Reprint

Far East Journal of Theoretical Statistics

PUSHPA PUBLISHING HOUSE
Vijaya Niwas, 198, Mumfordganj
Allahabad 211002, INDIA
Tel.: 0091-532-2250078
e-mail: arun@pphmj.com
BOOTSTRAPPING NONLINEAR REGRESSION
APPLIED TO ARPS HYPERBOLIC
DECLINE CURVE

SUTAWANIR DARWIS\textsuperscript{1}, AGUS YODI GUNAWAN\textsuperscript{2}, M. ALI ASHAT\textsuperscript{3},
SRI WARYUNINGSIH\textsuperscript{1,4} and NURTITI SUNUSI\textsuperscript{1,5}

\textsuperscript{1}Statistics Research Division
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung, Indonesia
e-mail: sdarwis@math.itb.ac.id

\textsuperscript{2}Financial and Industrial Mathematics Research Division
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung, Indonesia
e-mail: ayodi@math.itb.ac.id

\textsuperscript{3}Geothermal Laboratory
Faculty of Earth Sciences and Mineral Technology
Institut Teknologi Bandung, Indonesia
e-mail: ali_ashat@tm.itb.ac.id

\textsuperscript{4}Statistics Study Program
Faculty of Mathematics and Natural Sciences
Universitas Mulawarman
Samarinda, East Kalimantan, Indonesia
e-mail: sri_ws305@students.itb.ac.id

2000 Mathematics Subject Classification: 93E24, 62F40.

Keywords and phrases: nonlinear least squares regression, randomness adjustment, bootstrap method.

Received November 17, 2008
Abstract

Geothermal energy is actually the heat that can be extracted from the interior of the earth and considered as an alternative renewable energy. Regression analysis is one of statistical tools that are frequently used in geothermal data analysis. However, nonlinear regression is rarely used in analyzing steam mass flow data due to its tricky computational and inferential problems. A probabilistic method in reserve estimation suggests the use of nonlinear regression and bootstrap method for the estimation of the potential recovery of a geothermal field. A random resampling of stochastic components is used to generate a large number of mass flow data to be used in evaluation of production performance. This resampling scheme, called bootstrap method, does not rely on the assumption of normality. Bootstrap was developed based on one-sample model where a single unknown distribution produces the data by random sampling. Development of bootstrap method in decline curve analysis involves nonlinear regression of steam mass flow, and bootstrap nonlinear least squares algorithm was developed similar to bootstrapping residual linear regression. Tracer modeling aims to determine the degree of connectivity between injection and production wells, and non-linear regression was developed to estimate the parameter of tracer model. This paper aims to report the development of bootstrap nonlinear least squares regression to hyperbolic decline curve analysis. The main contribution is to show that: (a) the simulation of Arps hyperbolic decline curve leads us to consider a modification of Arps equations by introducing a new parameter which represents the randomness adjustment and (b) bootstrap method can be used to estimate the distribution of hyperbolic decline curve parameters.

1. Introduction

The bootstrap resampling approach has been used to the petroleum production data to obtain stochastic distribution of future production. The Arps hyperbolic decline curve analysis is the most popular method used in production data analysis. Parameters of the Arps equations are
estimated either by nonlinear least squares regression or by linearization using binomial expansion or Taylor expansion. Stochastic approaches have increasingly used to evaluate the uncertainty of the future production based on decline curve analysis because the historical data usually possess significant amount of noise. The normality assumption does not always hold in reservoir data analysis due to decline of reservoir pressure. This situation leads to data driven Monte Carlo methods to analyze reservoir production data. The bootstrap method is a random resampling of stochastic component used to generate a large number of statistics to investigate its sampling distribution. The method starts with generating $B$ independent bootstrap replications from the original sample. For a sample of size $n$, a sample replication of size $n$ is obtained by resampling $B$ times with replacement. Each bootstrap sample is fitted using nonlinear least squares regression to determined hyperbolic decline parameters, and extrapolated to forecast future production. The distribution of future production is determined from bootstrap samples. An amount of research has been done in the area of bootstrapping decline curve analysis. For example, Jochen and Spivey [8] developed a probabilistic reserves estimation using decline curve analysis with the bootstrap method. Cheng et al. [4] addressed an improved methodology using modified bootstrap and block resampling for probabilistic quantification of reserves estimates using decline curve analysis.

The bootstrap nonlinear least squares hyperbolic decline curve analysis presented in this article provides a nonparametric data driven method to obtain future production estimates. The production data was simulated using Arps hyperbolic equations with randomness adjustment $a$ and the parameters were estimated using nonlinear least squares regression method. The randomness adjustment represents the variation of steam mass flow due to the water injection and leads us to propose a modification of Arps equations by introducing a randomness adjustment parameter. Binomial expansion, Taylor expansion, and nonlinear least squares approach to hyperbolic decline curves will be reviewed, bootstrap regression residual is presented, a modification of traditional decline curve analysis is proposed, and some simulations’ result will be discussed.
2. Bootstrapping Nonlinear Regression

Consider the Arps hyperbolic decline curve (Arps [1])

\[
\frac{1}{q_t} \frac{du}{dt} = -Dq_t^b \quad \text{or} \quad q_t = q_0 (1 + bDt)^{-1/b},
\]

where \(q_t\) is the production rate at time \(t\), \(q_0\) is the initial production rate, \(b\) is the hyperbolic decline exponent, and \(D\) is the decline rate. The hyperbolic equation can be rewritten as

\[
\ln \frac{q_t}{q_0} = -\frac{1}{b} \ln(1 + bDt) \quad \text{or} \quad \left( \frac{q_t}{q_0} \right)^{-b} = 1 + bDt.
\]

Both transformations above will be used further in applying binomial and Taylor expansions. One assumes an estimate for \(bD\), and checks for the linearity. If the historical data fall on a straight line, the assumed estimate is correct. An iterative scheme is set up to a certain degree of accuracy of parameter updating. Following Spivey [11], using binomial expansion the mass flow data are fitted by a linear regression \(q_t = \beta_0^{(i)} + \beta_1^{(i)} t + \beta_2^{(i)} t^2 + \epsilon_t\), where \(i\) is the iteration indices, \(i = 0, 1, \ldots\). The updating approximation to decline exponent parameter \(b\) is given by

\[
b^{(i)} = b^{(i-1)}[1 - 2\beta_0^{(i-1)} / \beta_1^{(i-1)}].
\]

Once \(b\) has been determined, the other parameters can be obtained

\[
q_0 = [\beta_0^{(i)}]^{-1/\beta_0^{(i)}}, \quad D = \beta_1^{(i)}/[b^{(i)} \beta_0^{(i)}].
\]

A Taylor expansion of Arps hyperbolic decline curve leads to a regression model of the form (Spivey [11])

\[
\ln q_t = \beta_0^{(i)} + \beta_1^{(i)} \ln(1 + B^{(k)} t) + \beta_2^{(i)} \frac{t}{(1 + B^{(k)} t)} + \epsilon_t.
\]

The updating estimate of parameter \(B\), \(B^{(k)}\), \(k = 0, 1, \ldots\) is given by the equation

\[
B^{(k)} = B^{(k-1)} + \frac{\beta_2^{(i)}}{\beta_1^{(i)}}.
\]
Once the iteration converges, the other parameters can be estimated using the relationships

\[ q_0 = e^{e_0^{(i)}}, \quad D = -(\beta_1^{(i)} B^{(k)} + \beta_2^{(i)}), \quad b = -\frac{1}{\beta_1^{(i)}}. \]

Both methods converge linearly provided the initial estimate is good enough (Spivey [11]). Gentry's method can be used to obtain an initial estimate of \((q_0, b, D)\) (Gentry [7]).

A simulation model of the form (Spivey [11])

\[ q_t = (1 + a r) \frac{q_0}{(1 + b D t)^{1/b}} \]

is used to generate synthetic flow rate data, where \(r \sim U(-1, 1)\), \(a\) is a randomness adjustment parameter. Nonlinear regression is a statistical tool which may be used to estimate the Arps hyperbolic decline curve parameters \(q_0, b, D\). It is often constraint the values of parameters to lie within certain ranges. A common choice would be \(q_0 > 0, \quad D > 0, \quad 0 < b < 1\) (Jochen and Spivey [8]).

Any model that is not linear in parameters will be called a nonlinear model. Draper and Smith [5] introduced nonlinear least squares estimation procedure, which involved the fitting of nonlinear regression models by least squares technique. The normal equations are not linear, and the Gauss Newton minimization of the sum of squares is usually perform. A nonlinear model that can be transformed into linear form is termed as intrinsically nonlinear. An intrinsically nonlinear model \(Y = \beta_0 e^{\beta_1 X} e\) can be rendered in a linear form after a logarithmic transformation. A 2-parameter nonlinear model \(Y = \beta_0 e^{\beta_1 X} + e\) cannot be transformed to linear model because the error term is additive. Now, consider a nonlinear model

\[ E(Y) = f(\xi, 0) - Y = f(\xi, 0) + e, \]

where \(\xi = (\xi_1, ..., \xi_k)\) are the predictor variables, and \(\Theta = (\theta_1, ..., \theta_p)\) are the model parameters. The error sum of squares, SSE, of the model is
\[ \text{SSE}(\Theta) = \sum_{i=1}^{n} (Y_i - f(\xi_i, \Theta))^2. \]

The differentiation with respect to \( \Theta \) provides the normal equations

\[ \sum_{i=1}^{t} [Y_i - f(\xi_i, \Theta)] \left[ \frac{\partial f(\xi_i, \Theta)}{\partial \theta_j} \right] = 0 \]

for \( j = 1, \ldots, p \). A widely used method for minimization is Newton's method. The method requires the use of both the gradient vector and the Hessian matrix of \( f \). The requirement for second derivatives motivates the development of other methods that are less demanding. The Gauss-Newton consists of approximating the \( k \)-th derivative using only the first two terms in Taylor series expansion. Details of the development of the Gauss-Newton method can be found in Kennedy and Gentle [9].

Efron and Tibshirani [6] described the bootstrap methodology for assessing statistical accuracy of parameter estimate using sample empirical distribution. There are two ways of bootstrapping a regression model: bootstrapped the residuals and bootstrapped the pairs. The following will focus on bootstrapped the residuals. The data set for a regression model consist of \( n \) pairs \( (X_i, Y_i), i = 1, \ldots, n \). The structure of the model is expressed as \( Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \ldots, n \). The least squares estimate is the solution to the normal equations \( X^T X \beta = X^T Y \), \( \beta = (\beta_0, \beta_1) \). The probability model for regression has two components \((\beta, F)\), where \( \beta \) is the parameter vector of regression, and \( F \) is the cdf of error term. The estimate of \( F \) is the empirical distribution of \( \epsilon_i = Y_i - \hat{Y}_i \) \( \sim \) iid \( \hat{F} : \frac{1}{n}, i = 1, \ldots, n \). Let \( \epsilon_i^*, i = 1, \ldots, n \) be the bootstrap sample, the bootstrap responses be generated according to \( Y_i^* = \beta_0^* + \beta_1 X_i + \epsilon_i^* \), and the bootstrap estimate be the minimize of the error sum squares

\[ \beta^* = \text{arg min}_{\beta} \sum_{i=1}^{n} [Y_i^* - (\beta_0^* + \beta_1^* X_i)]^2. \]
The bootstrap version of normal equations is $X^iX\hat{\beta}^* = X^iY^*$. The assumption behind bootstrapping pairs is that the pairs were randomly sampled from bivariate normal distribution. The accuracy of bootstrapping residuals approaches the accuracy of bootstrapping pairs as the number of pairs grows large (Efron and Tibshirani [6]).

3. Results and Discussion

**Binomial expansion and Taylor expansion.** Figure 1 shows the steam flow rate of geothermal production well and an IRIS visualization of injection and production wells. Under certain circumstances the production well can operate as injection well. The standard approach to analyze this data was the Arps hyperbolic decline rate. However, the pattern of the flow rate differs significantly from an ideal Arps hyperbolic decline rate. Spivey [11] suggested selecting a range of data points which follows the Arps decline curve and ignored the other data points. The fluctuation of mass flow rate around the ideal Arps curves is due to the effect of fluid (water) injection to the reservoir in order to maintain the reservoir pressure and to increase well production. A modification of Arps decline rate should be investigated to obtain a realistic model. A simulation study of Arps equation suggests a modification of Arps equation. This paper focuses the discussion on standard Arps model. Figure 2 shows the influence of randomness adjustment to the pattern of flow rate; $\alpha = .001$ represents an ideal Arps equation and $\alpha = .25$ shows a significant variation around the ideal Arps equation. The flow rate profile for $\alpha = .25$ is close to the actual field flow rate observations (Figure 1(a)). Therefore, this simulation experiment suggests a modification of Arps equation by adding a new parameter to adjust randomness due to the injection of fluid (water) to the reservoir. Table 1 explores the bias of parameters estimate of binomial expansion and Taylor expansion of Arps equation. The decline exponent $b$ shows greater variability than does that for $q_0$ and $D$. 
Figure 1. (a) Plot of steam flow rate of geothermal reservoir; (b) IRIS visualization of injection and production wells. The flow rate decline differs significantly from an ideal Arps decline models due to fluid (water) injection and reservoir heterogeneity. The statistical modeling of geothermal solute injection process is still an active research problem. One possible approach is a combination of diffusion modeling, Brownian motion, nonlinear regression modeling, and bootstrap methodology.

Figure 2. Comparison of $q_t$ simulated, predicted $q_t$ using binomial expansion (a) and predicted $q_t$ using Taylor expansion (b) for randomness adjustment $\alpha = .001, .25$, initial flow rate $q_0 = 1000 (t/h)$, decline exponent $b = .01$, and decline rate $D = .05 (l/m)$, and (c) visualization of injection of fluid (water) to the reservoir. For $\alpha = .001$ the curves are smooth and represent an ideal hyperbolic decline rate, and for $\alpha = .25$ the curves are fluctuated around the ideal ($\alpha = 0$) hyperbolic decline curves. The simulated $q_t$ is well predicted by the binomial expansion and Taylor expansion. Comparison of simulated $q_t$ with actual field observation shows that the case of randomness adjustment $\alpha = .25$ related to treatment of fluid (water) injection to the reservoir.
Table 1. Parameters \((q_0, b, D)\) of hyperbolic decline rate, parameters
estimates using binomial expansion and Taylor expansion for period of
observations \(T = 6, 12, 24, 48, 96\) months, randomness adjustment
\(a = .01\) and \(.001\), decline exponent \(b = .01\) and \(.25\), decline rate
\(D = .05\) and \(.25\), and initial flow rate \(q_0 = 1000\text{ (t/m)}\). The accuracy of
the estimates requires good initial estimates of the hyperbolic
decline rate parameters. In order the binomial expansion and Taylor
expansion are reasonable approximation; the two methods must
converge to the actual parameters values. The bias of decline
exponent \(b\) shows greater variability than does that for the other two
parameters. The bias of Taylor expansion shows less variability than
does that for binomial expansion.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(T = 6)</th>
<th>(T = 12)</th>
<th>(T = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Binomial</td>
<td>Taylor</td>
<td>Binomial</td>
</tr>
<tr>
<td>.01</td>
<td>(q_0)</td>
<td>1000</td>
<td>1011.7</td>
</tr>
<tr>
<td>.01</td>
<td>(b)</td>
<td>.427</td>
<td>.422</td>
</tr>
<tr>
<td>.05</td>
<td>(D)</td>
<td>.055</td>
<td>.055</td>
</tr>
<tr>
<td>.001</td>
<td>(q_0)</td>
<td>1000</td>
<td>1012.2</td>
</tr>
<tr>
<td>.01</td>
<td>(b)</td>
<td>.054</td>
<td>.054</td>
</tr>
<tr>
<td>.05</td>
<td>(D)</td>
<td>.051</td>
<td>.050</td>
</tr>
<tr>
<td>.01</td>
<td>(q_0)</td>
<td>1000</td>
<td>986.6</td>
</tr>
<tr>
<td>.01</td>
<td>(b)</td>
<td>.015</td>
<td>.015</td>
</tr>
<tr>
<td>.15</td>
<td>(D)</td>
<td>.150</td>
<td>.151</td>
</tr>
<tr>
<td>.01</td>
<td>(q_0)</td>
<td>1000</td>
<td>1001.1</td>
</tr>
<tr>
<td>.25</td>
<td>(b)</td>
<td>.296</td>
<td>.296</td>
</tr>
<tr>
<td>.05</td>
<td>(D)</td>
<td>.051</td>
<td>.050</td>
</tr>
</tbody>
</table>

**Bootstrapping nonlinear least squares (nls) regression.** Figure 3 shows bootstrapping Arps decline curves for randomness adjustment.
a = .001 and a = .25. B = 500 bootstrap replications were generated for case a = .001, and 100 for case a = .25. The simulated flow rate for randomness a = .001 (Figure 3(a)) shows less variability than does that for randomness adjustment a = .25 (Figure 3(b)). Bootstrap confidence intervals for a = .001 show less variability compared to a = .25, and therefore consistent with intuitive judgment without having to assume distribution of (q0, b, D). For reasons of clarify only five bootstrap replications are shown together with 2.5%, 5%, 95%, and 97.5% bootstrap percentiles, the simulated flow rate, binominal approximation, and Taylor expansion. Following WPC (World Petroleum Congress, 1983, see Jochen and Spivey [8]), reserves estimation is classified as proved (90% percentile), proved + probable (50% percentile), proved + probable + possible (10% percentile) (Capen [3]). Figure 3 and Table 3 can be used to get these confidence intervals based only on a single data set. The fact offers an advantage of bootstrap method over Monte Carlo simulation where the probability distribution of the parameters must be known. As expected, the bias of bootstrap estimates is fairly narrow for case a = .001, and much wider confidence intervals for case a = .25 (Table 2). The values of randomness adjustment contributed significantly to the bootstrap summary and bootstrap percentiles.

Figure 3. Bootstrap replications and bootstrap percentiles for randomness adjustment a = .001 and .25. Bootstrap accuracy is smaller for a = .001 compared to a = .25 (Table 2). Bootstrap confidence interval was constructed based on empirical bootstrap percentiles (Table 3).
Table 2. Bootstrap summary \( q_0 = 1000, b = .01, D = .05 \)

<table>
<thead>
<tr>
<th></th>
<th>( a = .001 )</th>
<th>( a = .25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Bias</td>
</tr>
<tr>
<td>( q_0 )</td>
<td>1002.265</td>
<td>.532</td>
</tr>
<tr>
<td>( b )</td>
<td>.019</td>
<td>.002</td>
</tr>
<tr>
<td>( D )</td>
<td>.050</td>
<td>.0001</td>
</tr>
</tbody>
</table>

Table 3. Empirical percentiles \( q_0 = 1000, b = .01, D = .05 \)

<table>
<thead>
<tr>
<th></th>
<th>( a = .001 )</th>
<th>( a = .25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5%</td>
<td>5%</td>
</tr>
<tr>
<td>( q_0 )</td>
<td>999.579</td>
<td>1000</td>
</tr>
<tr>
<td>( b )</td>
<td>.069</td>
<td>.0094</td>
</tr>
<tr>
<td>( D )</td>
<td>.04998</td>
<td>.004999</td>
</tr>
</tbody>
</table>

4. Summary

This paper aims to explore the applicability of bootstrap method to study the sampling distribution of Arps decline curve parameters with randomness parameter \( a \). Randomness adjustment parameter represents the fluctuation of fluid flow around the Arps decline curve equation, and leads us to an improvement of the traditional Arps decline curve analysis by introducing a new parameter. Simulation of Arps decline curve shown that bootstrap yields a smooth confidence interval for the case randomness adjustment \( a = .001 \) and a significant variability for \( a = .25 \). The sampling distribution of Arps decline statistics was successfully simulated using bootstrap data driven method without prior distribution of \((q_0, b, D)\). The bootstrap samples influence significantly the estimate of the decline exponent \( b \). Bootstrap method can be used to estimate the sampling distribution of decline curve parameters and provide a prediction estimates of mass flow production based on decline curve equations. Monte Carlo simulation is a popular method for reservoir reserves estimation and based on two assumptions: a model which predicts the reservoir performance and the distribution of the
parameters of the model are known (Jochen and Spivey [8]). However, the bootstrap approach offers the advantages over Monte Carlo simulation: there is no prior assumptions of parameter distribution, the sampling distribution of parameter distribution is based on a single sample data set, i.e., nonparametric, the bootstrap estimate yields an interval prediction of mass flow rate, and can be extended to estimate energy reserve of geothermal reservoir. In recent years, bootstrap gains its popularity in engineering literature due its flexibility to normality assumption. Bootstrap methodology can be used to study the sampling distribution of nonlinear regression parameters of tracer modeling (Axelsson et al. [2] and Kocabas [10]).

Acknowledgements

The authors would like to thank the computational team for providing the decline curves computations. This work was supported by ITB research grant 2007.

References


