Cutting the Gordian Knot: Maximum Likelihood Estimation for Regression of Log Normal Error

Christian Smart, Ph.D., CCEA
Director, Cost Estimating and Analysis
Missile Defense Agency
christian.smart@mda.mil

Public Release 16-MDA-8878

Introduction

- In 333 BC in one of his early military campaigns Alexander the Great reached the city of Gordium
- When shown an intricate knot with its ends hidden, he cut the knot with his sword
  - The phrase "cutting the Gordian knot" has come to mean a direct and simple solution to a complex problem
  - We leverage the powerful sword maximum likelihood estimation to provide a simple alternative to log-transformed ordinary least squares

Linear Regression

- Given an equation of the form

$$Y = a + bX$$

- And a set of data

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$$

- The residuals are defined as:

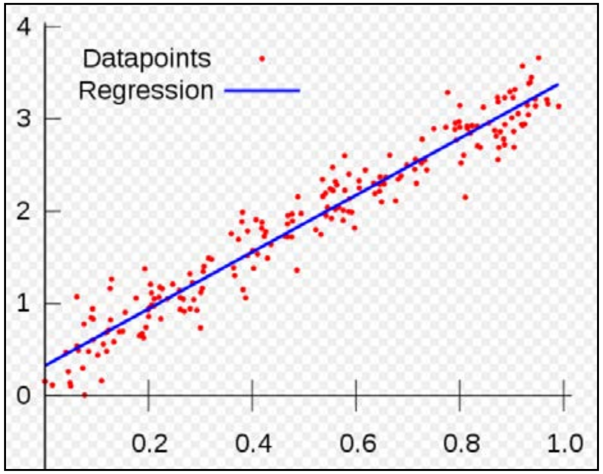
$$\varepsilon_i = Y_i - (a + bX_i) = Actual - Estimated$$

- This is also referred to as the "error" term since it is the difference between the actual cost and the estimated cost linear regression finds the "best fit" by finding the parameters *a* and *b* that minimize the sum of the squares of the residuals

$$\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} \left(Y_i - (a + bX_i)\right)^2 = \sum_{i=1}^{n} \left(Actual_i - Estimated_i\right)^2$$

Least Squares and Regression Analysis

- **The method of least squares was first developed by the mathematicians Legendre and Gauss in the early 19th century, who used it to predict the orbits of heavenly bodies using observed data**
- **Francis Galton later applied this technique to find linear predictive relationships between various phenomena, such as the relationship between the heights of fathers and sons**
  - **Galton found a positive correlation between these heights but found a tendency to return or "regress" toward the average height, hence the term "regression analysis"**



Nonlinear Regression

- In the spacecraft and defense industry it is more common to see nonlinear relationships between cost and cost drivers
- The power equation is ubiquitous

$$Y = aX^{b}$$

- In this case Y typically represents cost in $, but can also represent effort (hours, full-time equivalents)
- *X* typically represents weight or some other performance parameter
- The equation can also be modified to accommodate multiple cost drivers
- The value of the *b* parameter in the power equation is usually less than *1*, indicating economies of scale in design and production
- Linear regression is simple - the calculations can be done by hand, but nonlinear regression requires more sophisticated methods, often the use of a computer

Additive and Multiplicative Residuals

- The residuals of the power equation can either be additive or multiplicative
- Additive residuals have the form

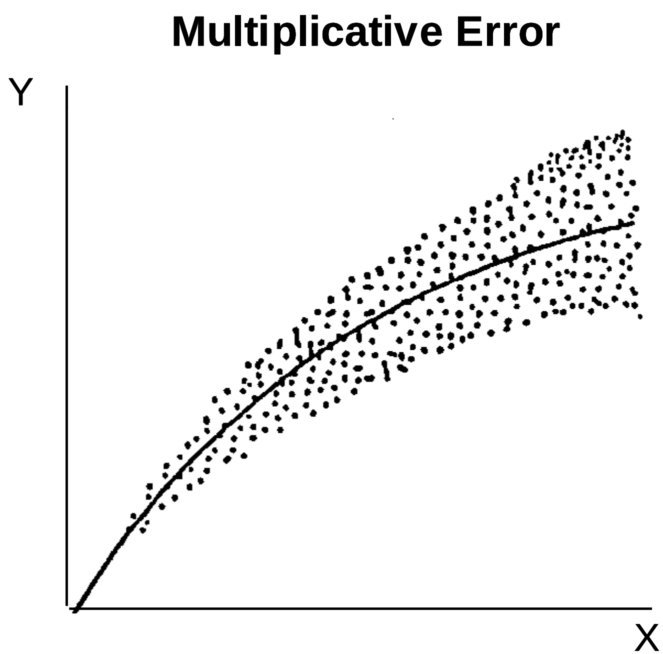
$$Y = aX^b + \varepsilon$$

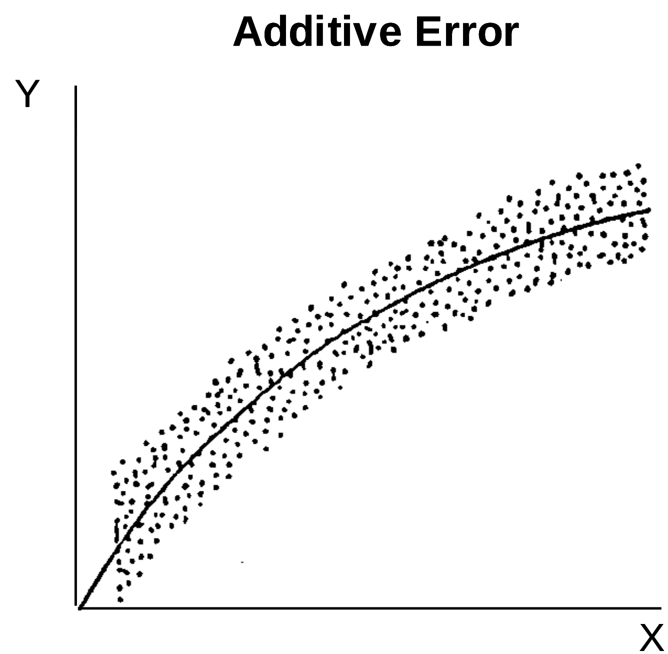
- Multiplicative residuals have the form

$$\boxed{Y = aX^b \varepsilon}$$

- Multiplicative residuals are more appropriate for the spacecraft and defense industry in most applications because of wide variations in size, scope, and scale of the systems that are estimated
  - As a result we are primarily interested in the percentage difference between actual and estimated costs, not the absolute difference
- For example, if historical data ranges from $50 million to $1 billion, better to analyze percentage differences

Residuals Comparison





- **The commonly-used regression techniques considered in this presentation are all based on the multiplicative error assumption**

Multiplicative Residuals

- **For the Power Equation with Multiplicative Residuals, i.e.,**

$$Y = aX^{b}\varepsilon$$

- **The Regression Estimates Vary Based on the Variation of the Residual**

$$\varepsilon = \frac{Y}{aX^{b}}$$

- **Also Common to Adjust This to Treat ε as a Percentage, i.e., Set**

$$Y = aX^{b}(1+\varepsilon)$$

$$\varepsilon = \frac{aX^{b} - Y}{aX^{b}} = \frac{Estimate - Actual}{Estimate}$$

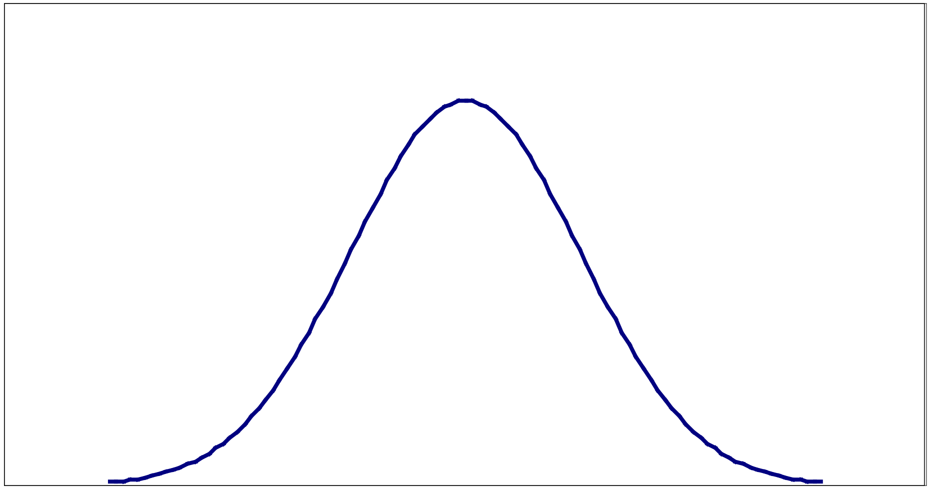
- **Actual Cost = Estimate +/- Percentage of Estimate**

Residuals are Random Variables

- For a "good" regression model, the cost drivers explain all (or most) of the variation in the historical data that can be explained
  - It is typically assumed that any remaining variation is random
    - Either due to non-repeatable random phenomena (e.g., test failures) that are truly random phenomena and will not help predict future cost, or due to our ignorance
- The multiplicative residuals that represent this unexplained variation are thus treated as random variables
- For linear regression, it is assumed that the additive residuals are normally distributed
- For nonlinear regression for CER development, residuals assumed to follow normal, lognormal, gamma, or treated without making such an assumption (non-parametric)

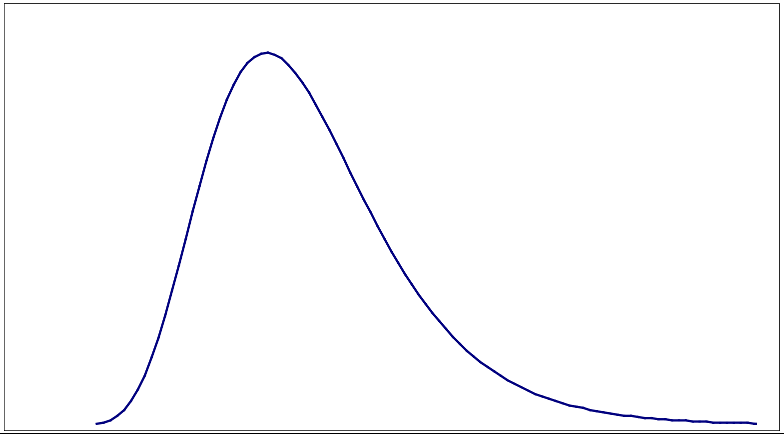
Normal Distribution

- **The most common probability distribution - many random phenomena follow this distribution**
- **Also called the "bell curve," noted for its symmetry and thin tails.**
- **If cost is a sum of many random independent phenomena, the central limit theorem indicates this may be the appropriate distribution.**



Lognormal Distribution

- **Lognormal distribution is skewed**
- **If *x* is lognormally distributed, *y* = *ln(x)* is normally distributed**
- **Fatter tails than the normal distribution**
- **Bounded below by zero, unbounded above.**
- **If cost is a function is multiplicative factors (e.g., test failues cause a percentage increase in cost rather than a fixed amount increase), then complex projects are likely to be lognormally distributed (multiplicative analog to central limit theorem)**
- **These aspects make the lognormal appealing for cost modeling**

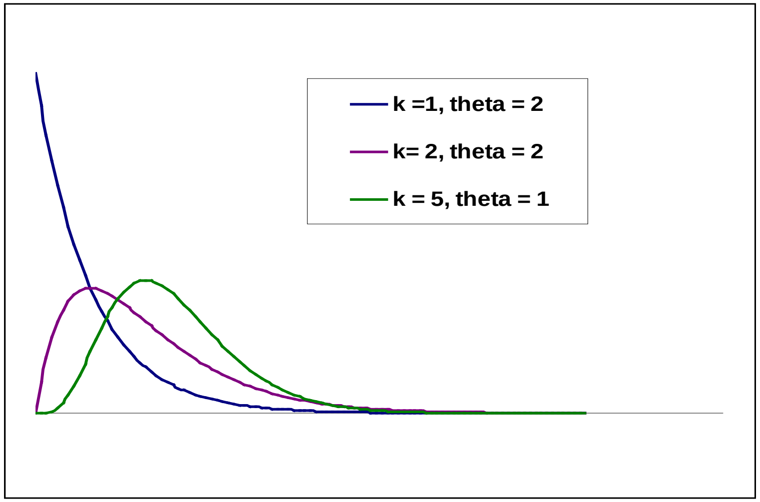


Gamma Distribution

- **The Gamma Distribution is a Flexible Distribution.**
- **Can Resemble a Lognormal, Can Also Resemble an Exponential Distribution.**
  - **Indeed the Gamma Distribution is the Sum of Independent Exponential Distributions.**

- **PDF is Given by:**

$$f(x) = x^{k-1} \frac{e^{-\frac{x}{\theta}}}{\Gamma(k) \cdot \theta^k}$$



Parameter Calculation

- **There are numerous ways to calculate the parameters of a cost-estimating relationship, but we consider one method – Maximum Likelihood Estimation (MLE)**
- **Maximum Likelihood Estimation is a widely used statistical technique that serves as a unifying framework for the CER methods we discuss**

Maximum Likelihood Estimation

- **Let $A_1, \ldots, A_n$ represent the observed data and $X_1, \ldots, X_n$ represent random variables where A_i results from observing the random variable X_i**
- **The likelihood function, which represents the likelihood of obtaining the sample results, is**

$$L(\theta) = \prod_{i=1}^{n} Pr\left(X_i = A_i \mid \theta\right)$$

- **The Maximum Likelihood Estimate of θ is the vector that maximizes the likelihood function**
- **Maximum Likelihood Estimation is an established popular statistical technique**
  - **Major advantage – likelihood function is almost always available**

Maximum Likelihood – Lognormal Residuals

- For $Y_i = f(X_i, \beta)\varepsilon_i$, where
  β = vector of coefficients of the CER
  Y_i = actual cost of the i^{th} data point
  X_i = vector of cost drivers for the i^{th} data point
  ε_i = residual of the i^{th} data point
- Probability density function for lognormal distribution

$$p(y, \mu, \theta) = \frac{1}{y\sqrt{2\pi\theta}} e^{-\frac{(\ln y - \mu)^2}{2\theta}}$$

- Note that μ is the log-space mean
- If we estimate $\mu = ln(Y)$ then in the case of the power equation

$$Y = \beta_0 X^{\beta_1}$$

  we are estimating the linear equation $\mu = ln\beta_0 + \beta_1 ln(X)$
- Note that $e^{\mu} = Y = \beta_0 X^{\beta_1}$ is the median in linear space

Estimating the Median

- **Conditional probability/likelihood of y_i given x_i and β:**

$$L(\beta,\theta)=\prod_{i=1}^{n} p\left(y_i \mid x_i;\beta_0,\beta_1,\ldots,\beta_p,\theta\right)=\frac{1}{y\sqrt{2\pi\theta}}e^{-\frac{\left(\ln y-\ln\beta_0-\beta_1\ln X_{i1}-\ldots-\beta_p\ln X_{ip}\right)^2}{2\theta}}$$

- **If we take the log-likelihood and maximize that, we find it is equivalent to minimizing**

$$\sum_{i=1}^{n}\left(\ln y_i-\ln\beta_0-\beta_1\ln X_{i1}-\ldots-\beta_p\ln X_{ip}\right)^2$$

which is Log-Transformed Ordinary Least Squares (LOLS)

- **Therefore LOLS is an MLE of the median**
- **Note that we can minimize the sum of squared errors for an arbitrary function $f(X_i,\beta)$ as well**

Estimating the Median Vice the Mean

- For skewed data the median is a better representative of a distribution's centrality
- However we rarely consider projects in isolation, but as part of a larger portfolio
- The medians (also known as 50^{th} percentiles) do not add, and the sum of 50^{th} percentiles is less than the 50^{th} percentile of the sum
- The mean is a better choice for estimation, so we say that LOLS is biased low
- There are factors that allow you to correct for this (e.g., "PING Factor"

Normal Residuals

- **For the Equation $Y_i = f(X_i, \beta)\varepsilon_i$, when the residuals are *normally* distributed, with mean = *1* and variance θ, the likelihood function, is**

$$L(\beta,\theta)=\frac{exp\left(\frac{-1}{2\theta}\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2\right)}{(2\pi\theta)^{\frac{n}{2}}\prod_{i=1}^{n} f(x_i,\beta)}$$

- **The negative log-likelihood function can be simplified to the following representation**

$$l^*(\beta)=\frac{n}{2}ln\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2+\sum_{i=1}^{n} ln\, f(x_i,\beta)$$

ZMPE

- **The normal residual MLE is very similar to the Zero-Percent Bias Minimum Percent Error (ZMPE) method developed by Dr. Steve Book**
- **ZMPE minimizes**

$$\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right)^2$$

subject to the constraint that the sample bias is zero, i.e.,

$$\sum_{i=1}^{n}\left(\frac{y_i - f(x_i,\beta)}{f(x_i,\beta)}\right) = 0$$

Normal MLE and ZMPE (1 of 2)

- The dominant term in the Normal MLE is the same as the ZMPE objective function, so the results of the two are often the same
- For the Data Displayed in the Table and Graphically Displayed in the Charts:
  - Normal MLE Fit is

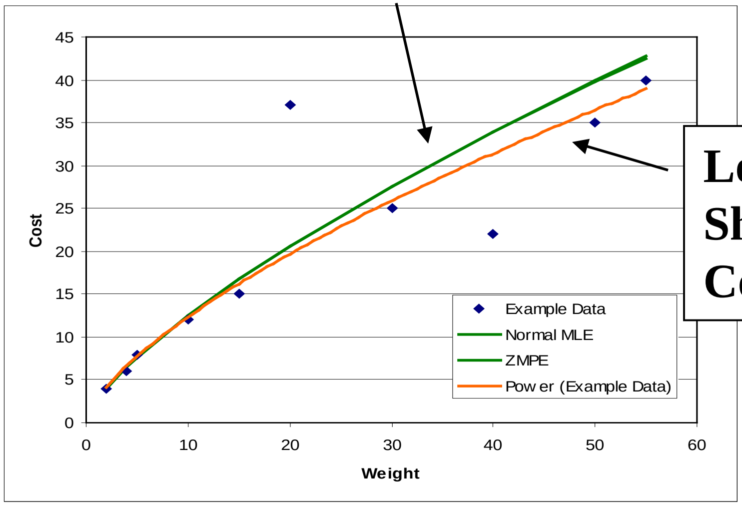
$$Est.Cost = 2.41Wt.^{0.716}$$

  - MPE-ZPB Fit is

$$Est.Cost = 2.39Wt.^{0.719}$$

Weight	Cost
2	4
4	6
5	8
10	12
15	15
20	37
30	25
40	22
50	35
55	40

MPE-ZPB and Normal MLE Fits are Coincident



Loglinear Fit Shown for Contrast

Normal MLE and ZMPE (2 of 2)

- **Normal MLE and ZMPE solutions are very similar since they are minimizing the same dominant term and are both "unbiased"**
  - **MLE solution is asymptotically unbiased (unbiased for "large" samples)**
  - **MPE-ZPB solution is unbiased regardless of sample size**
- **One advantage that MPE-ZPB has is lack of bias regardless of sample size**
  - **Cost estimates are often based on small samples, so MLE solution may be biased**
- **On the other hand, MPE-ZPB is tied to the assumptions of the normal MLE.**
  - **Need normally distributed (multiplicative) residuals to ensure consistent solutions in many cases.**

Gamma Residuals

- **When the residuals follow a gamma distribution, the negative log-likelihood function is**

$$l(\beta) = \sum_{i=1}^{n} \left(\frac{y}{f(x_i, \beta)} + ln\, f(x_i, \beta) \right)$$

- **This can be minimized by iteratively minimizing the sum of percent squared errors until the estimates converge:**

$$\sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \beta_k)}{f(x_i, \beta_{k-1})} \right)^2$$

- **Note *k* is the iteration number**
- **This method was first developed by Nelder (1968) and Wedderburn (1974), who called the method Iteratively Re-Weighted Least Squares (IRLS) and re-discovered by Hu in the 1990s, who called it Miminum Unbiased Percentage Error (MUPE)**

IRLS/MUPE

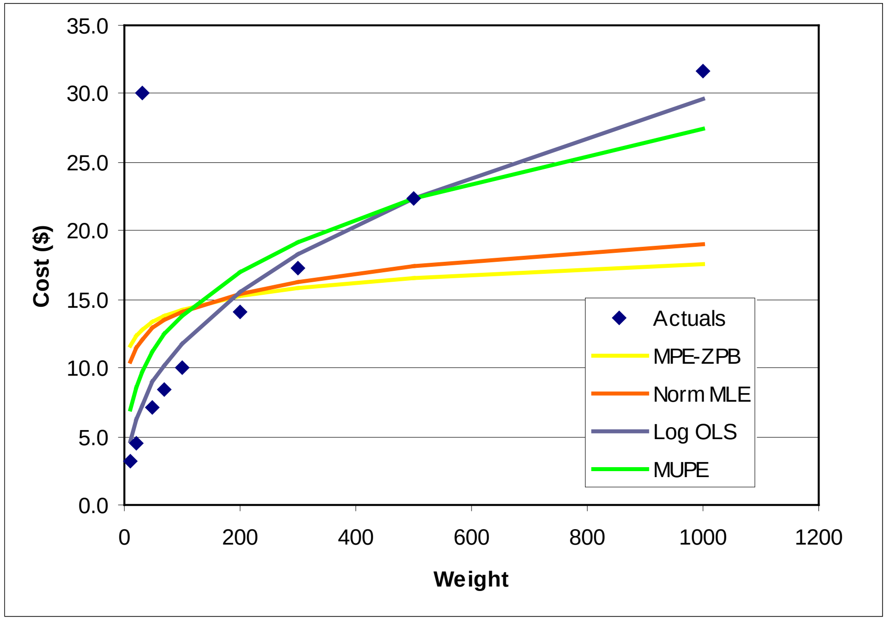
- In the case of gamma residuals, IRLS/MUPE is an MLE
  - also a Generalized Linear Model (GLM)
- However, IRLS/MUPE does not depend upon the assumption of gamma residuals.
- The likelihood method was generalized by Wedderburn to consider quasi-likelihood, which has good statistical properties

Summary of Three CER Methods

- Log-Transformed OLS, MPE-ZPB, and IRLS/MUPE all share a common connection in Maximum Likelihood Estimation
- Log-Transformed OLS is a Maximum Likelihood Estimator of the median of lognormally distributed multiplicative residuals
  - Parametric method
- MPE is a Pseudo-Likelihood Estimator of the Mean of normally distributed multiplicative residuals
  - Bias constraint added
  - Not directly parametric but has parametric properties
- IRLS/MUPE is a Maximum Likelihood Estimator of the mean of gamma distributed residuals
  - Also more general, quasi-likelihood
  - Parametric method

Comparing the Methods – an Example

Weight	Cost
10	3.2
20	4.5
30	30
40	7.1
70	8.4
100	10
200	14.1
300	17.3
500	22.4
1000	31.6



- **Although MPE-ZPB has the lowest standard percent error, the overall trend does not match the actual data**
- **MUPE and LOLS have similar fit**
  - **Similar to results reported by Mackenzie(2003)**

Lognormal – Theoretical Foundations (1 of 3)

- **Changes in cost over time are proportional to prior costs**
  - **Cost is more likely to increase than decrease over time**
  - **Over 80% of government projects experience cost growth**
- **Cost increases do not typically result in increased funding in the short term, so cost increases result in schedule slips**
- **Schedule slips mean a longer period in which personnel devoted to a project will charge to that project**
- **Larger projects have more people assigned to them so increases in cost will be proportional to the size of the project**

Lognormal – Theoretical Foundations (2 of 3)

- **The change in cost from time *t-1* to time *t* is**

$$X_t - X_{t-1} = \epsilon_t X_{t-1}$$

where the $\epsilon_t{}'s$ are mutually independent and independent of X_{t-1}

- **This is equivalent to**

$$\frac{X_t - X_{t-1}}{X_{t-1}} = \epsilon_t$$

- **Summing over t**

$$\sum_{t=1}^{n} \frac{X_t - X_{t-1}}{X_{t-1}} = \sum_{t=1}^{n} \epsilon_t$$

Lognormal – Theoretical Foundations (3 of 3)

- **Proportional changes can be approximated as**

$$\sum_{t=1}^{n} \frac{X_t - X_{t-1}}{X_{t-1}} \approx \int_{X_0}^{X_n} \frac{dX}{X} = ln(X_n) - ln(X_0)$$

- **Thus**

$$ln(X_n) - ln(X_0) \approx \sum_{t=1}^{n} \epsilon_t$$

or equivalently

$$ln(X_n) \approx ln(X_0) + \sum_{t=1}^{n} \epsilon_t$$

- **According to the Central Limit Theorem the sum of many independent random variables is normally distributed, so $ln(X_n)$ is normal for large n, thus X_n is lognormal**

The Use of the Lognormal in Other Industries

- Rule of thumb for environmental studies
- Pareto and lognormal are the only distributions found to fit actuals for size claims in property and casualty insurance
- Lognormal is widely used in health and labor economics
  - LOLS and IRLS are widely used

Empirical Evidence for the Lognormal

- For 32 LOLS CERs developed for the NASA/Air Force Cost Model between 2007 and 2009 the residuals were fit to a variety of distributions using the Crystal Ball® add-in for Excel
  - Using the Anderson-Darling test, the lognormal was not rejected for 30 out of 32 at the 10% critical value and was not rejected for 31 of 32 at the 5% critical value
- Mackenzie (2008) has provided empirical evidence for modeling residuals with the lognormal
- Smart (2015) showed that cost growth for 289 DoD and NASA programs closely follows a lognormal distribution
- The Joint Agency Cost and Schedule Risk and Uncertainty Handbook (2014) recommends the lognormal distribution as a default for modeling cost risk if no other information is known about the shape of the distribution

A New Method

- We have provided theoretical and empirical evidence for modeling the residuals as lognormal
- However the optimal method currently in use for modeling lognormal residuals is subject to criticism
- Because of the transformation LOLS estimates the median of the lognormal
  - There are factors to translate the median to the mean but these are approximations and may not be accurate outside the input data range
- The new method is to use MLE to directly estimate the mean of the lognormal without the use of transformation

MLE Regression of the Log Normal (MRLN) (1 of 4)

- **We term this direct estimation of the mean MLE Regression for Log Normal error (MRLN or "Merlin")**
- **We being with the lognormal likelihood but now we are going to estimate the mean rather than the median**
- **We are estimating the power equation**

$$\boldsymbol{Y = \beta_0 X_1^{\beta_1} \ldots X_p^{\beta_p}}$$

- **The mean of a lognormal distribution is**

$$\boldsymbol{e^{\mu + \frac{\theta}{2}}}$$

- **For the $\boldsymbol{i^{th}}$ observation**

$$\boldsymbol{e^{\mu_i + \frac{\theta}{2}} = \beta_0 X_{i1}^{\beta_1} \ldots X_{ip}^{\beta_p}}$$

MLE Regression of the Log Normal (MRLN) (2 of 4)

- **Taking log transformations**

$$\mu_i + \frac{\theta}{2} = ln\beta_0 + \beta_1 lnX_{i1} + \cdots + \beta_p lnX_{ip}$$

- **Therefore**

$$\mu_i = ln\beta_0 + \beta_1 lnX_{i1} + \cdots + \beta_p lnX_{ip} - \frac{\theta}{2} = ln\beta_0 + \sum_{j=1}^{p} \beta_i lnX_{ip} - \frac{\theta}{2}$$

- **Recall that the log-likelihood is**

$$L(\mu, \theta) = \prod_{i=1}^{n} \frac{1}{y_i\sqrt{2\pi\theta}} e^{-\frac{(lny_i - \mu_i)^2}{2\theta}}$$

MLE Regression of the Log Normal (MRLN) (3 of 4)

- **The log-likelihood (ignoring constants) is thus**

$$l(\mu,\theta) = -\frac{1}{2\theta}\sum_{i=1}^{n}(lny_i - \mu_i)^2 - \sum_{i=1}^{n} lny_i - \frac{n}{2}ln\theta$$

- **We substitute for μ to obtain**

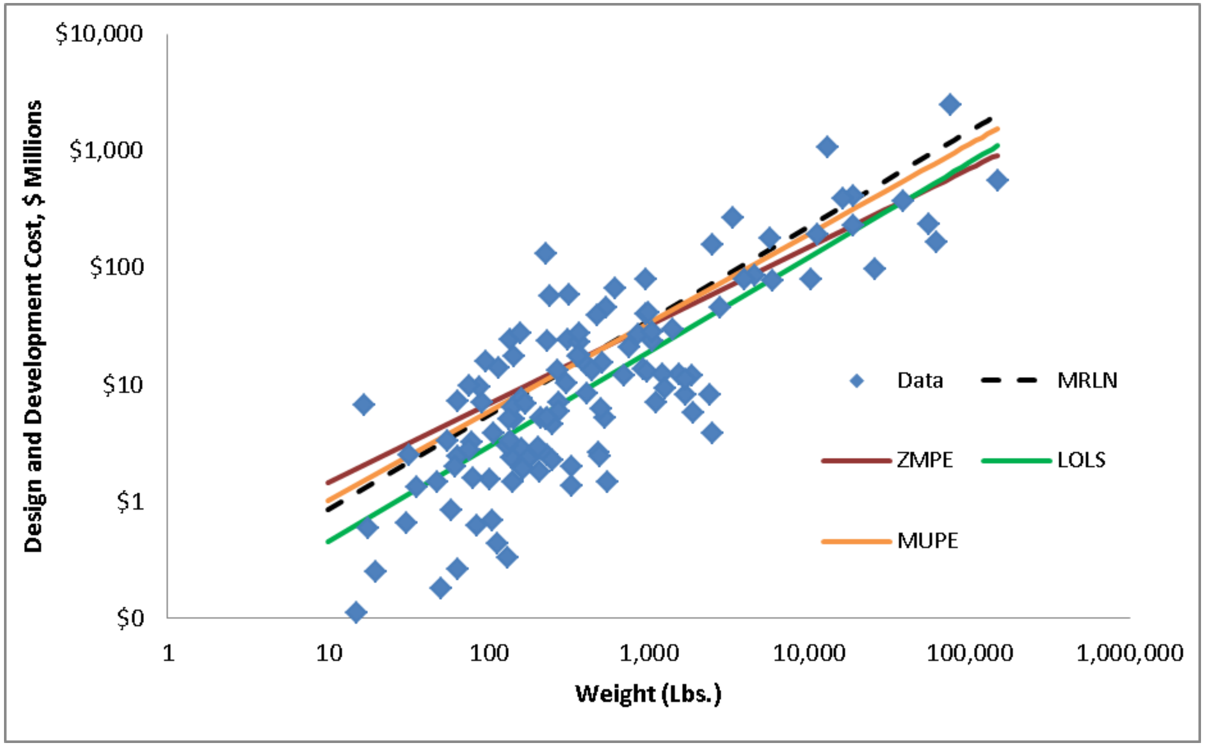
$$l(\beta_0,\beta_1,\dots,\beta_p,\theta)$$
$$= -\frac{1}{2\theta}\sum_{i=1}^{n}\left(lny_i - ln\beta_0 - \sum_{j=1}^{p}\beta_i lnX_{ip} + \frac{\theta}{2}\right)^2 - \sum_{i=1}^{n} lny_i - \frac{n}{2}ln\theta$$

MLE Regression of the Log Normal (MRLN) (4 of 4)

- **Ignoring constants we obtain**
  $$l(\beta_0, \beta_1, \dots, \beta_p, \theta)$$
  $$= -\frac{n}{2} ln\theta - \frac{1}{2\theta} \sum_{i=1}^{n} \left(lny_i - ln\beta_0 - \sum_{j=1}^{p} \beta_j lnX_{ij} + \frac{\theta}{2} \right)^2$$
- **There are no closed form solutions for the partial derivatives, so we need a numerical method, such as Newton-Raphson, to find the roots**
- **We can minimize the negative of the likelihood directly in Excel, using Solver, or by using the MLE package in R**

Example 1 (1 of 2)

- **121 data points for satellite and spacecraft Structures, design and development cost and weight**



Example 1 (2 of 2)

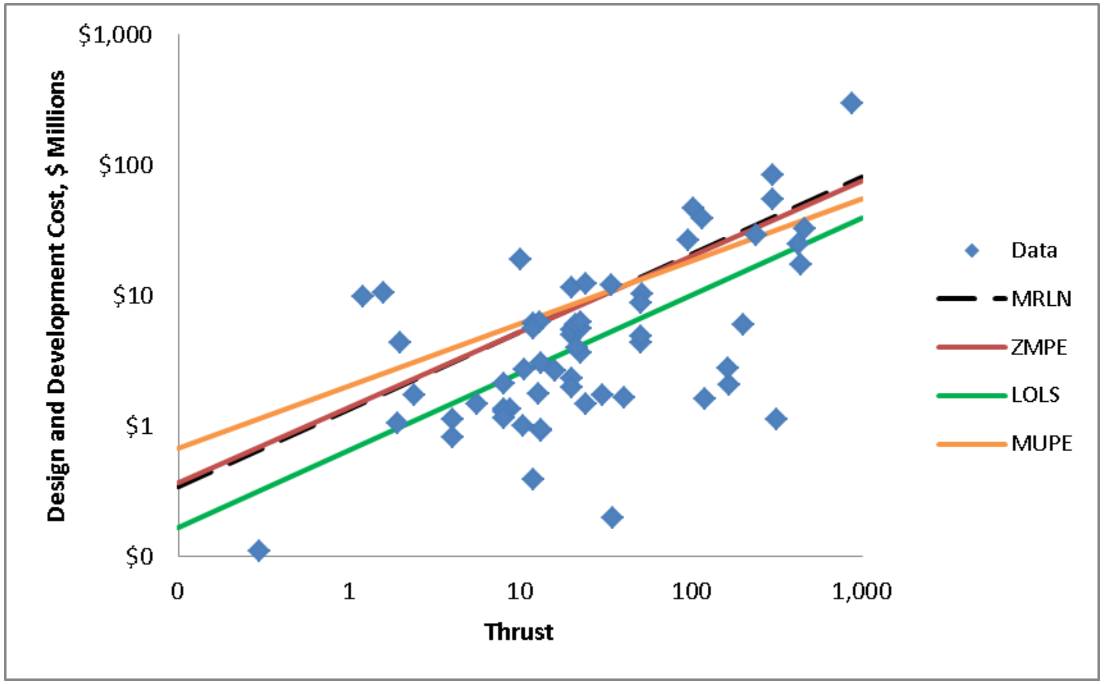
- The parameters and goodness-of-fit metrics are

Method	β_0	β_1	Pearson's R^2	Std % Error
MRLN	0.13	0.81	38%	146%
LOLS	0.07	0.81	38%	228%
MUPE	0.17	0.76	39%	143%
ZMPE	0.31	0.67	40.5%	140%

- The residuals follow a lognormal distribution (fail to reject at the 10% critical value for Anderson-Darling and Kolmogorov-Smirnov tests)
- MRLN, MUPE, and ZMPE all have similar fit statistics
- Bias is not an issue: sample bias is 0.2% for MRLN, 2.5x10-8 for MUPE and 8.7x10-7 for ZMPE

Example 2 (1 of 2)

- **For our second example we consider 62 reaction control subsystem – design and development cost and thrust for satellites and spacecraft**



Example 2 (2 of 2)

- The parameters and goodness-of-fit metrics are

Method	β_0	β_1	Pearson's R^2	Std % Error
MRLN	1.35	0.59	54%	144%
LOLS	0.65	0.59	54%	315%
MUPE	2.04	0.48	47%	157%
ZMPE	1.38	0.57	53%	147%

- The ZMPE and MRLN fit are similar for this case
- MUPE is the most different
- Residuals again follow a lognormal – fail to reject at the 5% critical value for either Anderson-Darling or Komolgorov-Smirnov

Example 3

- 77 command, control, and data handling subsystems costs weights, and % new design for satellites and spacecraft

Method	β_0	β_1	β_2	Pearson's R^2	Std % Error
MRLN	0.45	1.96	0.96	97%	63%
LOLS	0.39	1.96	0.96	97%	75%
MUPE	0.51	1.98	0.94	97%	62%
ZMPE	0.65	2.03	0.89	96%	61%

- The fits are all similar
- Data are better behaved with few outliers
- The residuals follow a lognormal, fail to reject at the 5% critical value for Anderson-Darling and Komolgorov-Smirnov

Summary (1 of 3)

- **We have discussed three popular CER methods**
  - **LOLS**
  - **MUPE/IRLS**
  - **ZMPE**
- **We have discussed Maximum Likelihood Estimation (MLE) and how each method has a connection to MLE when the residuals are multiplicative**
  - **LOLS is an MLE for lognormal residuals**
  - **MUPE is an MLE for gamma residuals (also it is a quasi-likelihood estimator)**
  - **ZMPE is similar to the MLE for normal residuals**

Summary (2 of 3)

- **We have also discussed the transformation controversy regarding LOLS**
- **We have presented logical/mathematical arguments and empirical evidence as to why the lognormal is the right model for the residuals**
- **We have presented a simple solution to avoid the transformation issue altogether by directly modeling the mean of the lognormal via MLEs - MRLN**
- **MRLN is easy to implement in Excel Solver or the MLE package in R**

Summary (3 of 3)

Wired	Tired	Expired
MRLN	ZMPE MUPE	LOLS

References (1 of 3)

1. Babyak, M.A., "What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models," Psychosomatic Medicine, 66, 411-421, 2004.

2. Bahadur, R.R., and L.J. Savage, "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," Annals of Mathematical Statistics 27, 1956, pp 1115-1122.

3. Basu, A. et al., "Comparing Alternative Models: Log Vs. Cox Proportional Hazard," Health Economics, 2004.

4. Book, S.A., and P.H. Young, "General-Error Regression for USCM-7 CER Development," The Aerospace Corporation, El Segundo, CA, 1995

5. Book, S.A., and P.H. Young, "General-Error Regression for Deriving Cost-Estimating Relationships," Journal of Cost Analysis, Fall 1997, pp. 1-28.

6. Book, S.A., and N.Y. Lao, "Deriving Minimum-Percentage-Error CERs Under Zero-Bias Constraints," The Aerospace Corporation, El Segundo, CA, July 1996.

7. Book, S.A., "IRLS/MUPE CERs Are Not MPE-ZPB CERs," Presented at the International Society for Parametric Analysts Annual Conference, Seattle, WA, May 23-26, 2006.

8. Book, S.A., and P.H. Young, "The Trouble with R^2," The Journal of Parametrics, Vol. 26, No. 1, Summer 2006, pp. 87-112.

9. Embrechts, P, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Fourth Printing, Berlin, 2003.

10. Eskew, H.L. and K.S. Lawler, "Correct and Incorrect Error Specifications in Statistical Cost Models," Journal of Cost Analysis, Spring 1994.

References (2 of 3)

11. Foussier, P. M. and P. Foussier, "Should We Use the Median Instead of the OLS?," Parametric World, Fall 2008, pp. 8-11.

12. Fu, L., and R. Moncher, "Severity Distributions for GLMs: Gamma or Lognormal?", 2004 CAS Spring Meeting, Colorado Springs, CO, 2004.

13. Gallin, J.H., "Net Migration and State Labor Market Dynamics," Journal of Labor Economics, 2004.

14. Garvey, P. R., Book, S.A., and Covert, R.P., *Probability Methods for Cost Uncertainty: A Systems Engineering Perspective*, CRC Press, Boca Raton, FL, 2016.

15. Goldberg, M.S., and A.E. Tuow, *Statistical Methods Learning Curves and Cost Analysis*, Institute for Operations Research and Management Sciences, Linthicum, MD, 2003.

16. Hu, S., "The Impact of Using Log-Error CERs Outside the Data Range and Ping Factor," Presented at the Annual Joint ISPA-SCEA Conference, Denver, CO, June, 2005.

17. Hu, S., "Fit, Rather Than Assume, a CER Error Distribution," Presented at the Annual ICEAA Professional Development and Training Workshop, New Orleans, LA, June 2013.

18. Ismail, N., and A.A. Jemain, "Comparison of Minimum Bias and Maximum Likelihood Methods for Claim Severity," Casualty Actuarial Society E-Forum, Winter 2009.

19. Kleiber, C., and S. Klotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley-Interscience, Hoboken, NJ, 2003.

20. Klugman, S.A., et al., *Loss Models*, 3rd Ed., John Wiley & Sons, Hoboken, 2008.

21. Lee, D.A., *The Cost Analyst's Companion*, Logistics Management Institute, McLean, VA, 1997.

References (3 of 3)

22. Mackenzie, D., "Cost Estimating Relationship Variance Study," AIAA Space Conference, Long Beach, CA, 2003.

23. Mackenzie, D., et al., "Top Level Spacecraft Cost Distribution Study," Joint Annual ISPA-SCEA Conference, Noordwijk, May, 2008.

24. Manning, W.G., and J. Mullahy, "Estimating Log Models: To Transform or Not to Transform?", Journal of Health Economics, 2001

25. Manning, W.G., et al., "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data," Journal of Health Economics, 2005.

26. Naval Center for Cost Analysis, *Cost Estimating Relationship Development Handbook*, 2016 (Draft).

27. Nelder, J. A., "Weighted Regression, Quantal Response Data, and Inverse Polynomials," Biometrics, Vol. 24 (1968), pages 979-985.

28. Seal, H.L., "Numerical Probabilities of Ruin When Expected Claim Numbers Are Large," Mitteilungen SVVM, 89-104, 1983.

29. Smart, C., "Bayesian Parametrics: How to Develop a CER

with Limited Data and Even Without Data," presented at the ICEAA Annual Conference, June 2014, Denver, CO.

30. Smart, C.B., "Covered with Oil: Incorporating Realism in Cost Risk Analysis," Journal of Cost Analysis and Parametrics, 8:3, 186-205, 2015.

31. Thomas, D. and A. Smith, *Joint Agency Cost Schedule Risk and Uncertainty Handbook*, 2014.

32. van Belle, G., *Statistical Rules of Thumb*, 2nd ed., Wiley, Hoboken New Jersey, 2008.

33. Wedderburn, R.W.M., "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," Biometrika, Vol. 61, Number 3 (1974), pages 439-447.