \beta_1 \cdot x + \Phi^{-1}(p)\sigma.
\]

(6)

which is linear in \( x \). This relationship is known as log-linear. Numerical methods are used to maximize this function, in which the first derivative respect to \( \beta_0, \beta_1, \sigma \) of the likelihood function \( L(\beta_0, \beta_1, \sigma; x) \) are set equal to zero and resolve for \( \beta_0, \beta_1, \sigma \). The quantile function can be used to calculate the weight percentiles for the shrimp size distribution for a specific value of \( x \) (days).

**Results**

**Model screening**

To identify the underlying distribution, several models were estimated. At first glance the relationship between shrimp weight gain and the intervening days seems to be linear. Several types of linear regressions were estimated for these data. That is, functional forms, like linear square root of the dependent variable, square root of the independent variable, exponential, logarithm s-curve, and x-square. For these models, the assumptions of normality were not met. For example, in the model square root of the independent variable (Equation 7) yielded one of the highest \( R^2 \) statistics; however, the analysis of residuals showed that residuals were not normally distributed. Table 1 shows the estimation of parameters for this model. Figure 2 and Figure 3 show the histogram and the normal probability plot for residuals of this model.

\[
\text{weight} = \beta_0 + \beta_1 \sqrt{\text{days}} + \epsilon
\]

(7)

**Table 1** Estimation of parameters for the square root of the independent variable model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-15.00</td>
<td>0.55</td>
<td>-16.09</td>
<td>-13.91</td>
</tr>
<tr>
<td>SqDays</td>
<td>3.36</td>
<td>0.06</td>
<td>3.24</td>
<td>3.47</td>
</tr>
<tr>
<td>Scale</td>
<td>3.38</td>
<td>0.09</td>
<td>3.21</td>
<td>3.56</td>
</tr>
</tbody>
</table>
As it can be seen in these two graphs, the residuals are far from following a normal distribution. Furthermore, all normality tests reject the null hypothesis that residuals follow a normal distribution at the 5% significance. Results of normality tests can be seen in Table 2.
If residuals do not follow an approximate normal distribution, then the follow-up questions are: What distribution do they follow? Is there a distribution that fits the residuals appropriately? In order to answer these questions, the authors performed a series of tests with different models. Residual plots were used to test for the normal distribution, lognormal distribution, the Weibull distribution, and the extreme value distribution. Figure 4 shows the probability plot for these four distributions. The extreme value distribution better fit the residuals (see Figure 4c). Aside of probability plots analytical tests were developed to verify which distribution model fits better to residuals. Tests of Goodness-of-fit failed to reject the null hypothesis, at the $\alpha=0.05$ level of significance, for the extreme value regression model, as it can be seen in Table 3. This table also shows the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The AIC and BIC also suggest the best model is the extreme value distribution. Therefore, we conclude shrimp weight can be modeled appropriately with an extreme value distribution.

### Table 2 Goodness-of-Fit Tests for Normal Distribution

<table>
<thead>
<tr>
<th>Normality Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>0.09</td>
</tr>
<tr>
<td>Cramer-von Mises</td>
<td>1.28</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>8.05</td>
</tr>
</tbody>
</table>

### Table 3 Goodness-of-fit Test for distributions

<table>
<thead>
<tr>
<th>Model</th>
<th>Log Likelihood</th>
<th>AIC</th>
<th>BIC</th>
<th>Kolmogorov-Smirnov</th>
<th>P-Value</th>
<th>Cramer-von Mises</th>
<th>W2</th>
<th>P-Value</th>
<th>Anderson-Darling A2</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>-1833.04</td>
<td>3672.1</td>
<td>3685.7</td>
<td>0.092</td>
<td>&lt; 0.0001</td>
<td>1.276</td>
<td>&lt; 0.0001</td>
<td>8.060</td>
<td>&lt; 0.0001</td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>-1815.90</td>
<td>3637.8</td>
<td>3651.4</td>
<td>0.095</td>
<td>&lt; 0.0001</td>
<td>1.361</td>
<td>&lt; 0.0090</td>
<td>7.124</td>
<td>&lt; 0.0090</td>
<td></td>
</tr>
<tr>
<td>LogNormal</td>
<td>-1891.67</td>
<td>3789.3</td>
<td>3803.0</td>
<td>0.100</td>
<td>&lt; 0.0001</td>
<td>1.497</td>
<td>&lt; 0.0090</td>
<td>8.442</td>
<td>&lt; 0.0090</td>
<td></td>
</tr>
<tr>
<td>Extreme Value</td>
<td>-1780.83</td>
<td>3567.7</td>
<td>3581.3</td>
<td>0.029</td>
<td>0.5998</td>
<td>0.089</td>
<td>&gt; 0.1000</td>
<td>0.551</td>
<td>&gt; 0.1000</td>
<td></td>
</tr>
</tbody>
</table>
AN EXTREME VALUE REGRESSION APPROACH TO THE PREDICTION OF SHRIMP SIZE DISTRIBUTION

Figure 4. From left to right counterclockwise: a) Log normal plot for standardized residuals, b) Weibull Probability plot for standardized residuals, c) Extreme Value Probability plot for standardized residuals, d) Normal Probability plot for standardized residuals.

To support the conclusion in a graphical manner, the histogram of standardized residuals and a line of the theoretical extreme value distribution are shown in Figure 5. The extreme value distribution line fits well to the shape of the histogram bars. Once the distribution of the residuals is found, the next step is estimating the extreme value linear regression.
The weight of shrimp for each sample (day) was analyzed to identify what kind of distribution follows them. In most cases, the model that best fits is the extreme value regression. Therefore, the location parameter $\mu = \beta_0 + \beta_1 x$ and $\sigma$ are estimated for each different day. Figure 6 shows the extreme value probability plots of the weight variable for each sample (day). The independent variable was transformed as the square root of day, due to the nature phenomena. It is assumed the shrimp will grow until the maximum limit for the species.

Next, a joint estimation for all days was executed; the estimate of the extreme value regression is shown in Table 4. The result model is:

$$\mu(x) = -14.7668 + 3.4940\sqrt{x}, \sigma = 2.7015$$ (8)
AN EXTREME VALUE REGRESSION APPROACH TO THE PREDICTION OF SHRIMP SIZE DISTRIBUTION

Table 4 Extreme value Regression Model $y = \sqrt{x}$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-14.77</td>
<td>0.42</td>
<td>-15.59</td>
<td>-13.95</td>
</tr>
<tr>
<td>$x = \sqrt{\text{days}}$</td>
<td>3.49</td>
<td>0.04</td>
<td>3.41</td>
<td>3.58</td>
</tr>
<tr>
<td>$\sigma = \text{EV Scale}$</td>
<td>2.70</td>
<td>0.08</td>
<td>2.55</td>
<td>2.86</td>
</tr>
</tbody>
</table>

An estimation of the data can be seen in the graph of the Figure 7. Note the slopes of the $\sigma$ estimates are the same. This is due to the constraint over $\sigma$.

![Extreme Value Probability Plot](image1)

![Extreme Value Percentiles](image2)

**Figure 7.** Extreme value Regression Model $y = \sqrt{x}$ assuming $\sigma$ same scale parameter for different days

Before this model can be used to predict the percentile of the shrimp weight, an analysis of residuals has to be performed. Results of goodness-of-fit tests can be seen in Table 5. Cramer-von Mises and Anderson-Darling test indicated that residuals follow an extreme value distribution. Also, an extreme value probability plot is shown in Figure 8. The residuals fit well in the center of the line and are misaligned on the left side of the distribution. Note that the location parameter is nearly zero and the scale parameter is equal to 1. This is equivalent to having exponential residuals follow an exponential distribution with parameter equal to 1.
Table 5 Extreme value regression model $y = \sqrt{x}$ residual assumptions

<table>
<thead>
<tr>
<th>Goodness-of-fit Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cramer-von Mises</td>
<td>0.089</td>
</tr>
<tr>
<td>Anderson-Darling</td>
<td>0.551</td>
</tr>
</tbody>
</table>

Figure 8. Extreme value probability plot

Now that the model has been evaluated, it can be used to estimate the size distribution of the shrimp. Table 6 contains the predicted weight of the 33 percentiles for several days 160, 170, 180, 190 and 200. This information is important for farmers for the reasons outlined in the literature review. For example, it is estimated for day 160 that 33% of shrimp will weigh less than 26.99 grams other 33% of the shrimp will weigh between 26.99 & 29.68g, and the other 33% of the shrimp will weigh between 29.68 and 35.43 grams. This information can be calculated from the quantile extreme value function, substituting the parameter estimates and the day. For example, for the first 33% shrimp the result is:

$$y_p = \mu(x) + [\log(\log(1 - p))](p)\sigma$$

$$y_p = \beta_0 + \beta_1 x + [\log[\log(1 - p)](p)\sigma$$

$$y_{33} = -14.7 + 3.4940(169) + [\log[-\log(1 - .33)](.33)(2.7015) = 26.99055$$
AN EXTREME VALUE REGRESSION APPROACH TO THE PREDICTION OF
SHRIMP SIZE DISTRIBUTION

Table 6 Shrimp size prediction for different days using the Extreme Value
Regression Model $y = \sqrt{x}$

<table>
<thead>
<tr>
<th>Obs</th>
<th>Day</th>
<th>Probability</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>160</td>
<td>0.33</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>160</td>
<td>0.67</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>160</td>
<td>1.00</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>170</td>
<td>0.33</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>170</td>
<td>0.67</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>170</td>
<td>1.00</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>180</td>
<td>0.33</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>180</td>
<td>0.67</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>180</td>
<td>1.00</td>
<td>38</td>
</tr>
<tr>
<td>10</td>
<td>190</td>
<td>0.33</td>
<td>31</td>
</tr>
<tr>
<td>11</td>
<td>190</td>
<td>0.67</td>
<td>34</td>
</tr>
<tr>
<td>12</td>
<td>190</td>
<td>1.00</td>
<td>39</td>
</tr>
<tr>
<td>13</td>
<td>200</td>
<td>0.33</td>
<td>32</td>
</tr>
<tr>
<td>14</td>
<td>200</td>
<td>0.67</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>200</td>
<td>1.00</td>
<td>41</td>
</tr>
</tbody>
</table>

Limitations

This study utilized secondary data, which made it impossible for the authors to
gather other independent variables that could potentially be used in the model.
Unlike Wyban, Walsh and Godin (1995), Ponce-Palafox, Martinez-Palacios, and
Ross (1997), and Esmaeili and Tarazkar (2011), this study did not have additional
variables to perform the statistical analysis. This is actually very common in an
applied setting; however, it can be considered a limitation of the study. Future
research may focus on utilizing different meaningful variables in the model in order
to evaluate the effect of one or more variables over the response variable.

Future Research

It is always desirable to have more and detailed information in order to have better
models. Thus, the addition of different independent variables would be a clear future research goal. Variables such as diet, water condition, and temperature could be added to this model to optimize model predictions. Additionally, the model can be improved substantially if there were biometric information at the time of planting, in order to have a better estimation on the left side of the growth curve. Additionally, other non-normal regression models could be explored such as the generalized extreme value distribution.

**Conclusions**

In this study, using empirical data, we explored the potential of the extreme value regression model to predict shrimp growth. The economic importance of shrimp growth goes hand-in-hand with the need of developing accurate statistical models to predict shrimp growth (Tian et al., 1993). Araneda et al. (2013) first introduced the idea of including size heterogeneity in a growth model. In this paper, the authors used the extreme value distribution and size heterogeneity to predict the shrimp size distribution in similar conditions. This model does not only estimate the average weight, but also the percentiles and the size percentages for a determined time. Also, the model is able to extrapolate information in order to predict the size distribution to determine time in the future. This allows certainty for strategic planning purposes and decision making for farmers and statisticians working in the aquaculture field. The most important implication of this study is that the shrimp size distribution can be better modeled using an extreme value distribution as shown by the results of this study.

**References**


AN EXTREME VALUE REGRESSION APPROACH TO THE PREDICTION OF SHRIMP SIZE DISTRIBUTION


