TECOLOTE RESEARCH, INC.
Bridging Engineering and Economics
Since 1973

Generalized Degrees of Freedom (GDF)

Addressing Equality Constraints

27 March 2017

Dr. Shu-Ping Hu

Supplemental Presentation

Derive GDF for Inequality Constraints

Dr. Boyan Jonov

■ Los Angeles ■ Washington, D.C. ■ Boston ■ Chantilly ■ Huntsville ■ Dayton ■ Santa Barbara
■ Albuquerque ■ Colorado Springs ■ Goddard Space Flight Center ■ Johnson Space Center ■ Ogden ■ Patuxent River ■ Washington Navy Yard
■ Ft. Meade ■ Ft. Monmouth ■ Dahlgren ■ Quantico ■ Cleveland ■ Montgomery ■ Silver Spring ■ San Diego ■ Tampa ■ Tacoma
■ Aberdeen ■ Oklahoma City ■ Eglin AFB ■ San Antonio ■ New Orleans ■ Denver ■ Vandenberg AFB

Outline

- Constrained Process (Background Info)
- Objectives
- Error Terms (Additive vs. Multiplicative)
- Multiplicative-Error Models
- ZMPE CER Unbiased?
- SPE Comparison: ZMPE vs. MUPE
- Definitions of DF and GDF
- Calculate Fit Statistics Using GDF
- Examples
- Conclusions

Note: SPE is standard percent error and MUPE stands for **m**inimum-**u**nbiased-**p**ercent error. Other acronyms will be explained on next page

Constrained Process (1/2) Introduction

- **Solver (an Excel add-in program) is a popular tool used to generate nonlinear cost estimating relationships (CER), especially when constraints are specified. A few examples are given below:**
  - Minimizing the sum of squared percentage errors under the Zero-Percentage Bias constraint (i.e., the **ZMPE** CER)
  - Minimizing the sum of squared residuals under the Zero-Percentage Bias constraint, (i.e., the mean of the % errors is zero)
  - Minimizing the sum of squared percentage errors or residuals in log space under the Zero-Bias constraint (i.e., the mean of the residuals is zero) using the Balance-Adjustment Factor (BAF)[1]
- **In the above examples, we may not have the degrees of freedom (DF) as given by the traditional definition when no constraints are specified**

1. Book, S., "Significant Reasons to Eschew Log-Log OLS Regression when Deriving Estimating Relationships," 2012 ISPA/SCEA Joint Annual Conference, Orlando, FL, 26-29 June.

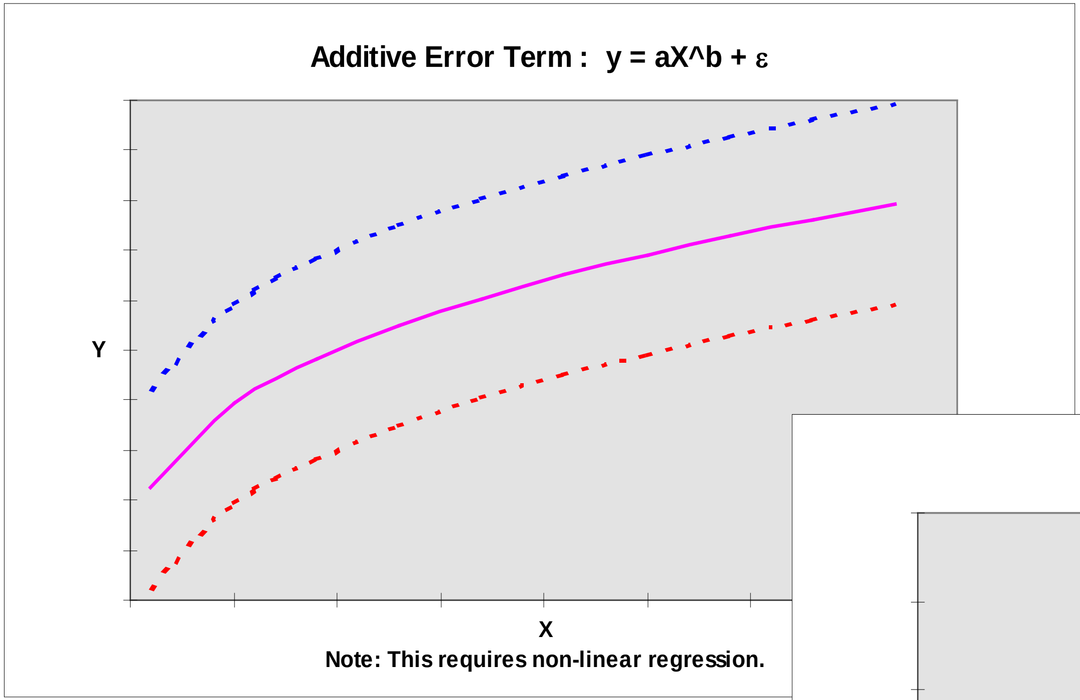
Constrained Process (2/2) Suggestions

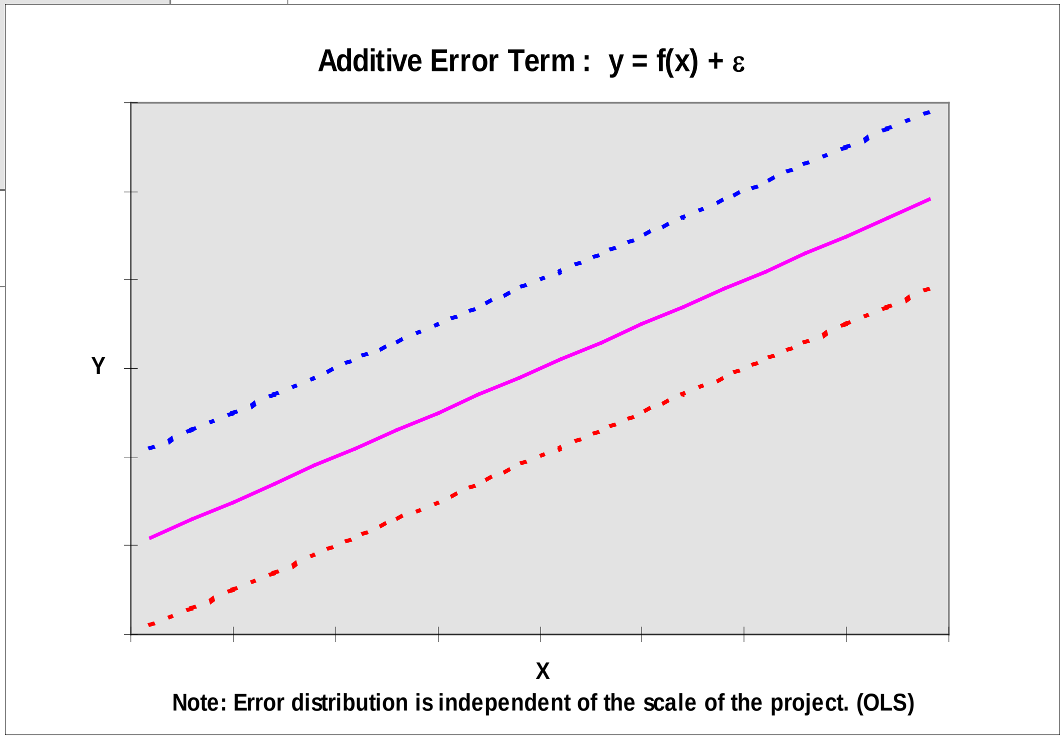
- **Do not abuse Solver**
  - Do not specify constraints excessively just because it is easy to do so in Solver
- **Explore different starting points to see if the solution stabilizes when using Solver**
  - Solver can be sensitive to starting points—different starting points may lead to different solutions
  - Solver can be trapped in local minima, especially when fitting complicated equations or the sample size is small
  - ZMPE equations seem to be less stable than their MUPE counterparts (Hu & Smith, 2007)
- **Specify "meaningful" constraints**
  - Make sure the constraints are necessary, logical, and statistically sound as DF can be reduced by additional constraints
- **Calculate the DF properly when constraints are specified**

Objectives

- **Explain why degrees of freedom (DF) should be adjusted if constraints are specified in the curve-fitting process**
- **Recommend a Generalized Degrees of Freedom (GDF) measure to compute fit statistics properly for constraint-driven equations**
- **Explain why ZMPE's standard error underestimates the spread of the CER error distribution**
  - We will illustrate how to calculate standard error properly for ZMPE CERs

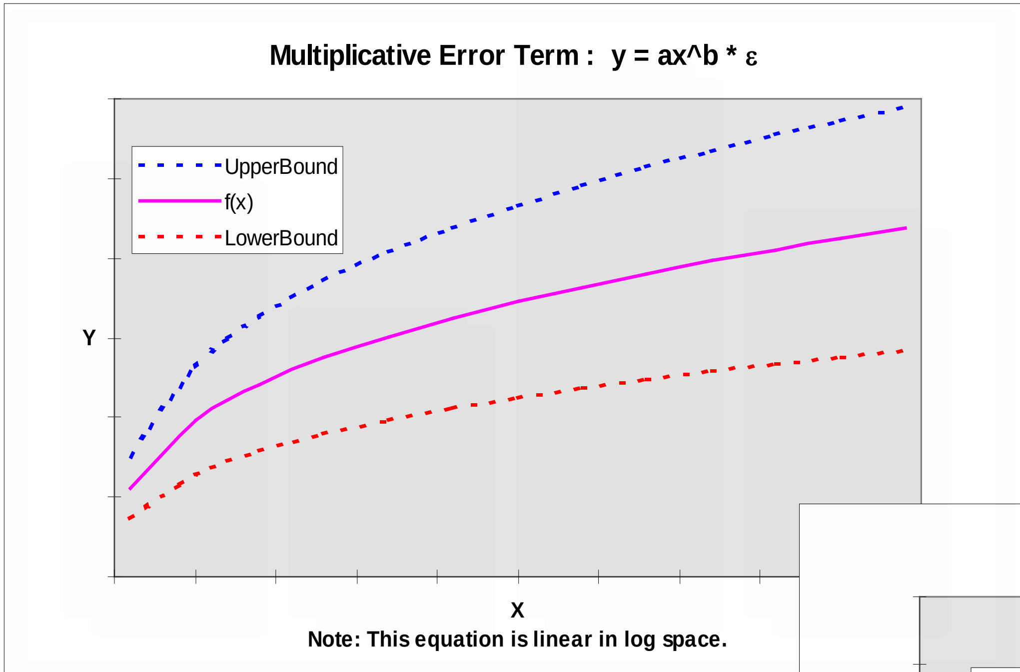
Additive Error Term: $Y = f(X) + \varepsilon$





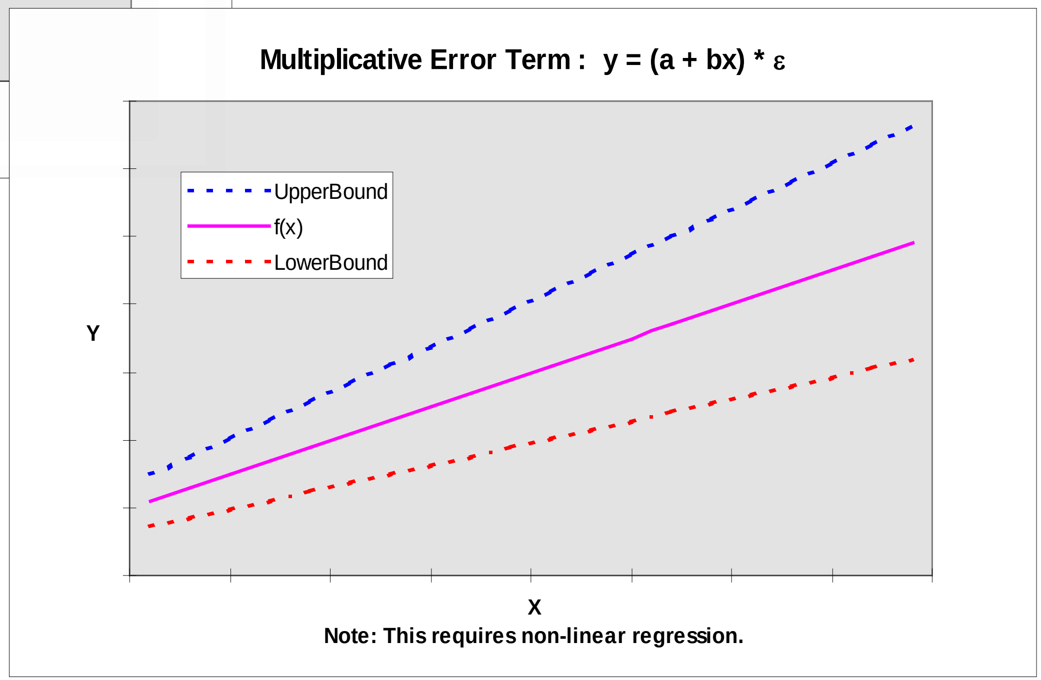
Cost variation is independent of the scale of the project

Multiplicative Error Term: $Y = f(X)*\varepsilon$



Multiplicative error assumption is appropriate when

- Errors in the dependent variable are believed to be proportional to the level of the function (the value of the variable)
- Dependent variable ranges over more than one order of magnitude



Cost variance is proportional to the scale of the project

Multiplicative Error Model: $Y = f(X)*\varepsilon$

- **Log-Error: $\varepsilon \sim LN(0, \sigma^2)$ ⇨ Least squares in log space**
  - Error = Log (Y) - Log f(X)
  - Minimize sum of squared errors; process is done in log space

If f(x) is linear in log space, it is termed log-linear or LOLS CER

- **MUPE: $E(\varepsilon) = 1$, $V(\varepsilon) = \sigma^2$ ⇨ Least squares in weighted space**
  - Error = (Y-f(X))/**f(X)**
  - Minimize sum of squared percentage errors iteratively (i.e., minimize $\Sigma_i \{(y_i - f(\mathbf{x}_i, \boldsymbol{\beta})/f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_{k-1}\}^2$, k is the iteration number)
  - MUPE (an iterative, weighted least squares) has zero sample bias

variance of error term

Note: $E((Y-f(X))/f(X)) = 0$
$V((Y-f(X))/f(X)) = \sigma^2$

- **ZMPE: $E(\varepsilon) = 1$, $V(\varepsilon) = \sigma^2$ ⇨ Least squares in weighted space**
  - Error = (Y-f(X))/**f(X)**
  - Minimize sum of squared % errors, $\Sigma_i \{(y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))/f(\mathbf{x}_i, \boldsymbol{\beta})\}^2$, with a constraint:

    $\Sigma_i (y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))/f(\mathbf{x}_i, \boldsymbol{\beta}) = 0$
  - ZMPE is a constrained minimization process
  - Average sample bias is eliminated by the constraint

We will focus on MUPE and ZMPE equations in this paper

ZMPE CER Unbiased?

Don't Know

- **Both MUPE and ZMPE methods have zero percentage bias (ZPB) for the sample data points:**

$$\frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{\hat{y}_i} = 0$$

y = actual value
ŷ = predicted value

  - For MUPE, this condition is achieved through the iterative minimization process; for ZMPE, ZPB is obtained by using a constraint

- **If a CER is unbiased, then $E(\hat{Y}) = E(Y) = f(X,\beta)$**

- **Does the "ZPB" property imply that the CER is unbiased?** *Not necessarily*
  - The ZPB constraint can be applied to any proposed methodologies (i.e., objective functions), but there is no guarantee that the CER result will be unbiased; namely, this condition "$E(\hat{Y}) = f(X,\beta)$" may not be satisfied

- **MUPE is the best linear unbiased estimator (BLUE) for linear models**
  - For linear CERs, e.g., $Y = (a + bX_1 + cX_2)*\varepsilon$, the MUPE method produces unbiased estimates of the parameters and the function mean; it also provides smaller variances for the parameters and for any linear function of the parameters

- **MUPE's parameter estimators are the quasi maximum likelihood estimators (QMLE) of the parameters; MUPE also provides consistent estimates of the parameters. ZMPE CERs, however, do not have statistical properties readily available.**

SPE Comparison: ZMPE vs. MUPE (1/4)

- **The standard percent error (SPE) for Y = f(X)*ε is given by**

$$SPE = \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2} * 100\%$$

SPE is CER's standard error of estimate, which is used to measure the model's overall error of estimation. It is the one-sigma spread of the MUPE or ZMPE CER.

  - n = sample size, p = total # of estimated parameters, y = actual value, and ŷ = predicted value

- **SPE^2 (i.e., MSE) is used to estimate σ^2, the variance of ε**
- **$SPE_{(ZMPE)} \le SPE_{(MUPE)}$**

Equal sign holds only for simple factor equations

  - ZMPE always produces a **smaller** SPE when compared to MUPE except for simple factor CERs (Book, 2006)

- **Is a smaller SPE better?**
  - **No**, not necessarily. If it is true, we should develop MPE CERs, which are proven to be over-estimating (see Hu and Sjovoid, 1994)
  - Beware of using SPE alone for selecting CERs; we should also evaluate other useful stats (see Hu, 2010)

SPE Comparison: ZMPE vs. MUPE (2/4)

$E(SPE^2_{(ZMPE)}) \le E(SPE^2_{(MUPE)}) = \sigma^2$

Q: Is ZMPE's SPE^2 (i.e., MSE) an unbiased estimator of σ^2? *No*

(I) When the CER is linear:

- **MUPE's SSE = $\Sigma_i w_i(y_i - \hat{y}_i)^2 = Z'(I - H)Z$**
  - MUPE can be converted to OLS in weighted space; w_i (= $1/(\hat{y}_i)^2$) is the weighting factor of the i^{th} observation
  - **Z** is the new vector variable in the weighted space ($z_i = \sqrt{w_i} y_i$). V(**Z**) = **I**σ^2 and **H** is **Z**'s hat matrix. See Morrison (1983) & Draper and Smith (1981).
- **For MUPE CERs: E(SSE) = σ^2(n – p) and $E(SPE^2_{(MUPE)}) = \sigma^2$**

  See proof in the backup section
  - $E(SSE/(n-p)) = E(MSE) = E(SPE^2_{(MUPE)}) = \sigma^2$
  - This equation is true regardless of the distribution type
  - This is an approximation if the CER is nonlinear

  Equal sign holds only for simple factor equations
- **ZMPE's SPE underestimates the true σ,** except for simple factor CERs
  - Since $SPE^2_{(ZMPE)} \le SPE^2_{(MUPE)}$, $E(SPE^2_{(ZMPE)}) \le E(SPE^2_{(MUPE)}) = \sigma^2$

Caution: ZMPE's SPE underestimates the spread of the CER error distribution

SPE Comparison: ZMPE vs. MUPE (3/4)

$E(SPE^2_{(ZMPE)}) \le E(SPE^2_{(MUPE)}) = \sigma^2$

Q: Is ZMPE's SPE^2 (i.e., MSE) an unbiased estimator of σ^2? *No*

(II) When the CER error term is also normally distributed:

- **MUPE's SSE = $\Sigma w_i(y_i - \hat{y}_i)^2 \sim \sigma^2\chi^2_{(n-p)}$**
  - See Morrison (1983) and Draper and Smith (1981) for details
  - The proof is also given in the backup section
- **For MUPE CERs: $E(SSE) = E(\sigma^2\chi^2_{(n-p)}) = \sigma^2(n - p)$**
  - $E(SSE/(n-p)) = E(MSE) = E(SPE^2_{(MUPE)}) = \sigma^2$
- **ZMPE's SPE underestimates the true σ,** except for simple factor CERs
  - Since $SPE^2_{(ZMPE)} \le SPE^2_{(MUPE)}$, $E(SPE^2_{(ZMPE)}) \le E(SPE^2_{(MUPE)}) = \sigma^2$
  - The equal sign holds only for simple factor equations

Caution: ZMPE's SPE underestimates the spread of the CER error distribution

SPE Comparison: ZMPE vs. MUPE (4/4)

Concerns About Underestimating CER Errors

- **The fitted regression coefficients may falsely become significant due to the biased low SPE**
- **Given $E(SPE^2_{(ZMPE)}) < \sigma^2$** (except for simple factor CERs), **using ZMPE CERs in cost uncertainty analysis may unduly tighten the S-curve**
  - Prediction Intervals (PI) are specified for cost uncertainty analysis
  - The smaller the SPE, the tighter the PI becomes
    - A PI is a function of the standard error of the regression (e.g., SPE), the sample size, the "distance" of the estimating point from the center of the database used to generate the CER, etc.
    - For a simple linear MUPE: $PI = \hat{y}_0(when\ X = x_0) \pm t_{(\alpha/2, n-2)} SPE \sqrt{\hat{y}_0^2 + \frac{1}{\sum w_i} + \frac{((x_0 - \bar{x}_w)/S_{wx})^2}{\sum w_i}}$
  - The impact on the risk session can be substantial when using underestimated SPEs in numerous work breakdown structure (WBS) elements

Q: Why is ZMPE's SPE biased low?

A: Its degrees of freedom did not get adjusted properly

Definition – Degrees of Freedom (1/2)

- **DF = number of data points (n) - number of parameters estimated in the regression equation (p)**
- **DF is used to characterize the number of "independent" pieces of information contained in a statistic**

  The term "independent" means "free to vary"
- **Any sum of squares (SS) has a DF associated with it**
  - DF indicates how many pieces of independent information from the n independent observations are needed to compile the sum of squares
- **Given a simple OLS model: $Y = \alpha + \beta X + \varepsilon$**
  - $SST = \Sigma_i(y_i - \bar{y})^2$ has (n – 1) DF because $\Sigma_i(y_i - \bar{y}) = 0$
  - $SSR = b^2\Sigma_i(x_i - \bar{x})^2 = (SS_{xy})^2/SS_{xx}$ has one DF
  - $SSE = \Sigma_i(y_i - \hat{y}_i)^2$ has (n – 2) DF as intercept and slope are estimated

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$b = \hat{\beta} = \frac{SS_{xy}}{SS_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

The DF for SSE is the excess data points that can be used to judge the quality of the fit

Definition – Degrees of Freedom (2/2) (a ZMPE Model)

- **The objective function, F, for "Y = (α + βX)*ε" is** $F = \sum_{i=1}^{n}\left(\frac{y_i}{\alpha + \beta * x_i} - 1\right)^2$
- **The solutions for α and β are derived by solving the two normal equations:**
  - $\frac{\partial F}{\partial \alpha} = 0 \rightarrow \sum_{i=1}^{n} \frac{y_i(y_i - \alpha - \beta * x_i)}{(\alpha + \beta * x_i)^3} = 0$
  - $\frac{\partial F}{\partial \beta} = 0 \rightarrow \sum_{i=1}^{n} \frac{y_i\, x_i(y_i - \alpha - \beta * x_i)}{(\alpha + \beta * x_i)^3} = 0$

  } **two** constraints

  Each normal equation in the system can be viewed as a constraint that restricts one DF
- **The total number of constraints becomes three when including an additional constraint in the system:**

$$\sum_{i=1}^{n} \frac{y_i - \alpha - \beta * x_i}{\alpha + \beta * x_i} = \sum_{i=1}^{n} \frac{y_i}{\alpha + \beta * x_i} - 1 = 0$$

← **ZMPE** CER applies the Zero-Percentage Bias constraint to the curve-fitting process.

- **Adjust DF accordingly for additional constraints: DF = n – 2 – 1**
  - Adjust DF when solutions are derived in a constrained domain. When additional constraints are present, we cannot search as freely as we can in an unconstrained domain to find a solution, which results in a loss of DF.

Definition – Generalized DF:
Adjusting DF for Constraint driven equations

- **DF should be adjusted to reflect (1) additional constraints and (2) redundancies in CER/PER development**
  - One restriction is equivalent to a loss of one DF
  - We should take redundancy into account when counting the DF
    - For example, if two additional constraints are specified for a regression model but one constraint can be derived by the other, then we should only count a loss of one DF rather than two. Another example, if a constraint is directly related to the normal equations, then it does not count towards a loss of DF
  - If an equality constraint is not redundant, it is considered to be **effective**

n = sample size
p = total # of estimated parameters

- **Generalized DF (GDF) = n – p – (# constraints) + (# redundancies)**
  = n – p – (# effective constraints)

- **GDF =** $\begin{cases} n - p & \text{for MUPE} \\ n - p - 1 & \text{for ZMPE (except for simple factor CERs)} \end{cases}$

See the last slide in the backup section

DF should be adjusted for ZMPE CERs

Q: Why is GDF important?

A: DF is the basis for fit measures. All fit statistics (e.g., SEE and SPE) should be updated by GDF if constraints are included.

Note: SEE and SPE are also used in cost uncertainty analysis.

Calculate SEE and SPE Using GDF

- **Standard error of estimate (SEE) for an additive-error model is given by:**

$$SEE = \sqrt{\frac{1}{GDF}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

y = actual value
ŷ = predicted value

- **Standard percent error (SPE) for a multiplicative-error model is given by:**

$$SPE = \sqrt{\frac{1}{GDF}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2} * 100\%$$

SPE, as well as SEE, is CER's standard error of estimate, which is used to measure the model's overall error of estimation

- **GDF = n – p when no constraints are specified**
- **GDF = n – p for MUPE; GDF = n – p – 1 for ZMPE** (except for factor CERs)
- **$SPE_{(ZMPE)}$ =**

$$\begin{cases} SPE_c\sqrt{\dfrac{n-p}{n-p-1}} \\ SPE_c \quad \text{for simple factor equations} \end{cases}$$

SPE_c stands for the current calculation, not using GDF

☺ *We can now compare MUPE's SPE with ZMPE's SPE*

Calculate Adjusted R^2 Using GDF

- **Adjusted R^2 for Additive-Error CERs, i.e., $Y = f(X)+\varepsilon$:**

$$Adj.\,R^2\,(Additive) = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / GDF}{\sum (y_i - \bar{y})^2 / (n-1)} = \frac{SEE_{\bar{Y}}^{\;2} - SEE_f^{\;2}}{SEE_{\bar{Y}}^{\;2}}$$

- **Adjusted R^2 for Multiplicative-Error CERs, i.e., $Y = f(X)*\varepsilon$:**

$$Adj.\,R^2\,(MUPE/ZMPE) = 1 - \frac{\sum ((y_i - \hat{y}_i)/\hat{y}_i)^2 / GDF}{\sum ((y_i - \bar{y})/\bar{y})^2 / (n-1)} = \frac{SPE_{\bar{Y}}^{\;2} - SPE_f^{\;2}}{SPE_{\bar{Y}}^{\;2}}$$

Adj. R^2 measures the % difference between the CER's estimated variance and the sample variance of an average CER. For example, if a CER's estimated variance is 0.35 while the sample variance of an average CER is 1.4, then the CER's variance is only 25% of the variance of an average CER. This reduction of variance, 75%, is the Adjusted R^2. The reduction of variance is considered to be an improvement when applying the CER (see Hu, 2010).

  - This adjusted R^2 compares MUPE and ZMPE's SPE^2 to its baseline (i.e., SPE^2 of an average CER, $\bar{Y}$); it is more pertinent to the fitting method
  - This measure "Adjusted $R^2_{(MUPE/ZMPE)}$" puts SPE^2 in perspective

- **GDF = n – p when no constraints are specified**
- **GDF = n – p for MUPE; GDF = n – p – 1 for ZMPE** (except for factor CERs)

☺ ***We can compare MUPE's Adjusted R^2 with ZMPE's Adjusted R^2***

Modify GRSQ (r^2) Using GDF (Within a CER)

- **Generalized R^2 (GRSQ): GRSQ is Pearson's r^2 between a CER's actual {y} and its predicted {ŷ} values in unit space, i.e., GRSQ = $r^2(y, \hat{y})$**

$$GRSQ = r^2(y,\hat{y}) = \frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})\left(\hat{y}_i-\bar{\hat{y}}\right)\right)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}\left(\hat{y}_i-\bar{\hat{y}}\right)^2}$$

$$\text{Definition}: r\,(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

  - GRSQ takes neither the sample size nor degrees of freedom into account

- **We should modify GRSQ (denoted by r^2) for DF using GDF:**

$$GRSQ_{(GDF)} = \begin{cases} r^2-(1-r^2)*\dfrac{p-1}{GDF} & if\ p>1 \\ r^2-(1-r^2)*\dfrac{1}{n-1} & if\ p=1 \end{cases}$$

Note: GRSQ, as well as *Pearson's correlation*, only measures the linear association between two sets of numbers rather than the actual deviation between them.

In OLS: Adjusted $R^2 = R^2 - (1-R^2)*(p-1)/(n-p)$

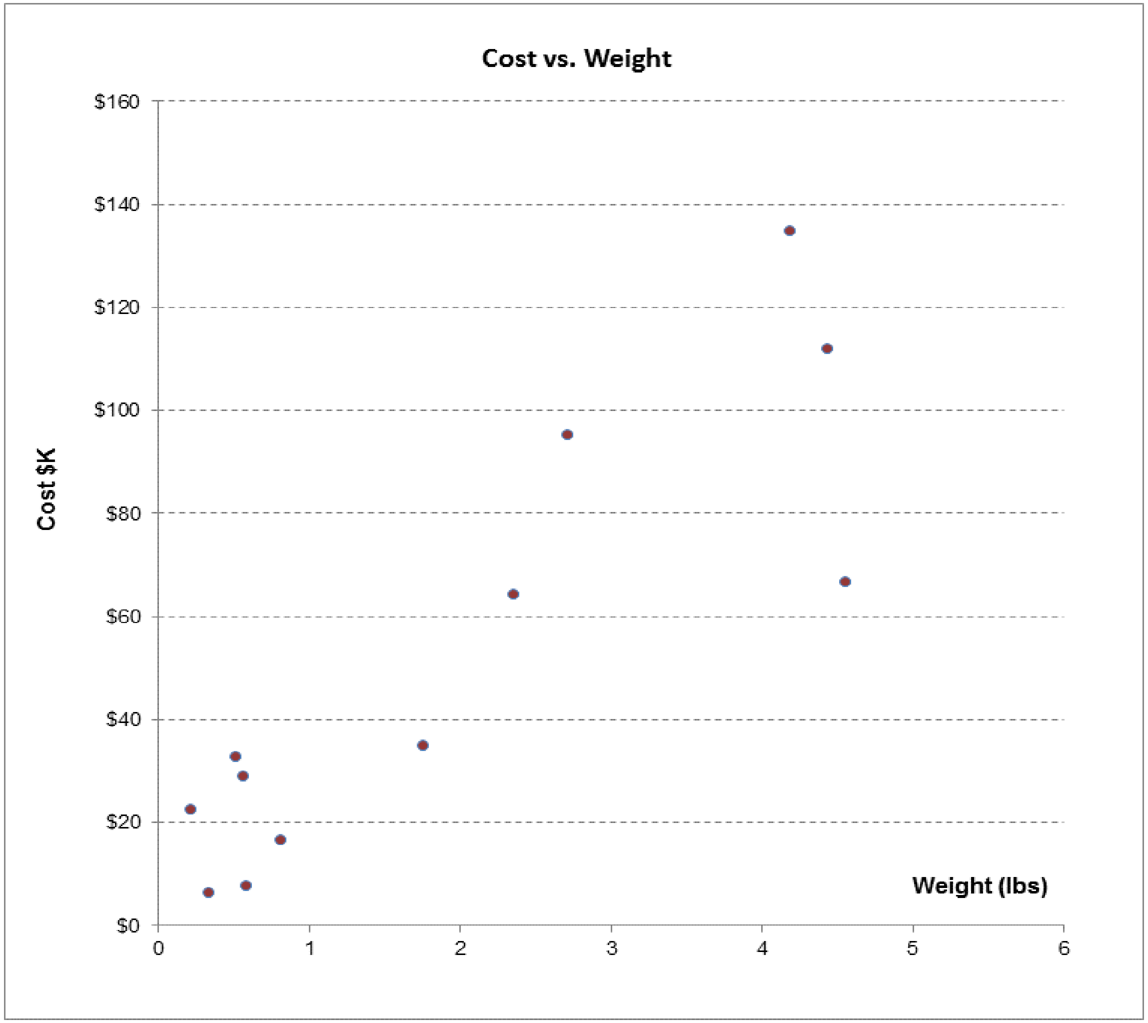
  - GDF = n – p if no constraints are specified
  - GDF = n – p for MUPE; GDF = n – p – 1 for ZMPE (except for factor CERs)

- **This modified GRSQ ($GRSQ_{(GDF)}$) takes the DF, numbers of estimated parameters, as well as the constraints, into consideration**

Example – Weight-Based CERs (1/3)

- A hypothetical data set to predict the cost of a black box using weight

Data Point	Cost $K	Weight (lbs)
Obs 1	135.0	4.18
Obs 2	6.5	0.32
Obs 3	8.0	0.57
Obs 4	64.6	2.34
Obs 5	32.9	0.50
Obs 6	95.4	2.70
Obs 7	67.0	4.54
Obs 8	112.2	4.42
Obs 9	29.2	0.55
Obs 10	22.7	0.20
Obs 11	16.9	0.80
Obs 12	35.0	1.75



Example – Weight-Based CER (2/3)

- **Four different CERs are generated using both MUPE and ZMPE methods**

ZMPE	CERs	SPE	$SPE_{(GDF)}$	Adj. R^2	Adj. $R^2_{(GDF)}$	GRSQ	$GRSQ_{(GDF)}$
Linear	12.794 + 19.16*Weight	45.7%	48.2%	68.9%	65.4%	77.5%	75.0%
LogLinear	36.889*Weight^(0.5882)	48.9%	51.5%	66.9%	60.4%	77.4%	74.9%
Semi-Log	17.881*(1.5314)^Weight	47.1%	49.7%	64.4%	63.2%	68.9%	65.4%
Triad	16.785+12.034*Weight^(1.349)	47.5%	50.4%	66.4%	62.2%	75.5%	69.4%

MUPE	CERs	SPE	Adj. R^2	GRSQ
Linear	10.528 + 21.2975*Weight	46.5%	67.8%	77.5%
LogLinear	36.2953*Weight^(0.6635)	50.2%	65.5%	77.7%
Semi-Log	16.756*(1.5835)^Weight	47.4%	66.5%	67.0%
Triad	15.026+12.9863*Weight^(1.386)	48.4%	65.1%	75.3%

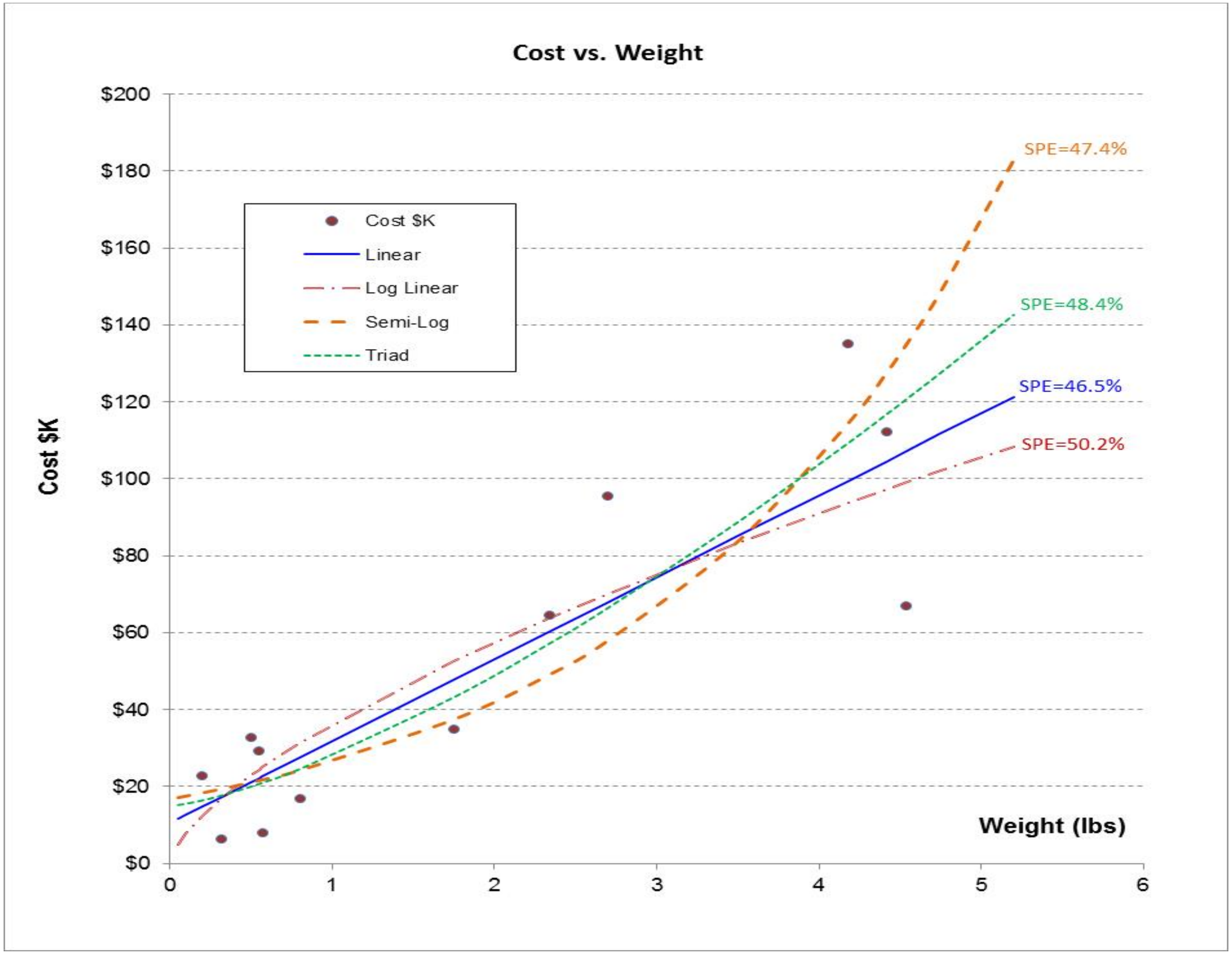
Although the semi-log equation is tighter than the log-linear and triad equations, its GRSQ is much worse than those of the log-linear and triad CERs.

Caution: Beware of using GRSQ alone for selecting CERs as it only measures the linear association between y and ŷ, not the actual difference between them.

- **The unadjusted SPE measures for these ZMPE CERs are all smaller than their respective SPE measures generated by the MUPE method**
- **ZMPE's updated SPE measures using GDF become larger than their MUPE counter-part (numbers in dark red)**
- **ZMPE's updated adjusted R^2 and GRSQ are all smaller than those generated by the MUPE method when GDF is applied (numbers in purple and green)**
- **MUPE outperforms ZMPE based upon all three statistics (SPE, Adj. R^2, & GRSQ)**

Example – Weight-Based CER (3/3) Scatter Plot for MUPE CERs

- **The linear CER seems to be the best choice among the four CERs**
- **Use the semi-log equation with caution because it goes up exponentially**



Conclusions

- **Make sure the constraints (if any) are meaningful, logical, and statistically sound when adding them to the curve-fitting process**
  - Explore different starting points in Solver to ensure the solution stabilizes
- **Adjust DF for constraint-driven equations**
- **Define GDF = n – p – (# effective constraints)**
  - Subtract one from DF for ZMPE equations except for simple factor CERs
- **Calculate SEE, SPE, Adjusted R^2, and GRSQ (r^2) using GDF**
- **No need to adjust DF and goodness-of-fit measures for MUPE CERs**
- **ZMPE's SPE underestimates the spread of the CER error distribution**
  - Using ZMPE CERs without adjustment in cost uncertainty analysis may unduly tighten the S-curve

Future Study Items for Constraint Driven Regression

- **Adding excessive constraints into the process may cause the unknown parameters in the CER to be determined completely by the constraints. (See the last slide in the backup section for an example.) When it happens, there is no need to run regression analysis. How do we define DF properly for this case?**
  - If # of constraints = # of estimated parameters, we do not use any curve-fitting methods to derive a solution. Consequently, we may have **no** degrees of freedom left to judge the quality of the fit due to lack of regression.
- **An inequality constraint may not be treated the same as an equality constraint. How to treat inequality constraints in the curve-fitting or distribution-finding process is another topic.**

References

- Book, S. A. (2006, May 23-26). IRLS/MUPE CERs are not MPE-ZPB CERs. 28th Annual ISPA International Conference, Seattle WA.
- Book, S. A. (2012, June 26-29). Significant Reasons to Eschew Log-Log OLS Regression when Deriving Estimating Relationships. ISPA/SCEA International Conference, Orlando, FL.
- Book, S. A. and N. Y. Lao. (1999). Minimum-Percentage-Error Regression under Zero-Bias Constraints. Proceedings of the 4th Annual U.S. Army Conference on Applied Statistics, U.S. Army Research Laboratory, Report No. ARL-SR-84, pp. 47-56.
- Draper, N. R. and H. Smith. (1981). Applied Regression Analysis (2nd ed.). New York, NY: John Wiley & Sons.
- Hu, S. (2001, June 12-15). The Minimum-Unbiased-Percentage-Error (MUPE) Method in CER Development. 3rd Joint Annual ISPA/SCEA International Conference, Vienna, VA.
- Hu, S. (2010). R^2 versus r^2. Journal of Cost Analysis and Parametrics, Volume 3, Issue 2, 13-27.
- Hu, S. (2010, June 8-11). Simple Mean, Weighted Mean, or Geometric Mean? ISPA/SCEA International Conference, San Diego, CA.
- Hu, S. (2016). Generalized Degrees of Freedom (GDF). Journal of Cost Analysis and Parametrics, Volume 9, Issue 2, 93-111
- Hu, S. and A. R. Sjovold. (1994, June 7-9). Multiplicative Error Regression Techniques. 62nd MORS Symposium, Colorado Springs, Colorado.
- Hu, S. and A. Smith. (2007, June 12-15). Why ZMPE When You Can MUPE. 6th Joint Annual ISPA/SCEA International Conference, New Orleans, LA.
- Morrison, D. F. (1983). Applied Linear Statistical Methods. New Jersey: Prentice-Hall, Inc.
- Searle, S. R. (1981). Linear Models. New York: John Wiley & Sons.
- Seber, G. A. F. and C. J. Wild. (1989). Nonlinear Regression. New York: John Wiley & Sons. pp. 37, 46, 86-88.
- Weisberg, S. (1985). Applied Linear Regression 2nd ed. New York: John Wiley & Sons. pp. 87-88.
- Wedderburn, R. W. M. (1974, December). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. Biometrika, Vol. 61, Issue 3, pp. 439-447.
- Young, P. H. (1991, September). GERM: Generalized Error Regression Model. 25th Annual DoD Cost Analysis Symposium, Leesburg, VA.
- Young, P. H. (1992, September). Generalized Coefficient of Determination. 26th Annual DoD Cost Analysis Symposium, Leesburg VA.

Derive GDF for Inequality Constraints

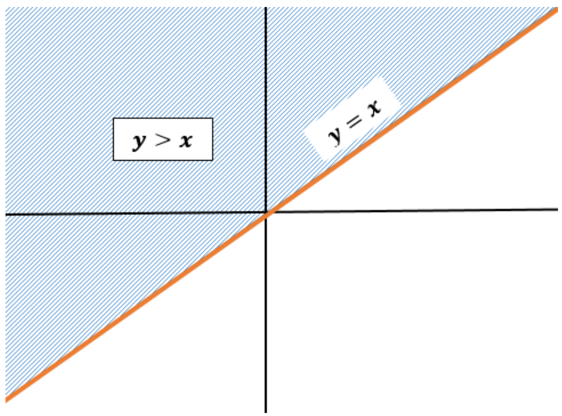
Dr. Boyan Jonov

Inequality Constraints

- **What we have learned so far …**
  - DF should be adjusted when constraints are introduced into the curve fitting process.
  - An equality constraint leads to a loss of one DF.
- **What if we have an inequality constraint?**
  - An inequality constraint can reduce the level of independence of the information contained in a statistic.
  - Therefore, DF should be adjusted (reduced) correspondingly.
  - Yet, an inequality is less restrictive than an equality constraint.
  - The question is … exactly how much DF do we lose?

Inequality Constraint Example

- **Let's consider this simple inequality constraint $y - x > 0$**
  - Two independent variables: x and y
  - The unconstrained domain is the entire x-y plane
  - The constrained domain is the "half" plane above the line y = x
  - Suggested DF adjustment: we lose ½ degree of freedom



- **Relate to the equality constraint $y - x = 0$**
  - The constrained domain is the line y = x
  - We lose 1 degree of freedom

Conclusion

- Our hypothesis is that an inequality constraint could result in a fractional loss of DF.
- We need to develop a measure that precisely captures the restrictive effect of inequality constraints.
- Our goal is to provide a more generalized definition of GDF that incorporates inequality constraints.

Backup Slides

SPE Comparison: ZMPE vs. MUPE

Formula to Prove $E(SSE_{(MUPE)}) = \sigma^2(n-p)$

- **If $E(y) = \mu$, $V(y) = \Sigma$ ➔ $E(y'Ay) = \mu'A\mu + trace(\Sigma A)$**

See Morrison (1983) for details

  - **y** is a vector of random variables; apostrophe is used to denote transpose
  - μ is **y**'s expected value; Σ is **y**'s variance/covariance matrix ($\Sigma = E(\mathbf{y}-\mu)(\mathbf{y}-\mu)'$)
  - **A** is a symmetric matrix
  - trace is defined to be the sum of the diagonal elements of a square matrix
  - This equation holds regardless of the distribution assumption

- **For WLS: If $y = X\beta + \varepsilon$, where $E(\varepsilon) = 0$ and $V(\varepsilon) = V\sigma^2$ ➔ there exists a nonsingular symmetric matrix P such that PP = V**
  - $\mathbf{Z} = \mathbf{P}^{-1}\mathbf{y} = \mathbf{P}^{-1}\mathbf{X}\beta + \mathbf{P}^{-1}\varepsilon = \mathbf{Q}\beta + \mathbf{P}^{-1}\varepsilon$ & $V(\mathbf{Z}) = V(\mathbf{P}^{-1}\varepsilon) = \mathbf{P}^{-1}V(\varepsilon)(\mathbf{P}^{-1})'\sigma^2 = \mathbf{I}\sigma^2$
  - $SSE_{(MUPE)}$ $= (\mathbf{Z}-\hat{\mathbf{Z}})'(\mathbf{Z}-\hat{\mathbf{Z}}) = (\mathbf{Z}-\mathbf{Q}\hat{\beta})'(\mathbf{Z}-\mathbf{Q}\hat{\beta}) = (\mathbf{Z}-\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Z})'(\mathbf{Z}-\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Z})$ $= (\mathbf{Z}-\mathbf{HZ})'(\mathbf{Z}-\mathbf{HZ}) = \mathbf{Z}'(\mathbf{I}-\mathbf{H})\mathbf{Z} = (\mathbf{Y}-\mathbf{X}\hat{\beta})'\mathbf{P}^{-1}\mathbf{P}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\beta})$
  - MUPE is a WLS. $\mathbf{P}^{-1}\mathbf{P}^{-1} = \mathbf{V}^{-1} = \mathbf{W}$ (weighting matrix), $\mathbf{H} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ (hat matrix), and $\hat{\beta}$ is the LS solution for the unknown parameters.
  - $E(SSE) = E(\mathbf{Z}'(\mathbf{I}-\mathbf{H})\mathbf{Z}) = (E(\mathbf{Z}))'(\mathbf{I}-\mathbf{H})E(\mathbf{Z}) + trace[(\mathbf{I}-\mathbf{H})V(\mathbf{Z})] = \beta'\mathbf{Q}'(\mathbf{I}-\mathbf{H})\mathbf{Q}\beta + \sigma^2(n - trace[\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}']) = 0 + \sigma^2(n - trace[(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Q}]) = \sigma^2(n-p)$

Theorem to Prove $SSE_{(MUPE)} \sim \sigma^2\chi^2_{(n\text{-}p)}$

- **Given a multivariate normal distribution y ~ N(μ,Σ), the quadratic form Q = y'Ay has the noncentral χ^2 distribution with r DF & noncentrality parameter μ'Aμ if and only if AΣ is an idempotent matrix**
  - μ is **y**'s expected value
  - Σ is **y**'s variance/covariance matrix (i.e., Σ = E(**y** – μ)(**y** – μ)')
  - A is a symmetric matrix of rank r
  - See Searle (1973) and Cochran (1934) for the proof of the theorem
- **If y = Xβ + ε, where ε ~ N(0,Vσ²) ➔ $P^{-1}y=P^{-1}X\beta + P^{-1}\varepsilon$ and $P^{-1}\varepsilon \sim N(0, I\sigma^2)$**
  - **PP** = **V**; $\mathbf{V}^{-1}$ (also denoted by **W**) is a weighting matrix
- **The SSE for MUPE CER can be expressed below**

$$
\begin{aligned}
SSE &= (\mathbf{P}^{-1}\mathbf{Y}-\mathbf{P}^{-1}\mathbf{X}\hat{\boldsymbol\beta})'(\mathbf{P}^{-1}\mathbf{Y}-\mathbf{P}^{-1}\mathbf{X}\hat{\boldsymbol\beta}) = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol\beta})'\mathbf{P}^{-1}\mathbf{P}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol\beta})\\
&= (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol\beta})'\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol\beta}) = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}-\hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}-\mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol\beta}+\hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol\beta}\\
&= \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}-\hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \mathbf{Y}'(\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{Y}\\
&= \mathbf{Y}'(\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{R}\mathbf{V}^{-1})\mathbf{Y} = \mathbf{Y}'(\mathbf{W}-\mathbf{W}\mathbf{R}\mathbf{W})\mathbf{Y}
\end{aligned}
$$

- **Since $(V^{-1} – V^{-1}RV^{-1})*V$ is an idempotent matrix and the noncentrality parameter is zero, SSE follows a χ^2 distribution with (n – p) DF**

Modify Pearson's r by Sample Size (Between CERs)

- **Consider sample size when applying correlation to a cost risk model**
  - A correlation of 0.8 derived from 30 data points is much more significant than the same correlation computed from just 5 data points
  - Neither the sample size nor the degrees of freedom (DF) adjustment is accounted for in the Pearson's correlation formula (see below)
- **In OLS: Adjusted $R^2 = 1 – (1–R^2)*(n–1)/(n–p) = R^2 – (1–R^2)*(p–1)/(n–p)$**
- **Compute Pearson's Adjusted r (Adj. r) for two sets of numbers, e.g., "cost vs. cost" or %error vs. %error correlation between two CERs:**

Not CER related

$$\bar{r}^2 = r^2 - (1 - r^2)/(n - 2)$$

Assume **p = 2** in this situation

$$Adj.\ r = \begin{cases} \text{sign}(r) * \bar{r} & if\ \bar{r}^2 > 0 \\ 0 & if\ \bar{r}^2 \le 0 \end{cases}$$

$$\text{Definition}: r\,(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- **Instead of sample correlations, suggest using Pearson's Adjusted r (Adj. r) for correlation analysis**

Constraints and DF adjustment are not relevant to derive *Adj. r* if not comparing y versus ŷ for a CER

Example of a Factor CER (y = βx*ε)

- **The MPE solution for β is derived by minimizing the objective function:**

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

- **The MPE solution for β is given by** $\hat{\beta} = \frac{\sum_{i=1}^{n} (y_i / x_i)^2}{\sum_{i=1}^{n} (y_i / x_i)}$

- **The ZMPE solution for β is given by** $\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} (y_i / x_i)$

- **Note: The ZMPE solution is derived by the ZMPE constraint, <u>not</u> the minimization process:**

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

See Hu (June 2010)

DF should be subtracted by one for ZMPE CERs except for simple factor CERs