NORTHROP GRUMMAN

DEFINING THE FUTURE

"To b or Not to b"
The y-intercept in Cost Estimation

Presented at SCEA, June 2007

Richard L. Coleman, Jessica R. Summerville, Peter J. Braxton, Bethia L. Cullis, Eric R. Druker

Northrop Grumman Corporation

The Society of Cost Estimating and Analysis

Background

- **We have found errors concerning the y-intercept to be a widely-occurring problem:**
  - Rampant among engineers, pricers, and others who use factors, rates, and data-based costing techniques
  - Less common (but not non-existent) among cost estimators
    - More common among engineers who overemphasize engineering or physics as the basis of CERs
- **In a prior paper[1] we discussed the dangers posed by the use of simple ratios of parameters in adjusting analogies**
- **In this paper we will discuss the implications of ignoring or suppressing y-intercepts:**
  - In cost estimating relationships (CERs)
  - In rates
  - In metrics or thumb rules
  - In analogies
- **First, let us consider the schools of thought in cost**

1. *Analogies: Techniques for Adjusting Them*, R. L. Coleman, J. R. Summerville, S. S. Gupta, So. MD SCEA Chapter, Feb 2004, ASC/Industry Cost/Schedule Workshop, Apr 04, SCEA 2004, MORS 2004

3 Schools of Thought on y-Intercepts

Y-Intercepts – 3 Schools of Thought

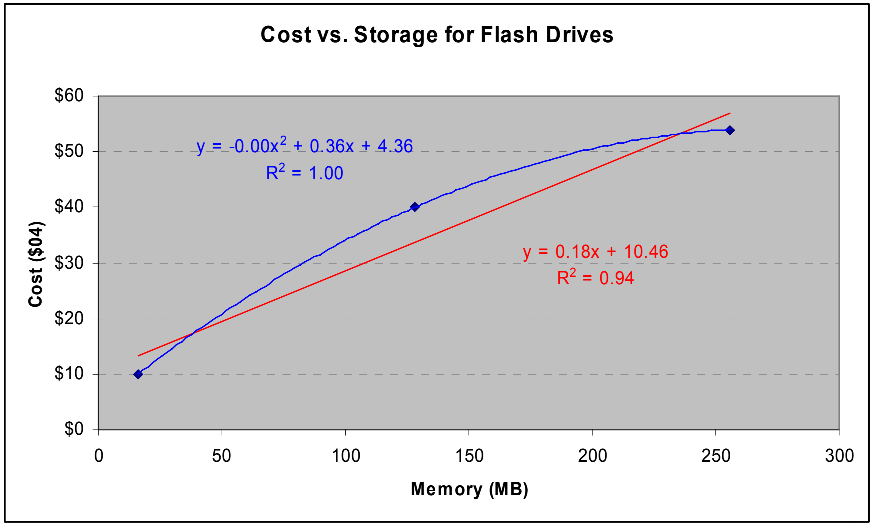
- **We have said that the y-intercept is sometimes missed inadvertently**
- **We will now discuss beliefs about the y-intercept when estimators do think of it**
- **We find that the y-intercept is a litmus test among cost estimators. There are about three schools of thought:**
  1. CERs must pass through the origin
  2. CERs which do not pass through the origin must have an explicable y-intercept
  3. CERs must be statistically derived, and if done properly, the y-intercept is just "what it is"
- **We'll discuss each briefly and then assume you are of school 2 or 3**

Warnings:
1- Almost anyone is from one of these schools of thought at heart. The writers are no exception.
2- The gulf between these schools is wide.

NORTHROP GRUMMAN

"CERs must pass through the origin"

- **Typical arguments**
  - "If I buy no product, I spend no money"
- **Pros:**
  - Sounds good
- **Cons:**
  - Doesn't seem to match the data. E. g., the price of Flash Drives
  - This is a simple data set, note that the two most likely curves both have a y-intercept between about $4 and $10



As a side bar … is the second relation ship better?

NORTHROP GRUMMAN

"Y-Intercepts must make sense"

- **Typical arguments**
  - "There must be physics-based arguments for CERs"
- **Pros:**
  - Helpful to think about it, within reason
- **Cons:**
  - If practiced to the extreme, good CERs can be rejected just because we do not yet understand them
  - Engineers, who hate cost estimation, can usually talk the analyst to a full stop

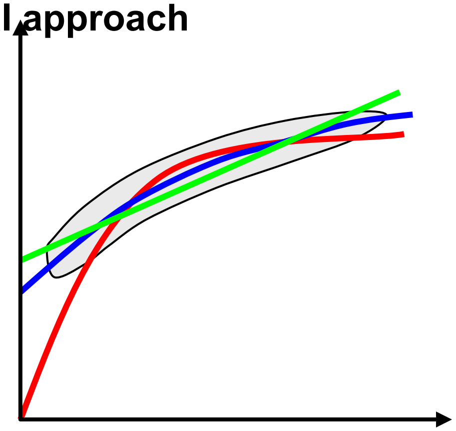
"The Y-Intercept is just what it is"

- **Typical arguments**
  - We are not trying to predict the y-intercept. We are trying to predict the cost of systems of <u>non-zero size</u>.
    - We should <u>take the best advice the data can give us</u>
    - We should extrapolate as little as we can.
  - If the data show that the y-intercept is non-zero, we should not reject a CER just because we do not know why
    - Galileo <u>believed the data</u>, even absent a theory of gravity. It took centuries before Isaac Newton knew why – but Isaac Newton wouldn't even have wondered without Galileo showing that there was an explanation missing
  - This approach is what the practice of statistics currently recommends
    - It should be noted that this argument gets little traction with engineers, who are trained to believe in literal meaning
- **Pros:**
  - Any existing system (i. e., one of the data points underlying the CER) is well-predicted
- **Cons**
  - There is no literal meaning to the y-intercept, which is not very satisfying

Y-Intercepts in CER Development

Omission of the y-Intercept Among Cost Estimators
Linear and Power Forms

- In an earlier paper[2] we noted that when fitting a power curve, estimators often forget that Ordinary Least Squares (OLS) cannot deal with an equation of the form $y = a * x^c + b$
  - To perform OLS, the data must be linear, but the log of this equation does not have a linear form
  - Not understanding this, analysts proceed with a mechanical approach
- Analysts usually just take the log of both sides and conduct OLS, but the fit will be poor, and the regression probably will not be significant, as the red curve portrays
  - The higher the y-intercept, the poorer the fit
  - The less arced (more linear) the data, the poorer the fit
- A linear form, $y = a * x + b$, like the green one, may fit
  - With a higher y-intercept, the linear model fits better than the power curve (red)
  - The more arced the data are, the worse it will fit
- The best fit would be the blue power curve
- There are several ways to fit the blue curve, including Excel Solver
  - There is no test of significance, but there is at least one approach to the fitting of confidence and prediction intervals[3]

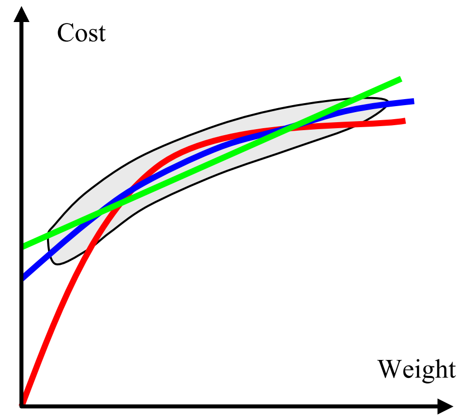


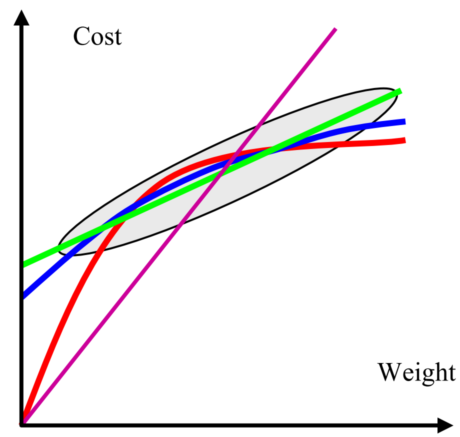
2. *Cost Response Curves - Their Generation, Their Use in IPTs, Analyses of Alternatives, and Budgets*; K. E. Crum, K. L. Allison, R. L. Coleman, R. Klion, 29th ADoDCAS, 1996
3. *Prediction Bounds for General-Error-Regression CERs,* Stephen A. Book, 39th ADoDCAS

NORTHROP GRUMMAN

Omission of the y-Intercept Among Cost Estimators
Linear and Power Forms

- **In an earlier paper[2] we noted that when fitting a power curve, estimators often forget that Ordinary Least Squares (OLS) cannot deal with an equation of the form $y = a * x^c + b$**
  - **To perform OLS, the data must be linear, but the log of this equation is not a linear form**
  - **Not understanding this, analysts proceed with a mechanical approach**
- **Analysts usually just take the log of both sides and conduct OLS, but the fit will be poor, and the regression probably will not be significant, as the red curve portrays**
  - **The higher the y-intercept, the poorer the fit**
  - **The less arced (more linear) the data, the poorer the fit**
- **A linear form, $y = a * x + b$, like the green one, may fit**
  - **With a higher y-intercept, the linear model fits better than the power curve (red)**
  - **The more arced the data are, the worse it will fit**
- **The best fit would be the blue power curve**
- **There are several ways to fit the blue curve, including Excel Solver**
  - **There is no test of significance, but there is at least one approach to the fitting of confidence and prediction intervals[3]**
- **Sometimes a similar result occurs because the analyst insists that the CER "should" go through the origin ("the cost of nothing should be zero"), and "must make sense"**
  - **This insistence can lead to rejection of the green line and the blue curve and choice of the red curve or the purple line - the red curve or purple line, though bad fits, are the best choice that go through the origin**





2. *Cost Response Curves - Their Generation, Their Use in IPTs, Analyses of Alternatives, and Budgets*; K. E. Crum, K. L. Allison, R. L. Coleman, R. Klion, 29th ADoDCAS, 1996
3. *Prediction Bounds for General-Error-Regression CERs*, Stephen A. Book, 39th ADoDCAS

The Y-Intercept and Rates

Also known as Factors

The Data

- Suppose we have the below data*
- What should we plot?
- What do we expect?

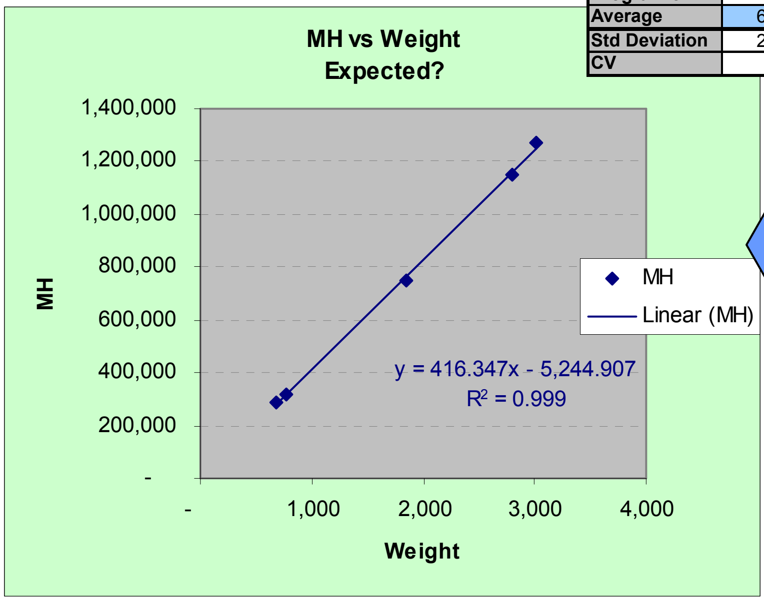
Program	MH	Weight	MH/Lb
Program 1	384,216	686	560
Program 2	368,537	765	482
Program 3	865,810	2,805	309
Program 4	857,160	3,011	285
Program 5	796,145	1,852	430
Average	654,374	1,824	413
Std Deviation	255,252	1,094	116
CV	39%	60%	28%

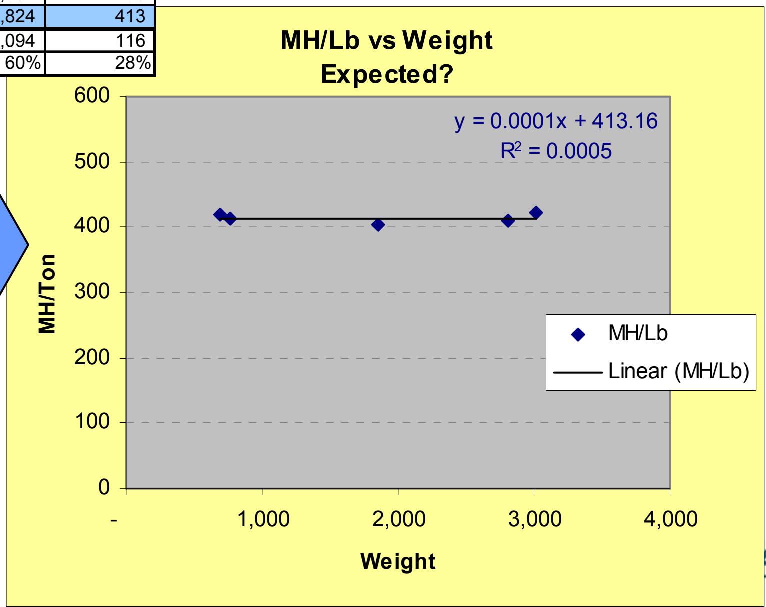
*Actual data, obscured to avoid identification.

What do we expect?

- **We were probably expecting some sort of rate**
  - We expect that MH are related to weight, and MH/Lb might be constant, so there might be a good rate to use for our estimate

Program	MH	Weight	MH/Lb
Program 1	384,216	686	560
Program 2	368,537	765	482
Program 3	865,810	2,805	309
Program 4	857,160	3,011	285
Program 5	796,145	1,852	430
Average	654,374	1,824	413
Std Deviation	255,252	1,094	116
CV	39%	60%	28%



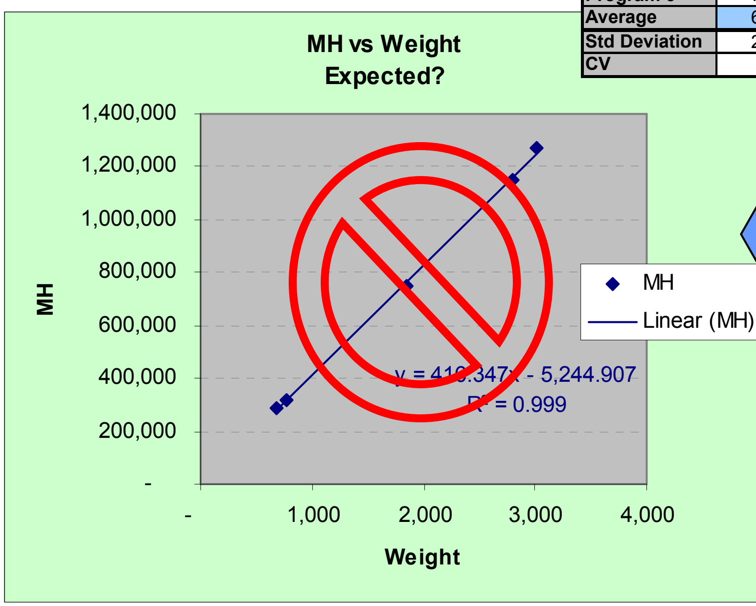


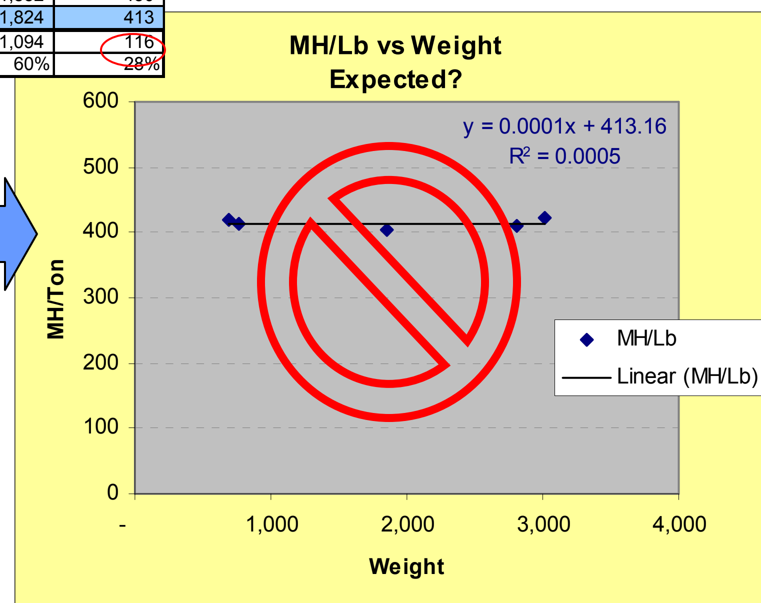
RLC, JRS, PJB, BLC, ERD, SCEA 2007

What do we expect? What do we get?

- **We were probably expecting some sort of rate**
  - We expect that MH are related to weight, and MH/Lb might be constant, so there might be a good rate to use for our estimate
- **We should look at the data before guessing**
  - The CV of the supposed rate is quite high, not much better than MH were to begin with, so we're not done
  - Those aren't really the graphs of the data, it was just what we expected

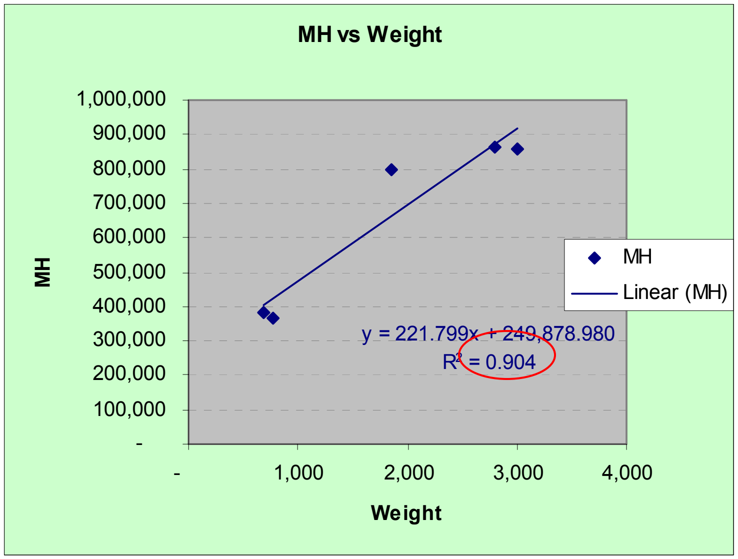
Program	MH	Weight	MH/Lb
Program 1	384,216	686	560
Program 2	368,537	765	482
Program 3	865,810	2,805	309
Program 4	857,160	3,011	285
Program 5	796,145	1,852	430
Average	654,374	1,824	413
Std Deviation	255,252	1,094	116
CV	39%	60%	28%





What do we expect? What do we get?

- **We were probably expecting some sort of rate**
  - We expect that MH are related to weight, and MH/Lb might be constant, so there might be a good rate to use for our estimate
- **We should have looked closer at the data before guessing**
  - We should always look at a scatter plot of the 2 variables in our rate … the real graph
  - The CV is quite high, a key signal (Our rule: CV should be < 10-15% to use a rate)
  - What we have is a failure to graph - we didn't think the rate was a function of weight but, it was
    - It is not a rate at all
    - There is a CER because we overlooked the pesky (large) y-intercept!
  - We now should consider a CER on MH as a function of weight
    - Given the r^2, we can expect a CV for our CER of 0.904 * 39% = 3.7%



Program	MH	Weight	MH/Lb
Program 1	384,216	686	560
Program 2	368,537	765	482
Program 3	865,810	2,805	309
Program 4	857,160	3,011	285
Program 5	796,145	1,852	430
Average	654,374	1,824	413
Std Deviation	255,252	1,094	116
CV	39%	60%	28%

NORTHROP GRUMMAN

The Y-Intercept and Metrics

Also known as Ratios or Thumb Rules

The Data

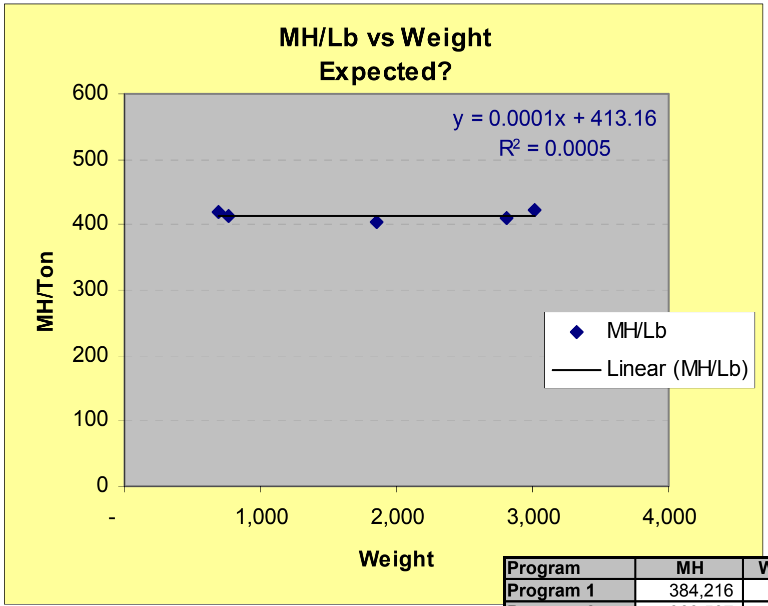
- Suppose we have the below data*
- What should we plot and what should we regress?
- What do we expect?

Program	MH	Weight	MH/Lb	Ln MH/Lb	Ln Wt
Program 1	384,216	686	560	6.328	6.531
Program 2	368,537	765	482	6.178	6.639
Program 3	865,810	2,805	309	5.732	7.939
Program 4	857,160	3,011	285	5.651	8.010
Program 5	796,145	1,852	430	6.064	7.524
Average	654,374	1,824	413	5.991	7.329

*Actual data, obscured to avoid identification.

What do we expect?

- **We were probably expecting some sort of thumb rule to emerge**
  - We expect that MH/Lb might be constant, there might be a good thumb rule we could use

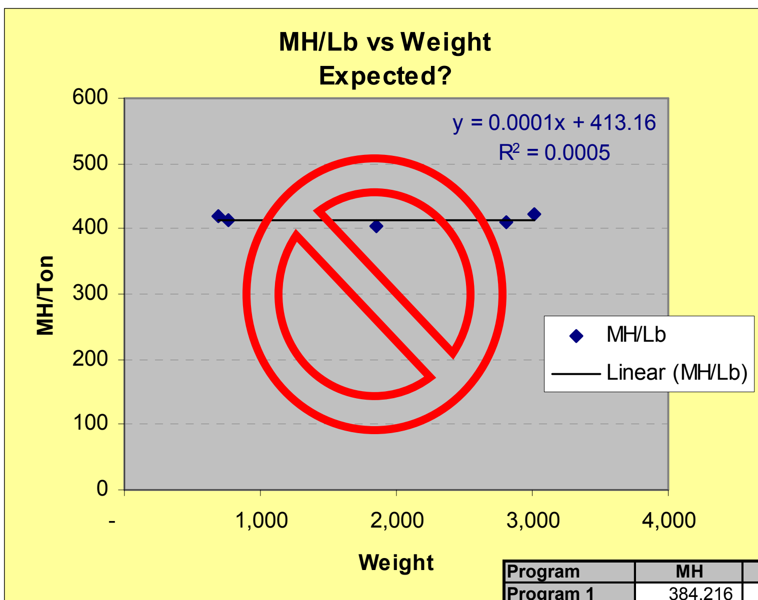


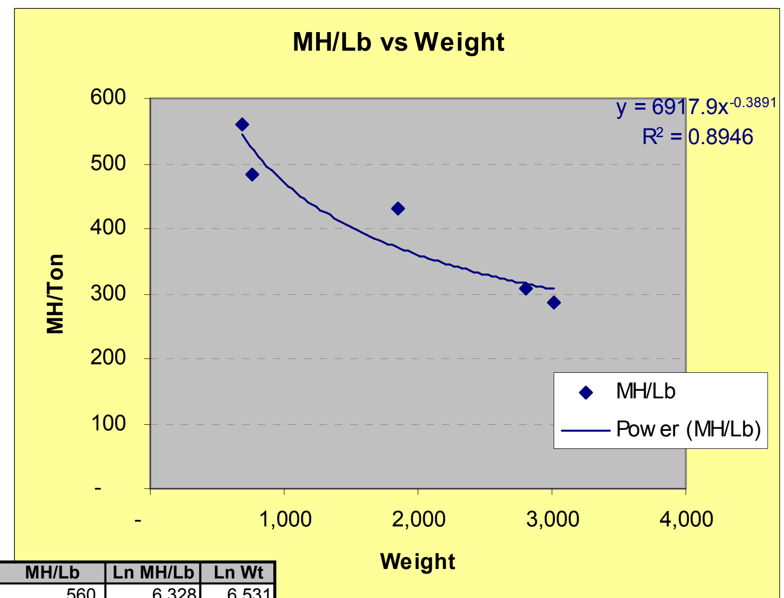
Program	MH	Weight	MH/Lb	Ln MH/Lb	Ln Wt
Program 1	384,216	686	560	6.328	6.531
Program 2	368,537	765	482	6.178	6.639
Program 3	865,810	2,805	309	5.732	7.939
Program 4	857,160	3,011	285	5.651	8.010
Program 5	796,145	1,852	430	6.064	7.524
Average	654,374	1,824	413	5.991	7.329

NORTHROP GRUMMAN

What do we expect? What do we get?

- **We were probably expecting some sort of thumb rule to emerge**
  - We expect that MH/Lb might be constant, there might be a good thumb rule we could use
- **We actually got the graph on the lower right … we got a very interesting, even compelling graphic**
  - MH/Lb clearly decreases, and seems to follow a smooth curve
- **What is happening here?**



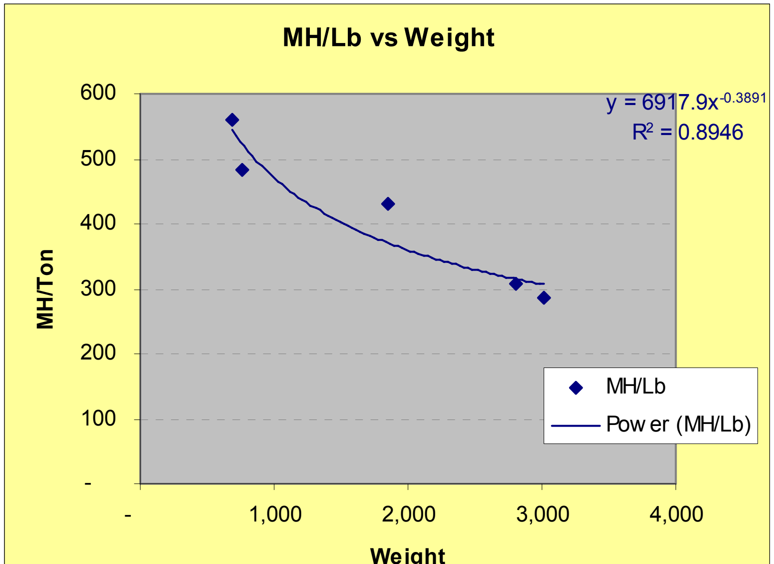


Program	MH	Weight	MH/Lb	Ln MH/Lb	Ln Wt
Program 1	384,216	686	560	6.328	6.531
Program 2	368,537	765	482	6.178	6.639
Program 3	865,810	2,805	309	5.732	7.939
Program 4	857,160	3,011	285	5.651	8.010
Program 5	796,145	1,852	430	6.064	7.524
Average	654,374	1,824	413	5.991	7.329

NORTHROP GRUMMAN

What Does this mean?

- **Hypotheses:**
  - Larger units are less complex, and so the work is less demanding
  - Larger units have less density, thus are easier to work in/on
  - Larger units have thicker structure, which is easier to work with, being less likely to deform, easier to weld, etc.
- **Think for a moment … which of these hypotheses is right?**



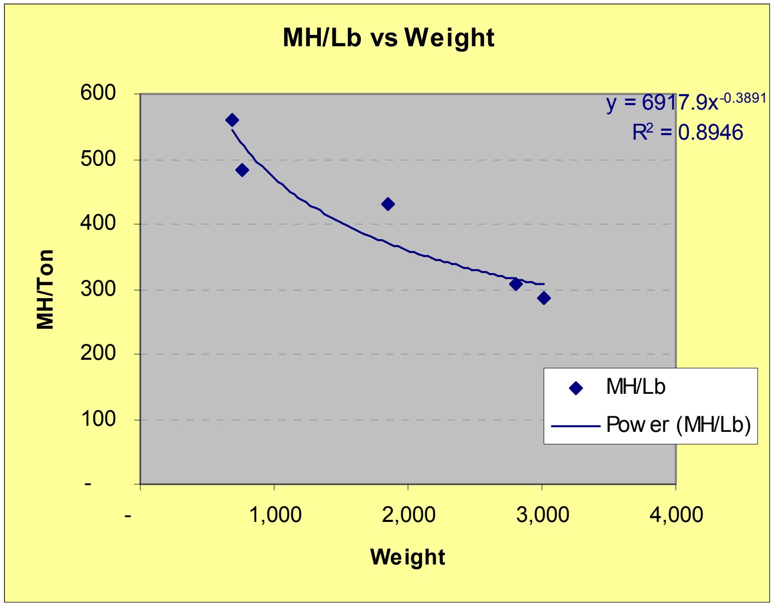
Program	MH	Weight	MH/Lb	Ln MH/Lb	Ln Wt
Program 1	384,216	686	560	6.328	6.531
Program 2	368,537	765	482	6.178	6.639
Program 3	865,810	2,805	309	5.732	7.939
Program 4	857,160	3,011	285	5.651	8.010
Program 5	796,145	1,852	430	6.064	7.524
Average	654,374	1,824	413	5.991	7.329

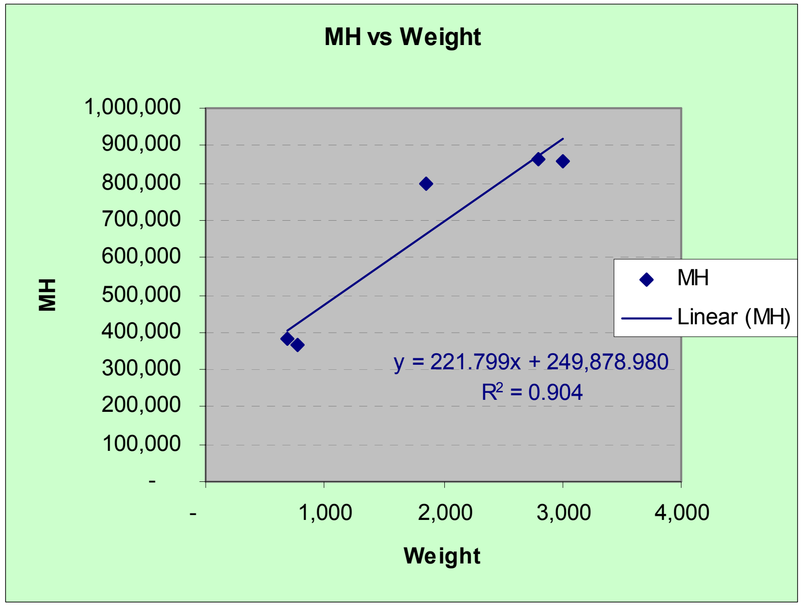
Note: These are actual hypotheses gleaned from actual cost presentations
These hypotheses are coupled with assertions like "we know this commodity, we have experience, you haven't built these"

NORTHROP GRUMMAN

MH and Weight – There are Two Graphs!!

- **We plot MH/Lb vs. Weight and MH vs. Weight and regress**
  - Both regressions seem good, which one do we choose?





Program	MH	Weight	MH/Lb	Ln MH/Lb	Ln Wt
Program 1	384,216	686	560	6.328	6.531
Program 2	368,537	765	482	6.178	6.639
Program 3	865,810	2,805	309	5.732	7.939
Program 4	857,160	3,011	285	5.651	8.010
Program 5	796,145	1,852	430	6.064	7.524
Average	654,374	1,824	413	5.991	7.329

MH and Weight – Two Choices

- **But, first, what does it mean to say MH/Lb is a power function of Weight (or Complexity or Density?)**
- **Let us see:**

Suppose $y/x = a * x^{c}$

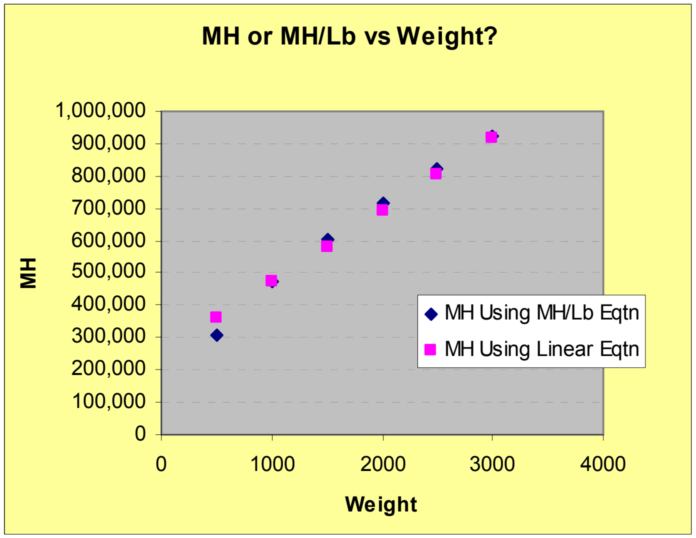
Then $y = a * x^{c+1}$

In our equation, y is MH ,x is weight

So, if $MH/Lb = 6918\ Wt^{-.3891}$

Then $MH = MH/Lb * Lb = 6918 * Wt^{-.3891} * Wt = 6918\ Wt^{.6109}$

- **So, let's plot the linear equation from the last slide and this new equation and see why they both work:**



- **They are almost identical, why choose the more complex power equation?**

MH and Weight – Two Choices

- **What does each equation say?**
- **Linear: MH = a * Wt + b**
  - As weight rises, cost rises, proportionately
  - Like most CERs, this one does not pass through the origin
    - As we noted, many analysts say it should, "If there's no weight, there's no cost"
  - But statisticians face this problem all the time. They are prone to dismiss any attempts to attach meaning to the y-intercept and accept it
    - Since a zero value of the explanatory variable is not part of the data, we cannot say much about the equation there … requiring any equation to do well at the zero-x value is too restrictive
  - In other words, the acceptability of <u>this</u> equation depends on our <u>beliefs</u>
- **Non-linear: MH/Lb = a * Wt ^{c}**
  - The usual explanation is that some sort of variable like density or complexity is driving a shift
    - This equation doesn't pass through the origin either
  - But, this equation has a hidden trap … if density is a driver, we should plot it, not just assert a smooth "density" variable that moves with size and just wave it away … but density is arcane and un-plottable, so we just wave our hands
  - In other words, the acceptability of <u>this</u> equation depends on our <u>beliefs</u>
- **The real point is that the odd form of the non-linear equation comes about because of the y-intercept … if there were no y-intercept, the graph would be flat**
- **We submit that the non-linear equation is merely an artifact, that the search for meaning is a wild goose chase**
  - But, of course, this comports with our beliefs

NORTHROP GRUMMAN

The Y-Intercept and Analogies[4]

4. *Analogies: Techniques for Adjusting Them*, R. L. Coleman, J. R. Summerville, Northrop Grumman; S. S. Gupta, IC CAIG, ASC/Industry Cost & Schedule Workshop Spring 2004, SCEA 2004 72nd MORSS - 2004

Discussion

- **Considerable attention is devoted to techniques in the development of Cost Estimating Relationships (CERs) for parametric estimating**
  - Research on CERs
  - Methods for calibrating
- **Considerable expertise is to be found in buildup techniques**
  - Many Original Equipment Manufacturers (OEMs) have large cost shops which practice buildup
- **Analogy, on the other hand, has been given little attention**

NORTHROP GRUMMAN

The Current Method

Typical Adjustments – By Ratio

- **Adjustments, in the analogy or buildup method, typically rely on an "obvious" characteristic**
  - The characteristic is most often weight
  - Sometimes weight of the new system is not known, and so another characteristic is used (often as a proxy for weight)
  - Sometimes a characteristic such as bore diameter of a gun is used
- **Usually the ratio of the values of the characteristic in the new system to the value in the old system is multiplied by the cost of the old system**
  - Sometimes called "j-ing up the estimate*"
- **Sometimes the characteristic is transformed in a way that is thought to make it proportional to weight**
  - E.g., the bore diameter of a gun, is cubed
  - In these cases, there may be a presumed relationship to weight

* Apparently from "ad**j**usting"

Implications of the Current Method

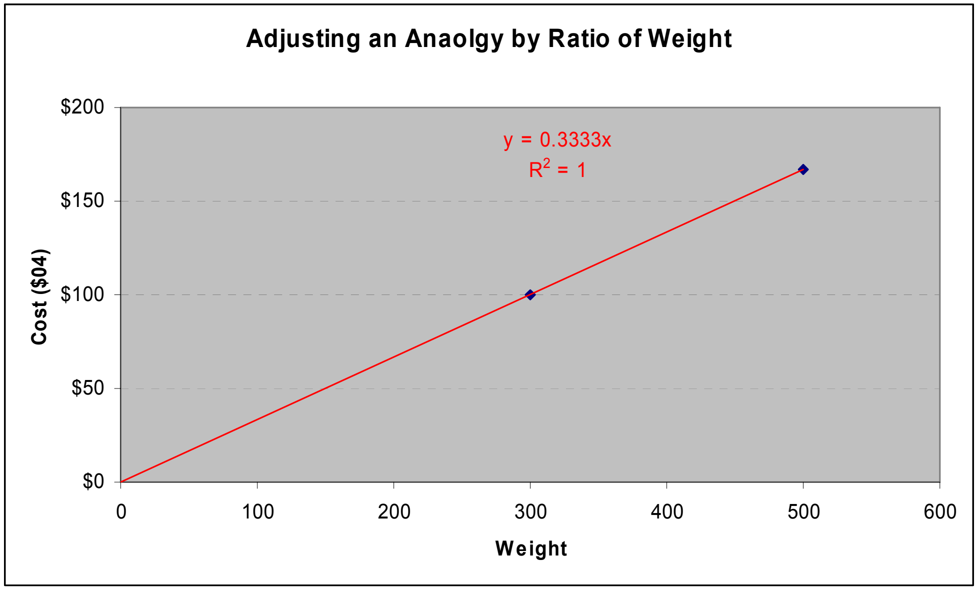
- **An example adjustment by ratio is:**
  - The analogy weighs 300 tons and costs $100M
  - The new system weighs 500 tons and so is assumed to cost (500/300) * $100M = $166.67M
- **This is a typical and familiar adjustment**
  - What is its implication?
  - Should we be inclined to believe it?
  - Is it in accord with what we believe?

… let's look at a graph to see what it implies … there is a surprise there for most of us … but first, force yourself to <u>predict what the line between the analogy and the prediction looks like</u> … where does it cross the y axis?

Adjustment by Ratio – The Graph

- **The below graph shows the previous adjustment**
  - The analogy weighs 300 tons and costs $100M
  - The new system weighs 500 tons and is assumed to cost (500/300) * $100M = $166.67M
  - Note that the line through the 2 points passes through the origin

Important Observation:
Straight adjustments by ratios *always* pass through the origin! Most observers fail to predict that, even though it is straightforward to show that it must.
Important Question:
Is this reasonable?



Proposal - Two New Methods

- **Borrowed slope[5] – a variant of the methods for calibrating CERs**
  - Adjust a "trusted analogy" by a "trusted slope"
- **Relational Correlation[6] – taking advantage of the geometry of the bivariate normal and regression**
  - Adjust a "trusted analogy" by a "best guess slope"
- **We will discuss the first method, but in the interests of time, we refer the reader to the original paper (or to the backup slides) for the more arcane second method**

5. *A Framework for Costing in a CAIV Environment*, R. L. Coleman, TASC; D. Mannarelli, Navy ARO, ASNE 1996, ADoDCAS 1996

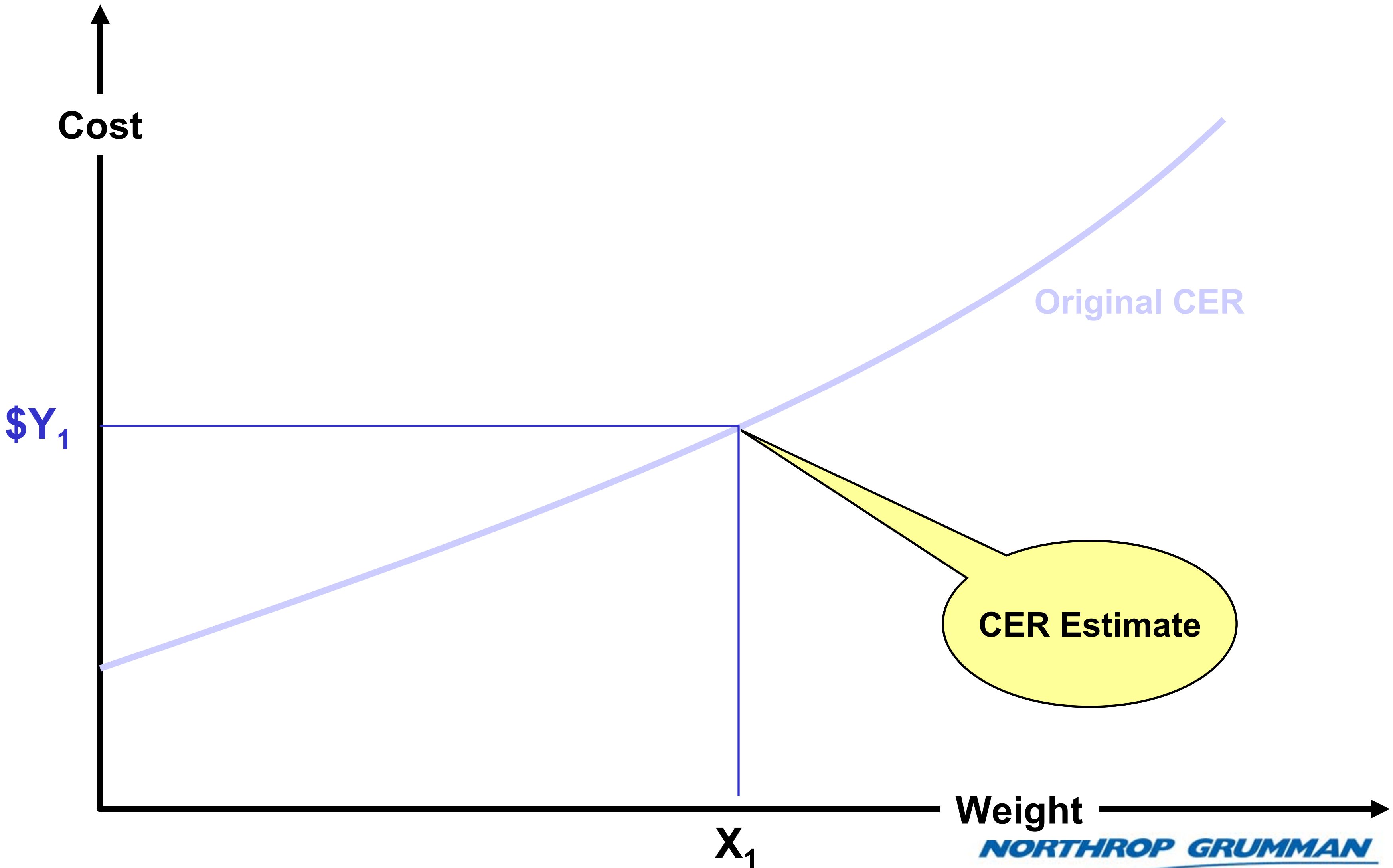
6. *Relational Correlation, What to do when Functional Correlation is Impossible,* ISPA/SCEA 2001, R.L. Coleman, J.R. Summerville, M.E. Dameron, C.L. Pullen; TASC, Inc., S.S.Gupta, IC CAIG

The Borrowed Slope Method

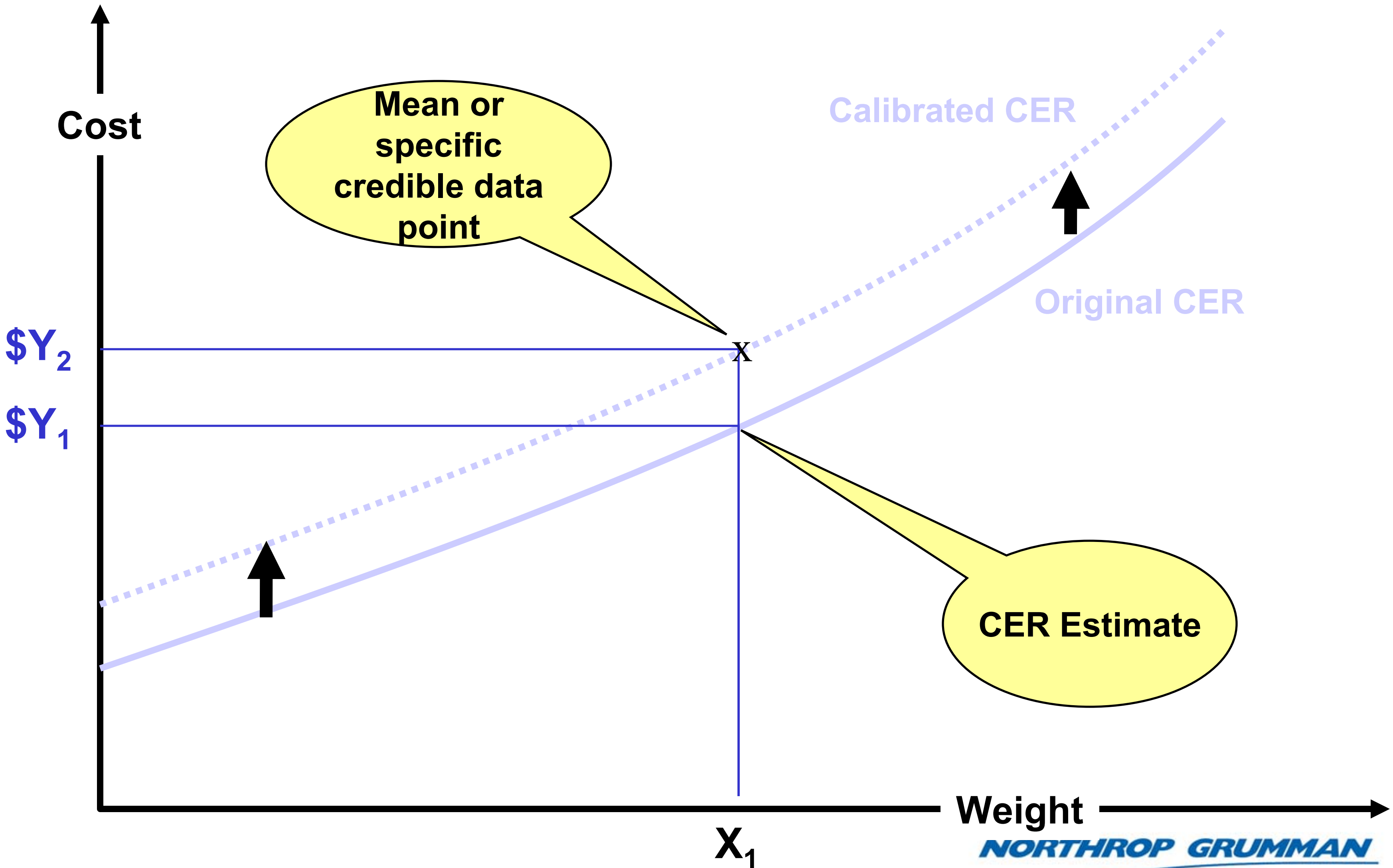
- **Based on "calibrating a CER"**
  - A CER is adjusted to "more-trusted," or industry, or company-specific data by moving the slope to pass through a point or set of points
  - Picture follows
- **To adjust an analogy, do precisely the same thing**
  - Instead of believing you are adjusting a CER to specific data, think of it as departing from "the most credible point" via "the most credible slope"

NORTHROP GRUMMAN

Calibrating CERs

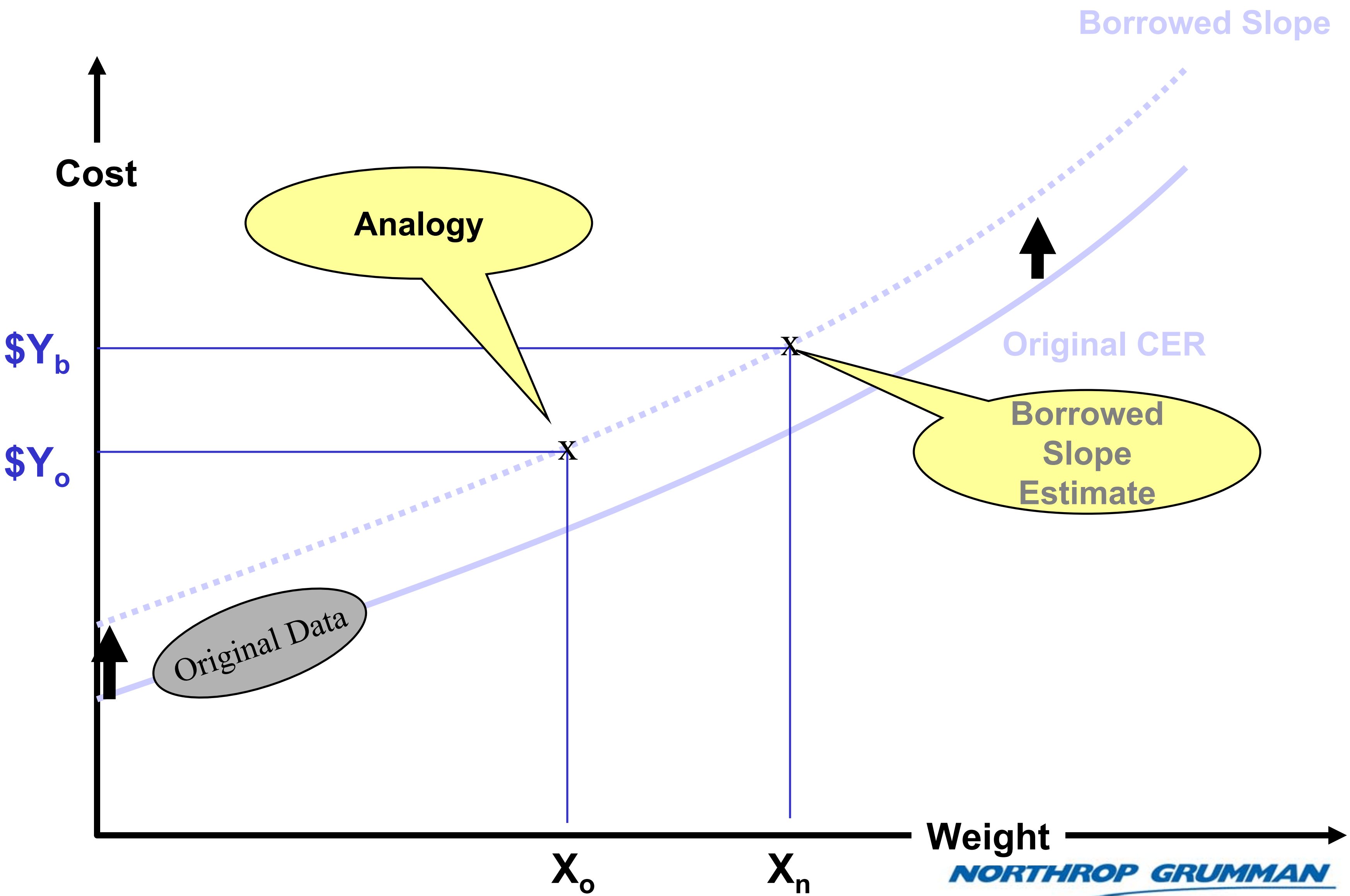


Calibrating CERs

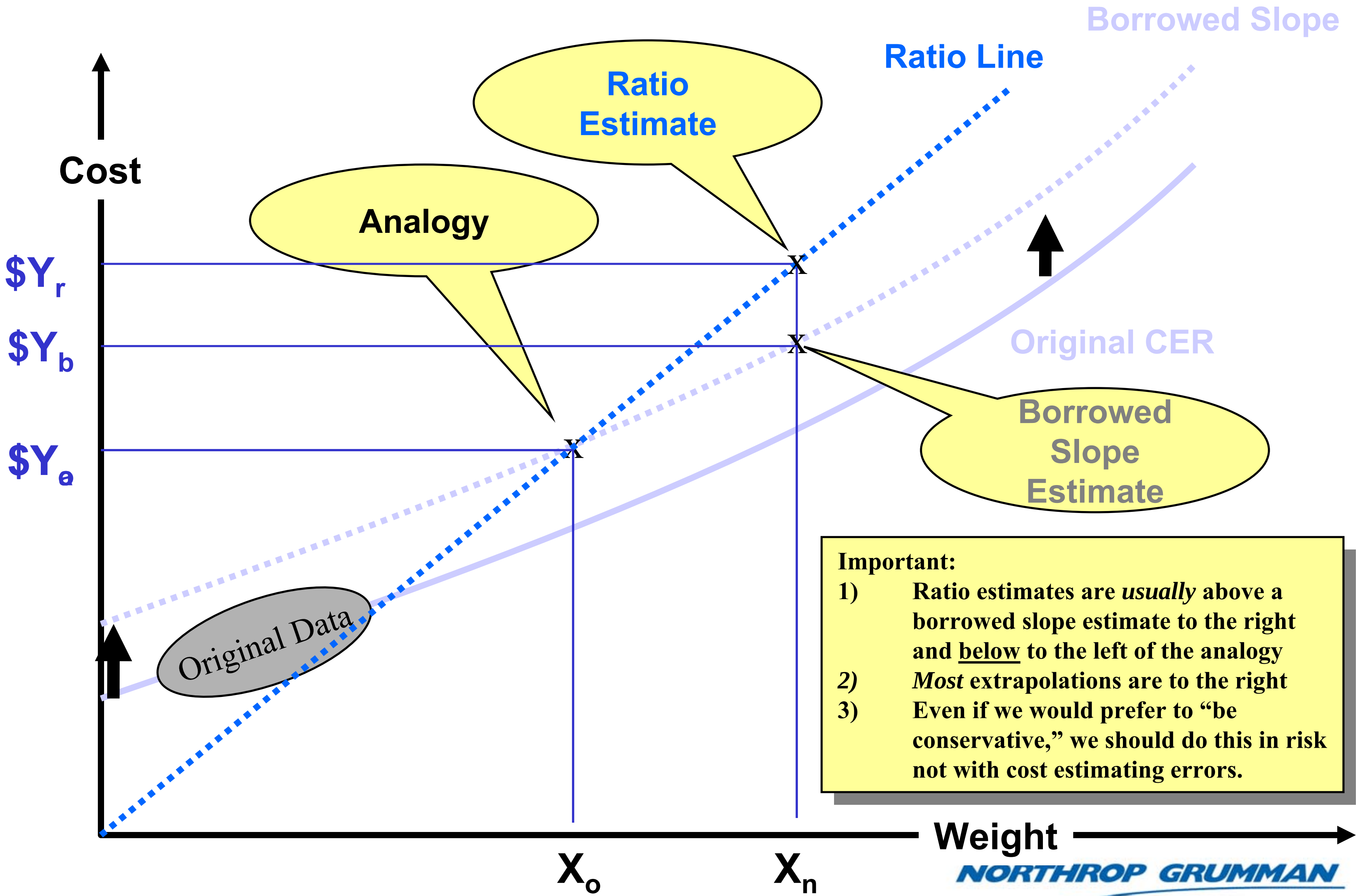


NORTHROP GRUMMAN

The Borrowed Slope Estimate



Comparison – Borrowed Slope & Ratio Estimates



Important:

1) Ratio estimates are *usually* above a borrowed slope estimate to the right and <u>below</u> to the left of the analogy
2) *Most* extrapolations are to the right
3) Even if we would prefer to “be conservative,” we should do this in risk not with cost estimating errors.

NORTHROP GRUMMAN

Conclusions

- **We discussed the 3 schools of thought with regard to y-intercepts**
- **We discussed CERs, rates, metrics (thumb rules), and analogies with regard to the y-intercept**
- **We urge you to think about this**
  - We are less offended at your holding beliefs (after all, we do!) than by sailing by this issue all unawares
- **If we had our way, we'd hope you believed in y-intercepts*, but in any event, think about it!**
  - Clap* if you believe in y-intercepts!

*In the second act of Walt Disney's "Peter Pan", Tinkerbell drinks poison that Peter is about to drink in order to save him. Peter turns to the audience and says, *"Tinkerbell is going to die because not enough people believe in fairies. But if all of you clap your hands real hard to show that you do believe in fairies, maybe she won't die."*
We all started to clap. I clapped so long and so hard that my palms hurt. Then suddenly the actress playing Peter Pan turned to the audience and she said, *"That wasn't enough. You did not clap hard enough. Tinkerbell is dead."* And then we all started to cry. The actress stomped off stage and refused to continue the production. They had to lower the curtain. The ushers had to come help us out of the aisles and into the street.
You hear that? CLAP LOUDER!

Definitions

- **Parametric Estimates: Estimates made by developing statistical "Cost Estimating Relationships" (CERs) based on one or more parameter and cost**
    - Estimates involving parameters but not based on statistical analysis are more properly called either "adjusted analogies" or "adjusted buildups"
- **Analogies: Estimation by assuming that the costs of a new system will be equal to (or similar to) the costs of a system that is similar in form**
    - "Adjustments" are almost always made
- **Buildups: Costed-out physical Bill of Materials (BOMs) and CAD-generated material lists and the like**
    - We do not mean "buildups" consisting entirely of Staffing levels * Duration. Such estimating techniques are little more than "engineering judgment" in fine detail
    - Buildups often include "adjustments" to allow for size differences
- **Composite methods: A method that involves at least two of the three other types**
- **Adjustments: Scaling of a cost by some physical, performance, or other such attribute**
    - Scaling is usually (in practice) directly proportional to the attribute
    - Scaling parameters are usually countable or measurable and intuitively tied to cost

Backup

Relational Correlation

- **A *much more esoteric* method is available, which borrows from**
  - Bivariate normality
  - The geometry of regression
- **This method is available when there is no “trusted slope” to borrow**

Warning: Serious math follows.

NORTHROP GRUMMAN

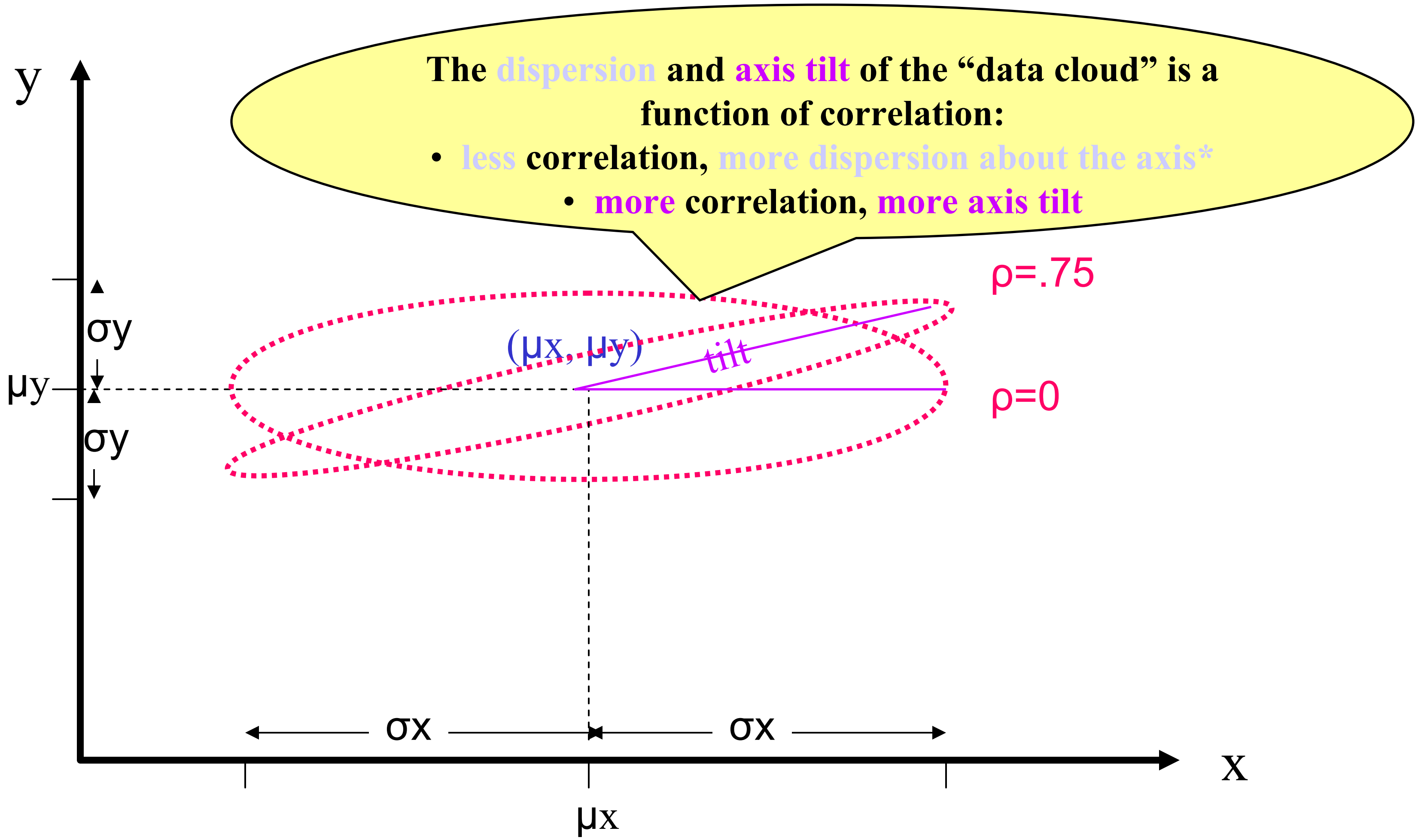
Bivariate Normality

Bivariate Normality

- **Suppose X and Y are distributed N(μ_x, σ_x) and N(μ_y, σ_y)**
- **Suppose X and Y are jointly bivariate normal with correlation ρ**

…. Then the graph of X and Y will appear as follows

The Bivariate Normal



The Geometry of Regression

Regression

- **The below facts are known to mathematicians, but obscure, and not remembered in cost analysis …**
  - For any two jointly distributed variables, there is a regression line
  - The slope is:

$$m = \rho*(\sigma y/\sigma x)$$

  - The y intercept is:

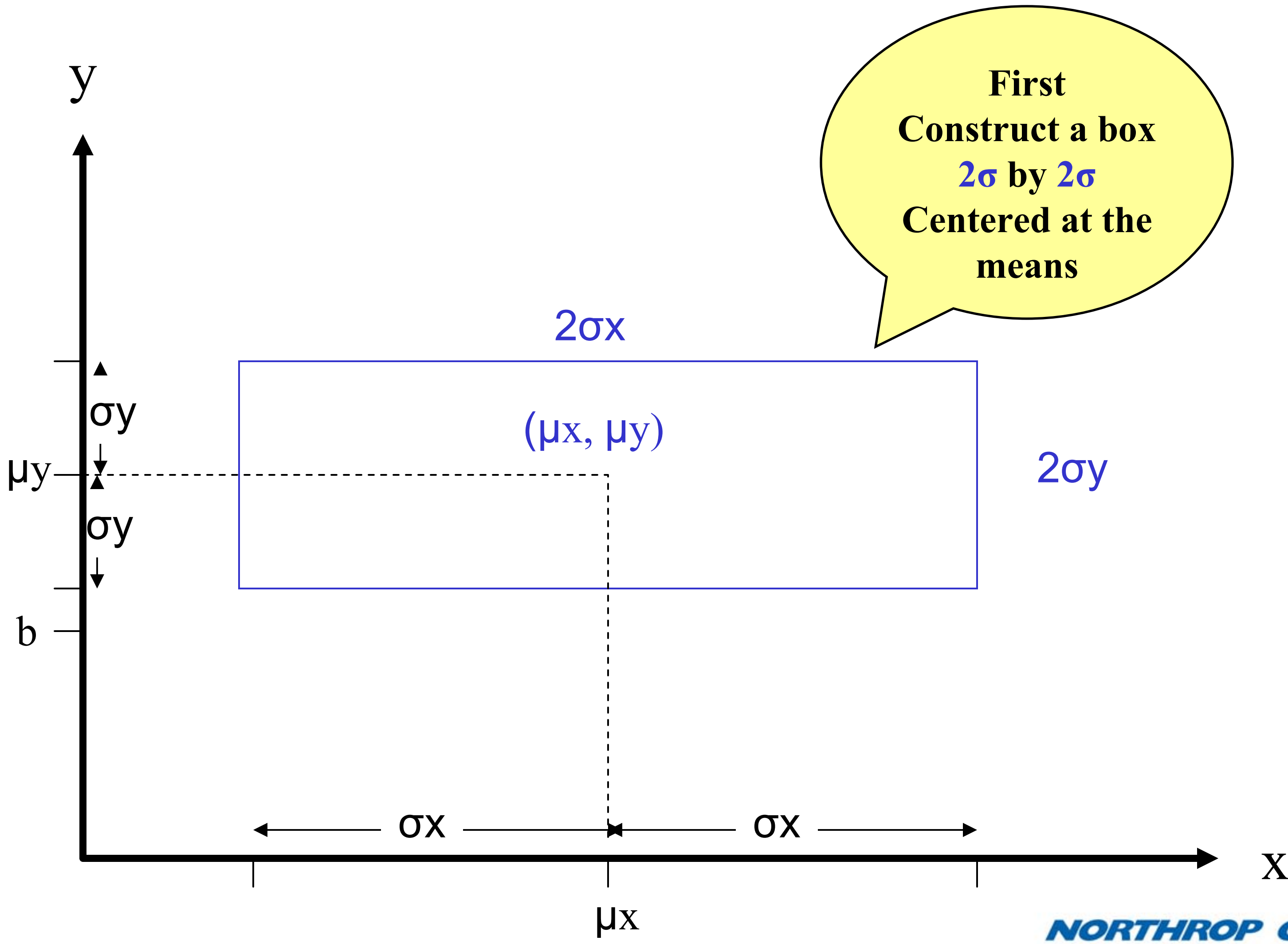
$$b = \mu y - \rho(\sigma y / \sigma x) * \mu x$$

  - If the variables are joint bivariate normal, then ρ is the correlation coefficient

Let’s look at the graph…

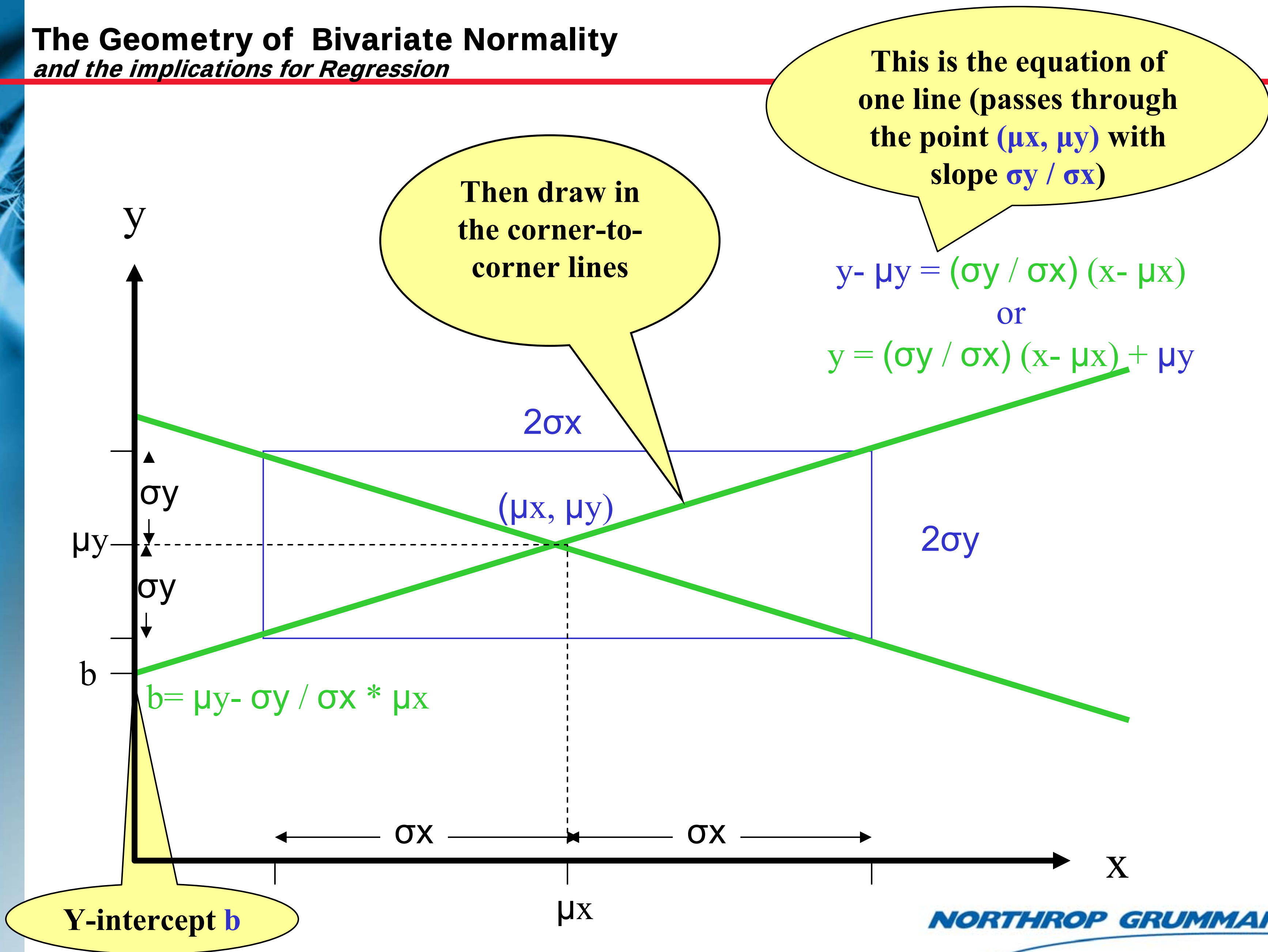
The Geometry of Bivariate Normality

and the implications for Regression



The Geometry of Bivariate Normality

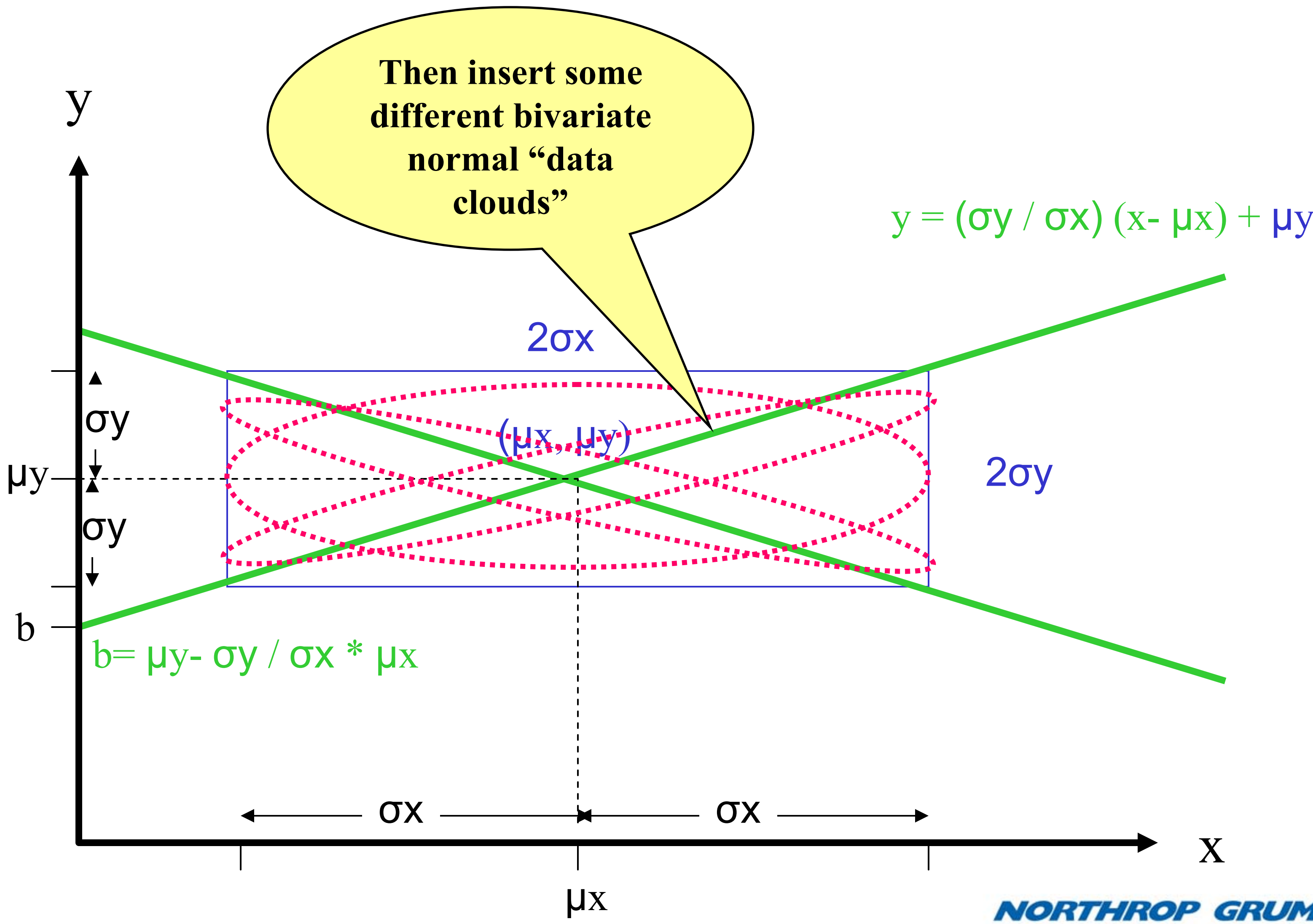
and the implications for Regression



richard.coleman@ngc.com, 11 May 2007, 46

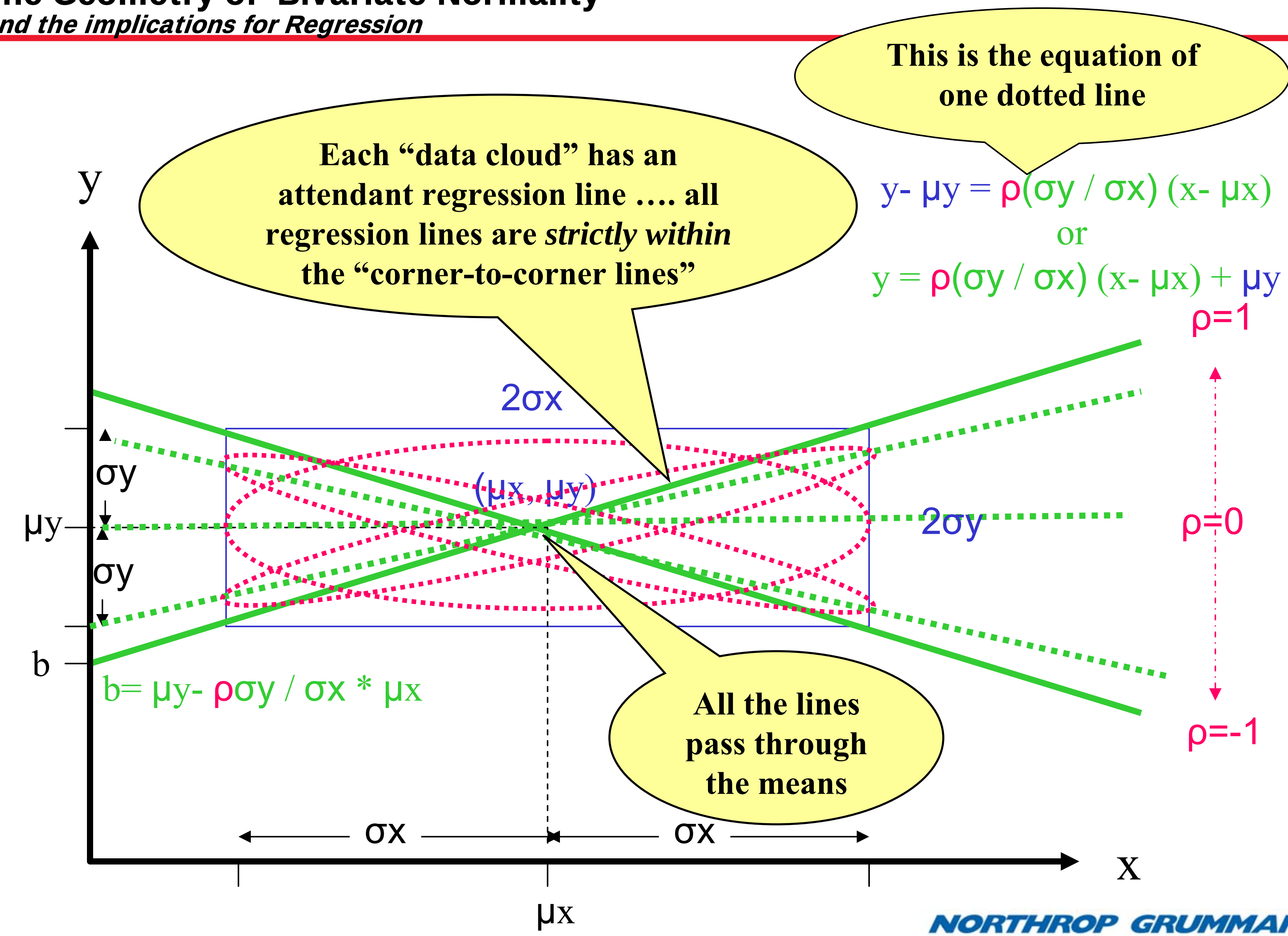
NORTHROP GRUMMAN

The Geometry of Bivariate Normality
and the implications for Regression



The Geometry of Bivariate Normality

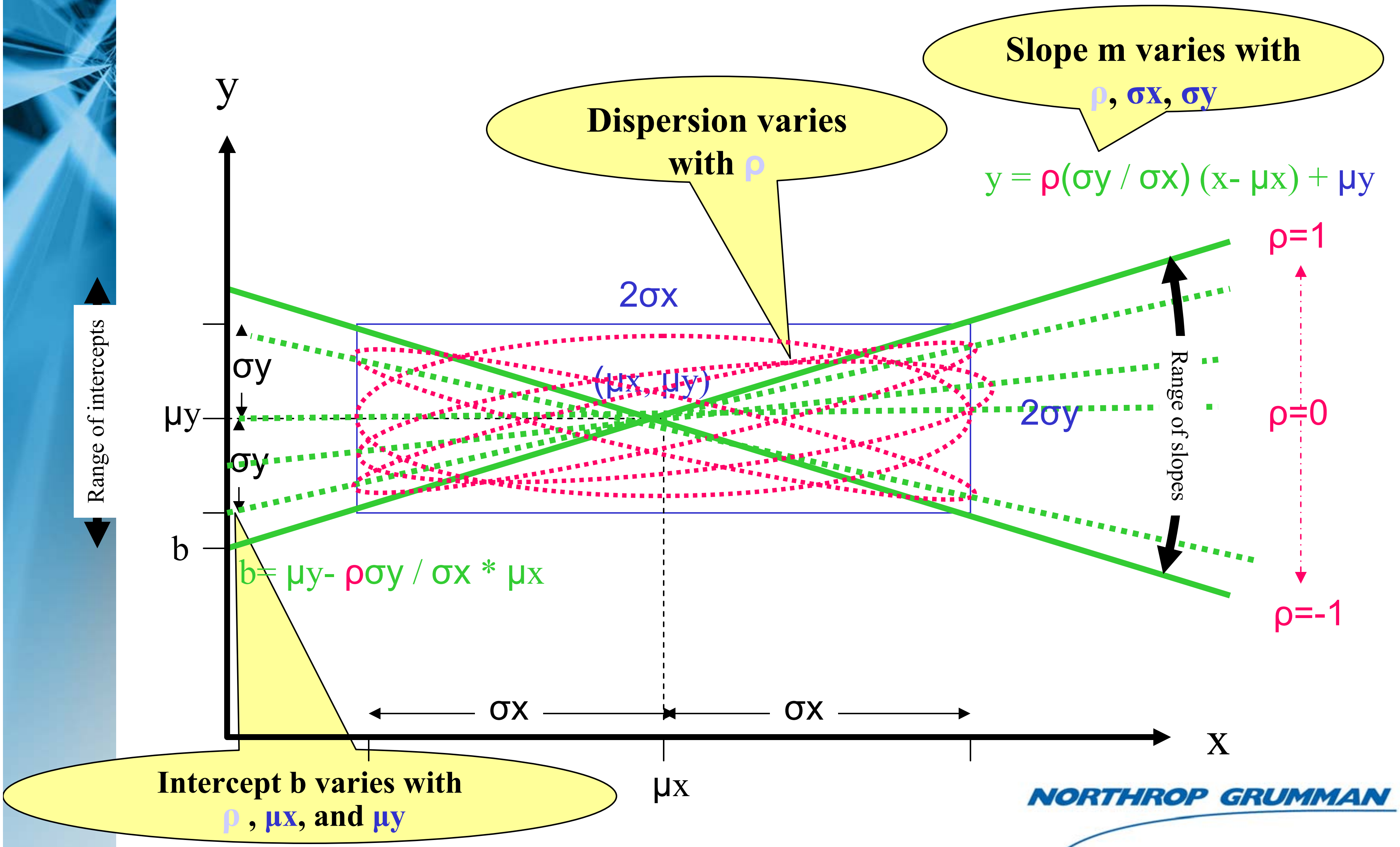
and the implications for Regression



RLC, JRS, PJB, BLC, ERD, SCEA 2007

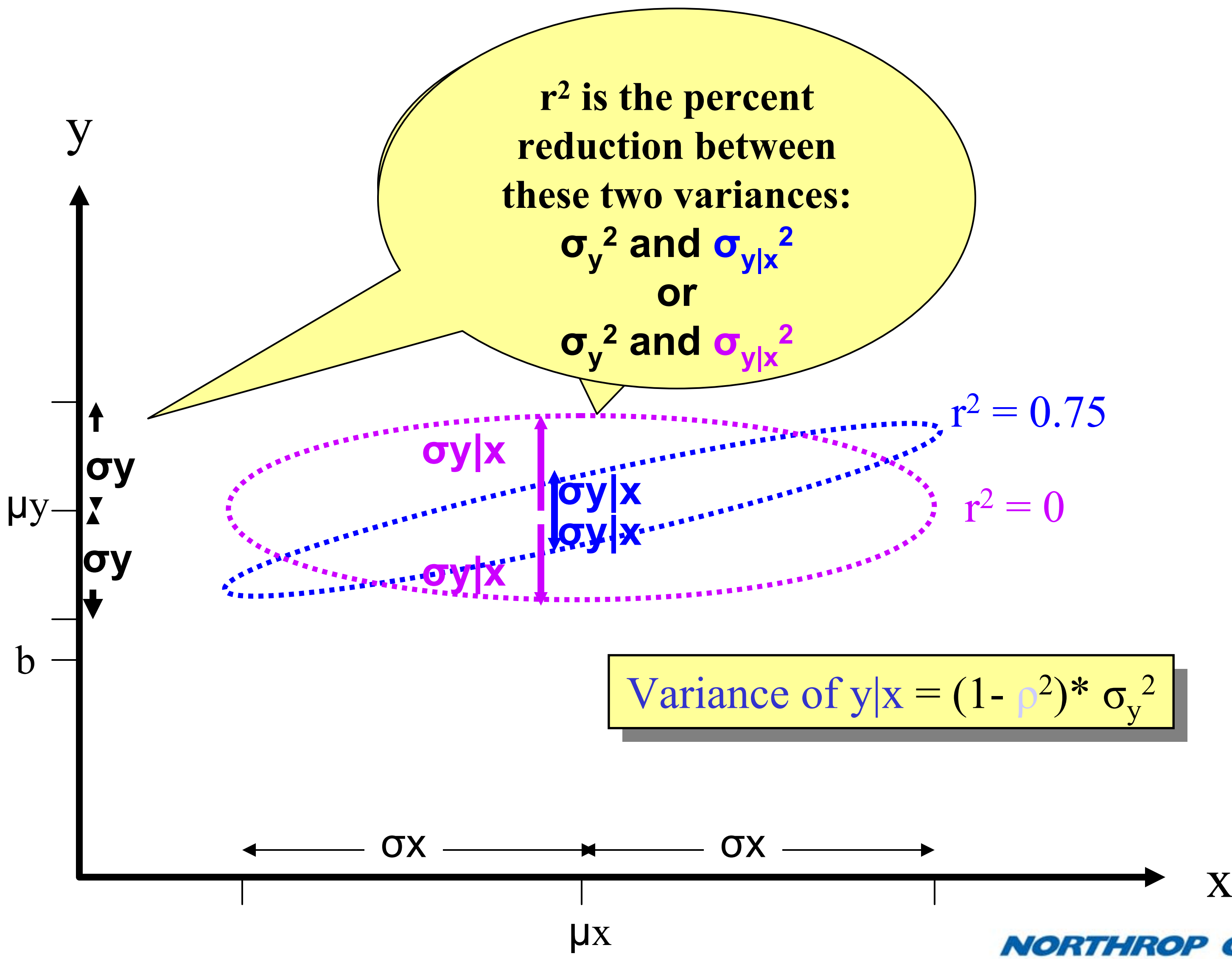
The Geometry of Bivariate Normality

and the implications for Regression



The Geometry of Bivariate Normality

and the implications for Regression

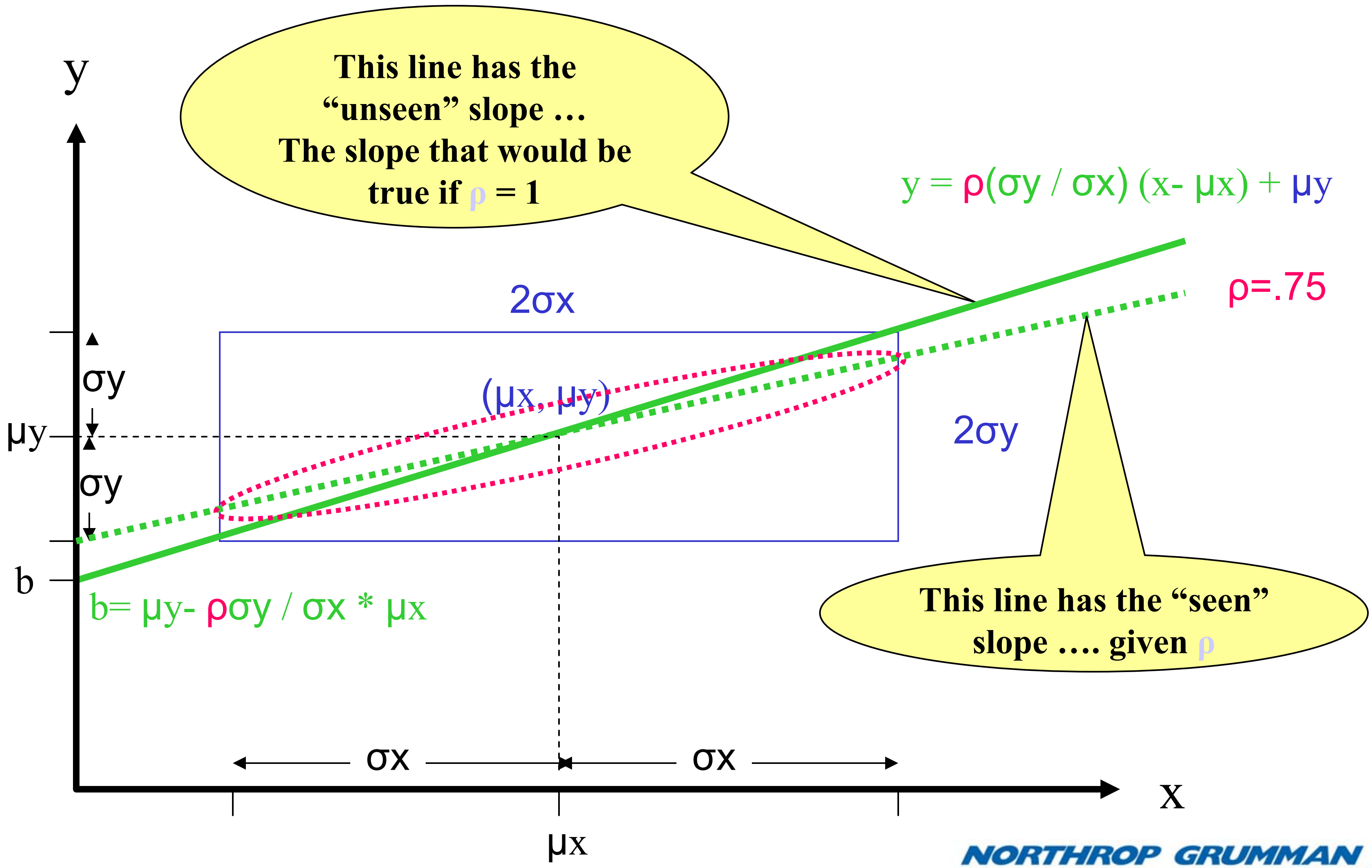


The implications

- **For every regression with apparent slope m, there is an unseen equation**
  - With steeper slope m/ρ which is the *unseen* slope of the two variables
  - With an unseen accompanying y intercept
- **Once we decide upon the means and the variances of x and y, the unseen line is fixed**
  - Once we pick ρ, the regression line is fixed

The Geometry of Bivariate Normality

and the implications for Regression

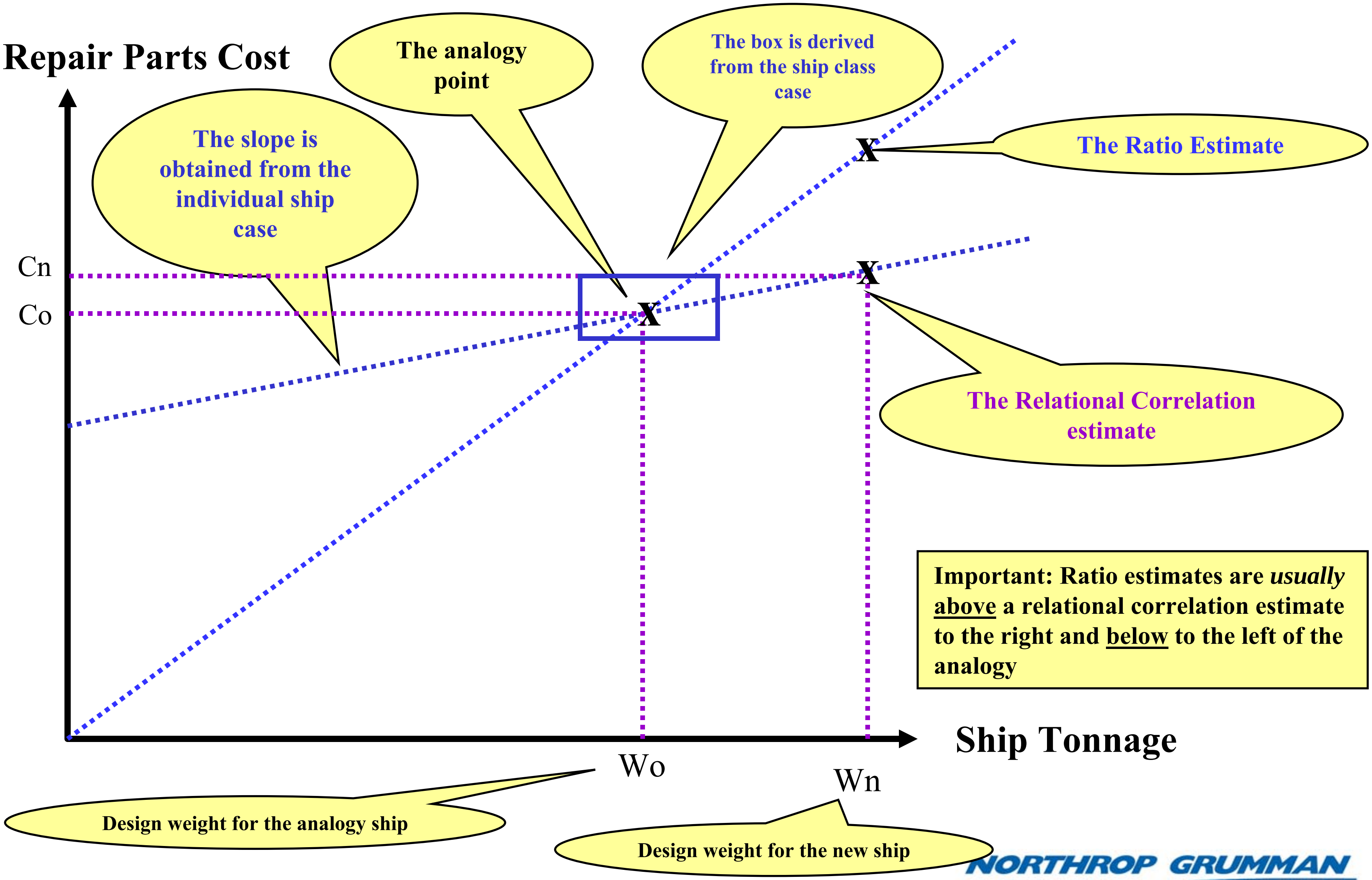


Implementing Relational Correlation for Analogies

Implementing Relational Correlation for Analogies

- **For Single Point Analogies**
  - Determine a reasonable (preferably historically-based) standard deviation for the x and y variable
  - E.g, to estimate ship repair parts as a function of tonnage you'll need:
    1. The standard deviation for the analogy ship class repair parts cost
    2. The standard deviation for the tonnage within the ship class
    3. The standard deviation of repair parts for a single ship of the class
    - The ratio of 1 and 2 gives you the unseen slope
    - The relationship of 3 and 1 will yield r^2 (Variance of y|x = $(1-\rho^2) * \sigma y^2$)
- **For buildups, do as above, but use an analogy for the values of the standard deviations, and apply it to your buildup using percents**

The Relational Correlation Method



Important: Ratio estimates are *usually* <u>above</u> a relational correlation estimate to the right and <u>below</u> to the left of the analogy

Backup & Old

How to Implement Relational Correlation for Expert-Based CERs
The problem

- **You have two WBS elements**
  - Warhead cost
  - Motor cost
- **You know their historic means and standard deviations – for both cost and the driving parameter, say weight**
  - You know these values from independent data bases
    - So, you cannot get correlation
- **You *do* have a CER to predict warhead cost**
- **You *do not* have a CER to predict motor cost**
  - You believe weight is a driver, but a CER cannot be derived
  - And, the data you have is too far away from your program, it needs to be adjusted … but how?
    - You do not wish to simply factor the cost by the weight change
- **This is a typical problem, and is closely related to the risk problem just described**
- **We will try to predict motor cost as a function of warhead cost … a useful equation as well as a helpful CER**

How to Implement Relational Correlation for Expert-Based CERs

1. Ask the engineer: *How much leeway in % do you typically have for weight (or cost) of the motor* *<u>if design has not yet begun</u>?* (The unconstrained case)
   - *Note: this may differ from the historic variation, but we will use it only in a relative sense*
   - *We will translate the weight fluctuation into cost fluctuation*
2. Ask the engineer: *How much leeway in % do you have for weight (or cost) of the motor,* *<u>if design of the warhead is complete</u>?* (The constrained case)
3. This will give you r^2:
   - You already knew the "unseen slope", σy/σx, now you know the "seen" slope ρ(σy / σx), and you know b = μy- ρσy/σx * μx
   - The percent reduction in the variance of y is the r^2, and the square root of that is r (Variance of y|x = (1- ρ^2)* σy^2)
4. Implement the result as a CER, by passing the slope through the analogy or average of your data.

Note: We do not advocate using such a CER in lieu of a standard CER, only if there is no other recourse

How to Implement Relational Correlation for Expert-Based CERs