Midpoint Formulas

Jason T. Lee
Technomics, Inc
June 2007

Table of Contents

Table of Equations

Midpoint Formulas

Jason T. Lee
Technomics, Inc
June 2007

The general learning curve, or cost progress curve, is commonly given by the formula

$$c_q = T_1 q^b . \quad (1)$$

Equation 1 defines both the *unit* cost curve and the *cumulative average* cost curve, depending on how one defines the independent and dependent variables. Under unit learning theory, c_q is the cost of unit q; under cumulative average learning theory, c_q is the average cost of the first q units. In both cases, T_1 is the cost of the first production unit and b is the algebraic slope of the equation. The slope b is related to the percent decrease in unit cost when quantity doubles by the equation

$$b = ln(s) / ln(2), \quad (2)$$

where s represents the percent decrease in unit cost. Thus the algebraic slope b can be understood more easily as a percent: s = (Cost of 2q) / (Cost of q). In the context of "cost progress," $0 < s < 1$. In practice, $0.5 < s < 1$, although mathematically speaking, this is really only necessary when using certain midpoint estimating equations. Unless otherwise noted, throughout the following discussion the slope will be restricted to $0 < s < 1$.

True Algebraic Midpoint

Given Equation 1, the total cost of any particular production lot, say lot L, is given by

$$C = \sum_{q=A}^{B} T_1 q^b , \quad (3)$$

where C is the total cost of the lot, A is the first unit of lot L and B is the last unit of L. Instead of explicitly calculating the summation in Equation 3, we can calculate the total lot cost by multiplying the average unit cost of the lot by the total number of units in the lot. Since the number of units in lot L is $(B - A + 1)$, the *average unit cost* of L, or $\overline{C}$, is given by

$$\overline{C} = \frac{\sum_{q=A}^{B} T_1 q^b}{(B - A + 1)}. \quad (4)$$

The average cost of the lot can also be calculated using the learning curve equation and the *algebraic midpoint*. The algebraic midpoint is the unit at which the average cost of the lot occurs, and is represented by k in Equation 5:

$$\overline{C} = T_1 k^b. \quad (5)$$

The midpoint unit k is a function of the parameter b, because b dictates the severity of curvature in the cost progress curve—that is, how steep or shallow the learning curve is. To solve for k, set Equations 4 and 5 equal:

$$T_1 k^b = \frac{\sum_{q=A}^{B} T_1 q^b}{(B - A + 1)}$$

$$k = \left(\frac{\sum_{q=A}^{B} q^b}{(B - A + 1)} \right)^{\frac{1}{b}}. \quad (6)$$

Equation 6 yields the *true algebraic lot midpoint*.

Approximation with Known Slope: Asher's Approximation (H. Asher, RAND Corp.)

Obviously, Equation 6 is cumbersome because it requires the calculation of the summation of q^b. A shortcut to deriving k without the summation operation is attributed to Harold Asher of RAND[1]. By employing a bit of calculus we can derive the sum of q^b over A through B rather quickly.

As a brief review of the concept, if a curve is drawn by a function $f(q)$, the area under the curve over the interval $[A, B]$ can be estimated by summing the areas of n rectangles under the curve. Each rectangle's base is a particular partition of $[A, B]$, say $[q_{p-1}, q_p]$, and each rectangle's height is given by the point $(c_p, f(c_p))$, if c_p is a point within the subinterval $[q_{p-1}, q_p]$. If the width of each subinterval is given by $\Delta q_p = q_p - q_{p-1}$, then each rectangle's area is $f(c_p) \cdot \Delta q_p$ (see Figure 1).

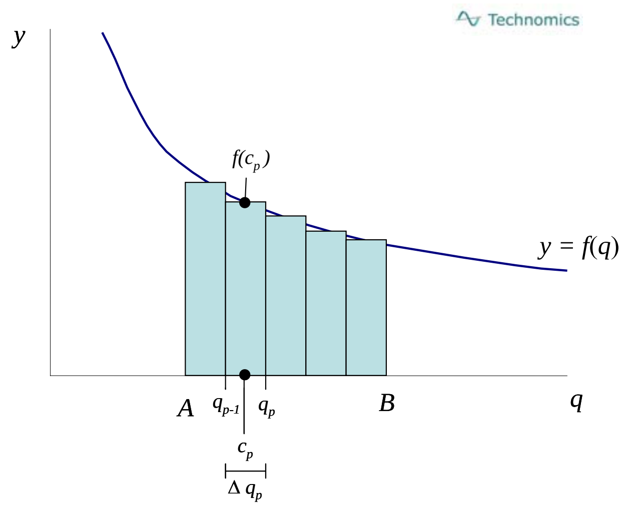


Figure 1. Estimation of the area under a curve on [*A*, *B*].

The sum of the n products of $f(c_p) \cdot \Delta q_p$ is called a *Riemann sum*. As the partitions of the interval $[A, B]$ become smaller, the number of rectangles, n, becomes larger, and the area under the curve approximated by the Riemann sum becomes more accurate. As the limit of the largest partition in the interval $[A, B]$ (called the *norm* of partition P, or $\|P\|$) approaches zero, the base of the largest rectangle approaches zero, and we may say that

$$\lim_{\|P\| \to 0} \sum_{p=1}^{n} f(c_p)\Delta q_p = \int_A^B f(q)dq, \quad (7)$$

[1] Asher, H. Cost-Quantity Relationships in the Airframe Industry. RAND, 1956. While this method was in use at the time in the airframe industry and by Air Force cost estimators, a nod is given to Asher for detailing the derivation in his celebrated book.

provided the limit exists. Therefore, instead of summing over all the units in lot L to derive the total lot cost, we may calculate the area under the curve by evaluating the definite integral of the cost progress curve over the interval [A, B]:

$$\int_A^B q^b dq = \frac{B^{1+b}}{1+b} - \frac{A^{1+b}}{1+b}$$
$$= \frac{B^{1+b} - A^{1+b}}{1+b}. \quad (8)$$

The integration is straightforward enough; however, there is a disconnect between the continuous integral and the summation of discrete unit costs. Notice that the limits of the integral in Equation 8 are from A to B. This can be represented by the example given in Figure 2 below, assuming $A = 1$ and $B = 10$.

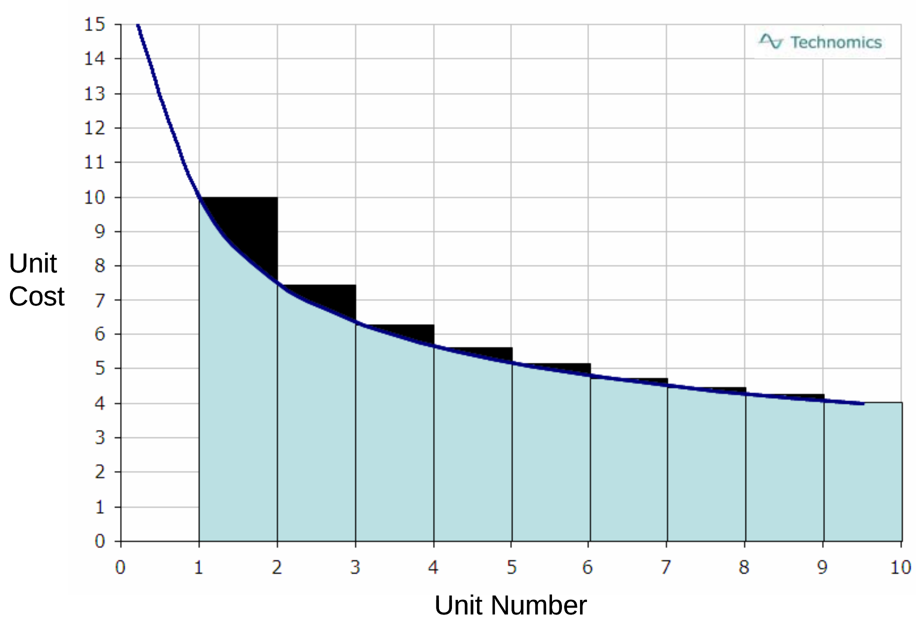


Figure 2. Integral over the interval [0, 10] for $A = 1$, $B = 10$.

The example given in Figure 2 is a cost progress curve with 75 percent learning and a first unit cost of 10. As can be seen in the figure, if $A = 1$ and $B = 10$, the total cost will be underestimated by the integral of x^b with respect to x by the black shaded areas. Furthermore, note that we have 10 units with costs, but the integral is estimating the area over 9 discrete rectangles. However, if we integrate over the interval [0, 10], then the integral *over*estimates the total cost by the black shaded areas shown in Figure 3.

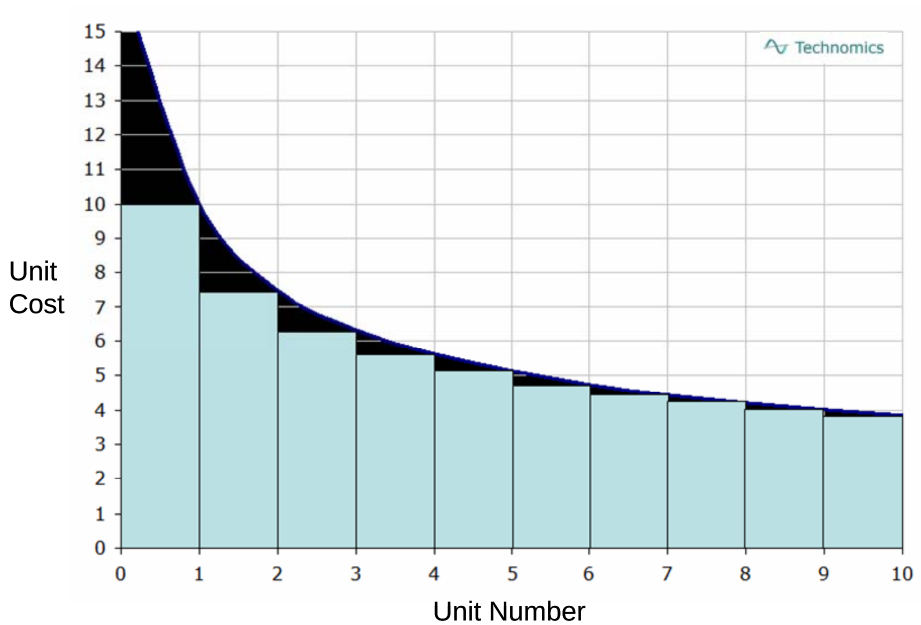


Figure 3. Integral over the interval [0, 10] for $A = 1$, $B = 10$.

Figure 3 introduces the additional confusion of having to integrate over the interval [0, 10] when our units begin at 1 and end at 10. As may be gleaned from Figures 2 and 3, the discrepancy between over- or underestimating discrete units using a continuous integral becomes more severe as slope s increases.

The accepted solution, and one that has been in use by the airframe industry for almost as long as cost progress theory, is to integrate over the interval starting with the A^{th} unit but shift the entire interval by $\pm$ ½ unit.[2] This shift is represented by Figure 4 below.

[2] See Asher (34).

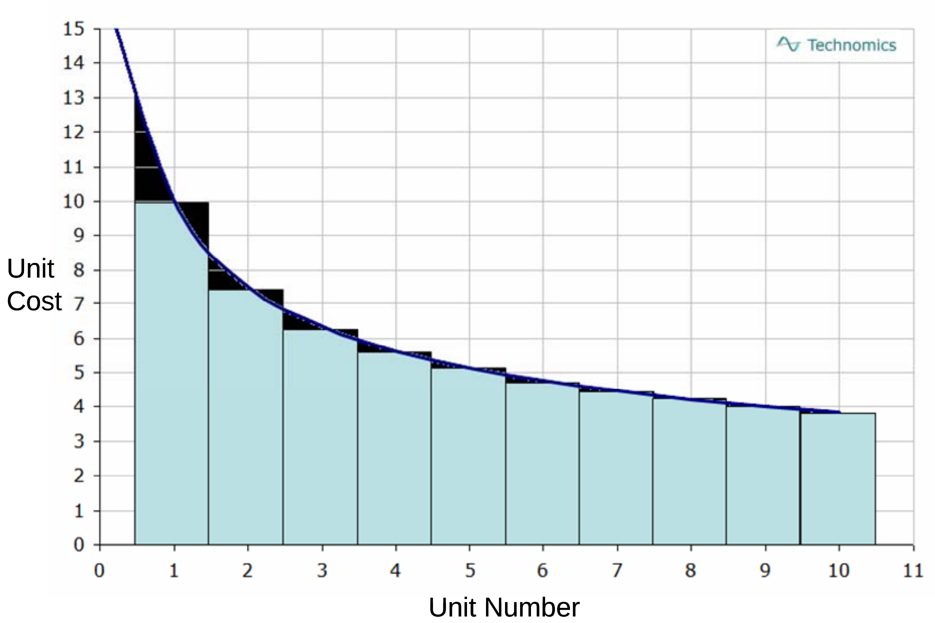


Figure 4. Integral over [A = 1, B = 10] with adjusted endpoints.

In effect, the area underestimated by the integral (the black shaded area under the curve) approximately cancels out the area overestimated by the rectangles (the black shaded area outside the curve). This is represented mathematically by

$$\sum_{q=A}^{B} q^b \approx \frac{(B+0.5)^{1+b} - (A-0.5)^{1+b}}{1+b}. \quad (9)$$ [3]

Substituting Equation 9 into Equation 6, we arrive at Asher's lot midpoint approximation:

$$k \approx \left(\frac{(B+0.5)^{1+b} - (A-0.5)^{1+b}}{(B-A+1)(1+b)} \right)^{\frac{1}{b}}. \quad (10)$$

This approximation is highly accurate in most cases, and suffers only for lower unit numbers if the slope is steep, say < 70 percent. Table 1 provides several midpoints calculated for a selection of lot units and learning curves. The approximation increases in accuracy as the severity of the slope decreases and/or the number of units increases.

[3] In its original form, the equation's numerator is represented by Asher as $(n_0 + n_1 + .5)^{1+b} - (n_0 + .5)^{1+b}$, where n_1 is the number of units after n_0 have been produced. With our definition of A and B, $n_0 + n_1 = B$ and $n_0 + 1 = A$. Notice that if $A = n_0 + 1$ then $(n_0 + .5) = (A - .5)$

First Unit A	Last Unit B	Slope	True Midpoint k	Asher's Approx k	Error
1	3	70%	1.722	1.658	3.716%
1	3	98%	1.812	1.774	2.081%
1	10	70%	3.950	3.878	1.834%
1	10	98%	4.497	4.464	0.726%
10	50	70%	25.915	25.910	0.016%
10	50	98%	27.281	27.279	9.05E-05
1,000	1,500	70%	1,237.198	1,237.198	4.27E-08
1,000	1,500	98%	1,241.312	1,241.312	2.86E-08

Table 1. Error associated with Asher's approximation.

Previously we defined the slope s as $0 < s < 1$. However, this definition does not hold when using Asher's approximation because Equation 10 is undefined if the slope $s = 0.5$ (that is, $b = -1$).

Approximation with Known Slope: Asher's Approximation with Correction Terms (D. Lee, LMI Corp.)

As shown in Table 1 above, Asher's approximation becomes less accurate as the lower unit A becomes smaller and the slope becomes steeper. In his book The Cost Analyst's Companion, Dr. David Lee applies the Euler-Maclaurin summation formula to the evaluated integral in Equation 8 to increase the accuracy of the approximation.[4]

The Euler-Maclaurin formula is used in mathematics to approximate finite sums using integrals, or vice versa. The difference between the summation and the integral is expressed in terms of the difference between the derivatives of the function at the interval endpoints and a remainder term. As shown in Lee (38), when the Euler-Maclaurin formula is applied to the Asher approximation from Equation 9, we arrive at

$$\sum_{q=A}^{B} q^b \approx \frac{(B+0.5)^{1+b} - (A-0.5)^{1+b}}{1+b} - \sum_{r=1}^{p} \frac{1-2^{1-2r}}{(2r)!} B_{2r} \left(f^{(2r-1)}(B+0.5) - f^{(2r-1)}(A-0.5) \right) + R \,.$$

(11)

In other words, the summation of q^b is approximated by the definite integral of $q^b\, dq$ from A to B, minus a correction term, plus a remainder. The term $f^{(2r\text{-}1)}$ gives odd-numbered derivatives of the function $f(q) = q^b$; B_{2r} is the $2r^{th}$ Bernoulli number; and the

[4] Lee, David, The Cost Analyst's Companion, LMI, 1997.

remainder R is an error term that is usually small for suitable values of p. The remainder R is a truncated error term expressed as an integral; as stated in Lee (39), this remainder may be neglected in this application.

While Asher's approximation suffers in accuracy for low unit numbers and steep slopes, the Euler-Maclaurin correction allows for eight-figure accuracy for A as low as 5 and for slopes near 50 percent (40). This method is executed by a Visual Basic routine written by Dr. Lee and published in his book. The routine is used as Dr. Lee intended, taking into consideration that approximations are generated only "when they give the desired accuracy in less computing time than would be required to sum the defining series" (41). In light of this, the corrected Asher's approximation is used when the lower unit A is greater than 4. If the upper unit B is less than 5, the summation $\sum q^b$ is calculated directly. If A is less than 4 and B is greater than 5, the operations are split to maximize accuracy and minimize execution time. For example, if the lower unit is 2 and the upper unit is 10, the direct sum operation is performed for units 2 through 4, while the corrected approximation is performed for units 5 through 10. The sum of these operations provides a result that is much more accurate than using Asher's approximation and much quicker than evaluating the summation explicitly. Table 2 below is a reprint of Table 1, but now with the corrected approximations alongside the true midpoint and Asher's approximation. The accuracy of the corrected midpoints has increased dramatically. Note that the two corrected approximations in Table 2 for $B = 3$ are not shown. The actual error given by the algorithm would be zero, because when $B < 5$ the summation is directly evaluated rather than approximated.

First Unit A	Last Unit B	Slope	True Midpoint k	Asher's Approx k	Asher's Error	Approx with Correction k	Corrected Error
1	3	70%	1.722	1.658	3.716%	-	-
1	3	98%	1.812	1.774	2.081%	-	-
1	10	70%	3.950	3.878	1.834%	3.950	3.96E-10
1	10	98%	4.497	4.464	0.726%	4.497	1.08E-10
10	50	70%	25.915	25.910	0.016%	25.915	1.02E-11
10	50	98%	27.281	27.279	9.05E-05	27.281	2.24E-12
1,000	1,500	70%	1,237.198	1,237.198	4.27E-08	1,237.198	1.47E-15
1,000	1,500	98%	1,241.312	1,241.312	2.86E-08	1,241.312	2.34E-14

Table 2. Error associated with Asher's approximation and the corrected approximation.

Because this method is a correction to Asher's approximation, again the initial definition of the slope s as $0 < s < 1$ does not hold because Equation 11 is undefined if $s = 0.5$.

Parameter-Free Midpoint Formulas

While the Asher and corrected Asher techniques described above produce highly accurate estimates of the true lot midpoint in most cases, they are both calculated assuming the slope of the learning curve is known. Since the true lot midpoint is a function of the algebraic slope b, clearly a method of estimating k that employs the actual slope will be more accurate than any method that does not consider the slope term. With that said, there are techniques that can arrive at a relatively accurate estimate of k without making any assumptions on the parameter b. These techniques are useful when one does not wish to make any inferences on the rate of learning.

Parameter-Free Approximation: Rule-of-Thumb

A standard rule-of-thumb for estimating the lot midpoint without a learning curve is given by

$$k \approx \frac{B - A + 1}{2} + A - 1 . \quad (12)$$

This is a straightforward estimate of any interval's midpoint, where the average number of units in lot L, $(B - A + 1) / 2$, is added to the number of units prior to lot L (that is, $A - 1$). Equation 12 becomes more accurate as the number of units increases, because the learning curve will flatten out regardless of slope. For lower values of A, however, the rule-of-thumb can be quite dangerous, and can create errors near 50 percent when applied in certain situations. For example, the true midpoint of a first lot with 15 units and 60 percent learning is 5.011; but the rule-of-thumb gives an estimate of 7.5. The solution to this first-lot problem is to estimate the first lot with a separate rule-of-thumb, usually

$$k_1 \approx \frac{B}{3} + 1 . \quad (13)$$

While Equation 13 does indeed reduce the error of the basic rule-of-thumb (it estimates a lot midpoint of 6 in the example just given), this technique is still not recommend for anything other than back-of-the-envelope estimates.

Parameter-Free Approximation: Mean (D. Nussbaum, Naval Postgraduate School)

In his article *Evaluation of an Alternative Estimator of Learning Curve Lot Midpoints*, Dr. Daniel Nussbaum proposes a novel technique for estimating a lot midpoint without knowing the parameter *b*.[5] Nussbaum employs a property known as the Arithmetic-Geometric Means Inequality to, in effect, narrow the possibilities of how the true lot midpoint could vary if it needs to be estimated without knowing *b*.

The geometric mean of a series of numbers is given by

$$g = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}}, \quad (14)$$

and simplifies to $\sqrt{AB}$ when calculating the geometric mean of lot *L*:

$$g = \left(\prod_{q=A}^{B} q \right)^{\frac{1}{2}}$$

$$g = (AB)^{\frac{1}{2}}.$$

The arithmetic mean of *A* and *B* is simply $a = (A + B) / 2$. If *k* represents the true lot midpoint, it can be shown that

$$g \le k \le a. \quad (15)$$ [6]

Dr. Nussbaum suggests that if the true lot midpoint is bounded by the geometric and arithmetic means, an estimator also within the bounds of *g* and *a* should be a good predictor of *k*. The obvious estimator is the average of *g* and *a*, which thus becomes the non-parameterized lot midpoint estimator:

---

[5] Nussbaum, Daniel, "Evaluation of an Alternative Estimator of Learning Curve Lot Midpoints", Society of Cost Estimating and Analysis, Spring 1994.

[6] Note if $f(q) = T_1 q^b$ then $f(g) \ge f(a)$ if $b < 0$.

$$k \approx \frac{g+a}{2}$$
$$= \frac{\sqrt{AB} + \frac{(A+B)}{2}}{2}$$
$$= \frac{A+B+2\sqrt{AB}}{4}. \quad (16)$$

As is to be expected with non-parameterized estimators, the error increases as the slope increases and/or *A* decreases. Dr. Nussbaum himself suggests that Equation 16 is a "poor estimator" for lot 1 (12). In light of this, it has been suggested that a rule-of-thumb be employed for lot 1, similar (although a different equation) to that given in the previous estimator "Rule-of-Thumb" above for lot 1.[7] This rule-of-thumb for lot 1 is

$$k \approx \begin{cases} B/2 & if \ B < 10; \\ B/3 & otherwise \end{cases} . \quad (17)$$

However, the error in the rule-of-thumb given in Equation 17 is highly dependent on the slope and the size of the first lot, much more so than even the mean estimator given in Equation 16. Table 3 gives the true midpoint, the "Mean" estimator of Equation 16, and the rule-of-thumb estimator for lot 1 of size *B* = 2, 3, …, 15 and a learning curve slope of 60 percent. Figure 5 is a graph of the information.

60% Slope

Last Unit B	True Midpoint k	"Mean" Estimated k	ROT Estimated k	"Mean" Error	ROT Error
2	1.35	1.46	1.00	8%	26%
3	1.68	1.87	1.50	11%	11%
4	1.99	2.25	2.00	13%	0.3%
5	2.30	2.62	2.50	14%	9%
6	2.59	2.97	3.00	15%	16%
7	2.87	3.32	3.50	16%	22%
8	3.15	3.66	4.00	16%	27%
9	3.43	4.00	4.50	17%	31%
10	3.70	4.33	3.33	17%	10%
11	3.97	4.66	3.67	17%	8%
12	4.23	4.98	4.00	18%	5%
13	4.49	5.30	4.33	18%	4%
14	4.75	5.62	4.67	18%	2%
15	5.01	5.94	5.00	18%	0.2%

Table 3. Non-parameterized midpoint estimator error at 60 percent slope, lot 1, L = {2, 3, …, 15}.

[7] "Introduction to Learning Curves and Cost Estimating Relationships," Society of Cost Estimating and Analysis, Washington Area Chapter Presentation, March 2002.

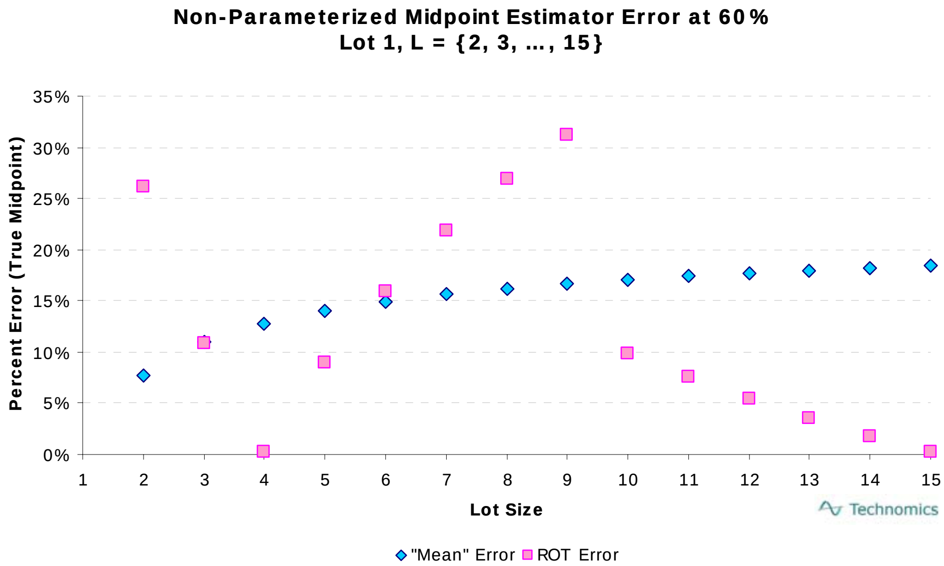


Figure 5. Non-parameterized midpoint estimator error at 60 percent slope, lot size versus error.

The mean estimator is indeed a poor estimator at a 60 percent slope for lot 1, but its error is far less erratic than the error given by the recommended rule-of-thumb for lot 1. In the case of a 60 percent slope, one would want to employ k ≈ B / 2 if B < 7, not B < 10. Compare the 60 percent slope case to the 99 percent slope case, given in Table 4 and Figure 6.

99% Slope

Last Unit B	True Midpoint k	"Mean" Estimated k	ROT Estimated k	"Mean" Error	ROT Error
2	1.41	1.46	1.00	3%	29%
3	1.81	1.87	1.50	3%	17%
4	2.21	2.25	2.00	2%	9%
5	2.60	2.62	2.50	1%	4%
6	2.99	2.97	3.00	0.4%	0.5%
7	3.37	3.32	3.50	1%	4%
8	3.75	3.66	4.00	2%	7%
9	4.13	4.00	4.50	3%	9%
10	4.51	4.33	3.33	4%	26%
11	4.89	4.66	3.67	5%	25%
12	5.27	4.98	4.00	5%	24%
13	5.65	5.30	4.33	6%	23%
14	6.02	5.62	4.67	7%	22%
15	6.40	5.94	5.00	7%	22%

Table 4. Non-parameterized midpoint estimator error at 99 percent slope, lot 1, L = {2, 3, …, 15}.

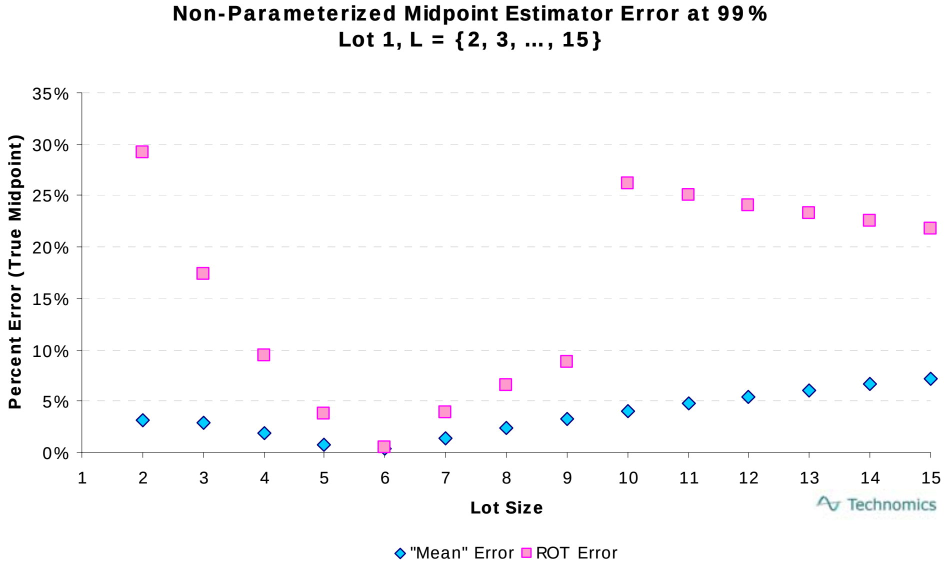


Figure 6. Non-parameterized midpoint estimator error at 99 percent slope, lot size versus error.

In the case of a 99 percent slope, the mean estimator is a good predictor of the true lot midpoint for small lots, with its error gradually becoming higher as the lot size increases. The rule-of-thumb is again erratic, sometimes giving almost 30 percent error, sometimes almost no error. For the rule-of-thumb in this case, for a lot size less than 15, always dividing by 2 would reduce the error.

It is therefore recommended that the analyst ignore the rule-of-thumb and use the mean estimator for any lot size, even lot 1. Any non-parameterized estimator will fare poorly for lot 1, but the rule-of-thumb presented in Equation 17 is far too erratic an alternative to be trusted for general cases.

Parameter-Free Approximation: AGM (J. Lee, Technomics)

As shown in the previous section, the true lot midpoint is bounded by the geometric and arithmetic means of the first and last units of lot L (Eq. 15), so that one non-parameterized estimate of the midpoint is the average of both means. The result (Eq. 16) is a good estimator that can be calculated very quickly using a simple equation. With a computer, however, it is possible to take Equation 16 one step further and find the number at which the geometric and arithmetic means converge. As will be shown, in most trial cases with steep slopes, this number provides a better estimate of the true lot midpoint than simply taking the average of the means.

As explained above, the Arithmetic-Geometric Means Inequality tells us that the geometric mean will always be less than or equal to the arithmetic mean. That is, if $g_0 = \sqrt{AB}$ and $a_0 = (A + B) / 2$, then $g_0 \le a_0$. If we continue these sequences so that $g_{n+1} = \sqrt{a_n g_n}$ and $a_{n+1} = (a_n + g_n) / 2$, then

$$g_n \le g_{n+1} \le a_{n+1} \le a_n . \quad (18)$$

It can be shown that the sequences in Equation 18 will converge to a number between A and B. This number, called the Arithmetic-Geometric Mean and denoted AGM(A, B), is the number such that

$$AGM(A, B) = \lim_{n \to \infty} a_n, g_n . \quad (19)$$

Practically, Equation 19 means calculating the series $\{a_n\}$ and $\{g_n\}$ until the difference between g_{n+1} and a_{n+1} is less than some ε.[8] While it is not mathematically proven here that the AGM is a better or worse estimator than the average of the geometric and arithmetic means for any one slope, empirical evidence shows that it may be better for steeper slopes, for example, a slope less than 70 percent, where the mean method is prone to large errors.

Lot sizes and endpoints were created randomly from A= 1 to B = 500 for each trial, with 5,000 trials for each slope of 50 percent, 60 percent, 70 percent, 80 percent, 90 percent, and 99 percent. For each trial lot, a midpoint was calculated using the true algebraic midpoint, the Mean method detailed in the previous section, and the AGM method. The errors associated with the Mean and AGM methods are summarized and compared in Table 5.

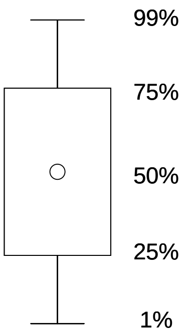
[8] In the Midpoint Calculator, ε = 1E-9.

	Average Error		Max Error	
Slope	Mean	AGM	Mean	AGM
50%	2.22%	1.77%	80%	40%
60%	1.06%	0.62%	30%	14%
70%	0.13%	0.33%	6%	21%
80%	0.58%	0.92%	11%	31%
90%	1.19%	1.55%	20%	40%
99%	1.55%	1.86%	26%	44%

Table 5. Errors associated with the Mean and AGM methods for slopes 50 percent, 60 percent, …, 99 percent.

The average and maximum errors of the AGM midpoint estimates are lower than the Mean method for slopes up to 70 percent. After 70 percent, the Mean method tends to give more accurate midpoint estimates.

Figure 7 displays median box and whisker diagrams for the Mean and AGM errors associated with a slope of 50 percent. The box and whisker diagrams presented here are defined by the following percentiles:



That is, the "boxes" in the plots are defined by the 25th and 75th percentiles, with the 50th percentile (the median) displayed as a point in the center of the box. The "whiskers" are defined by the 1st and 99th percentiles. Each box and whisker combination summarizes data within categories; in Figure 7 the first units of each of the 5,000 trials is bucketed into groups of 50.

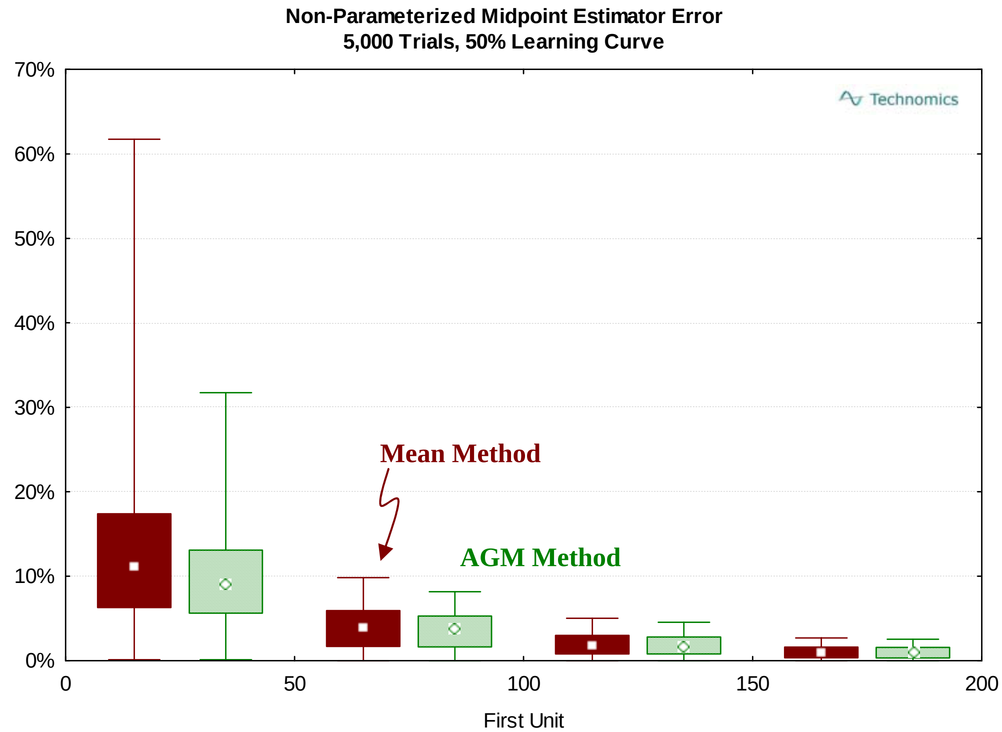


Figure 7. Mean and AGM errors by first unit for a slope of 50 percent, 5,000 trials.

As shown in Figure 7, for a the 50 percent learning curve the Mean method has the potential to produce much larger errors, and to produce them more often, than the AGM method. Figure 8 shows the same graph but for a 99 percent learning curve. As shown in Table 5, as the slope of the learning curve becomes less steep, the midpoint estimation error becomes greater for the AGM method than the Mean method. Both methods produce less error for 99 percent learning curves than for 50 percent learning curves.

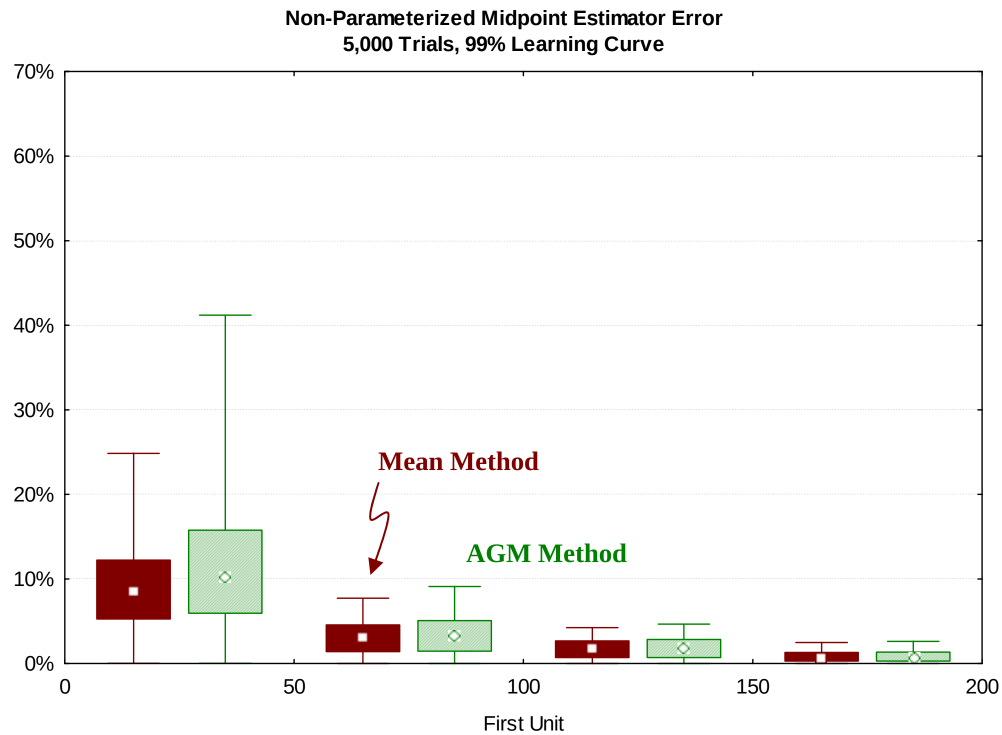


Figure 8. Mean and AGM errors by lot size for a slope of 99 percent, 5,000 trials.

Deciding on Which Midpoint Approach to Use

The previous discussion makes no recommendations on which midpoint estimators should be used. Which approach depends on the following, listed in no particular order:

- Whether or not the slope is known
- Steepness of the slope
- Cumulative quantity
- Required precision of the midpoint estimate
- Spreadsheet tracking and visibility.

Whether or not the slope is known

If historical data on the same system or an analogous system suggest a specific learning slope, a formula that incorporates that learning slope should be used. An equation that incorporates known learning slope information will provide a more accurate estimate of the lot midpoint than any method that does not. However, guessing a learning curve slope may be more dangerous than estimating a lot midpoint without any slope information.[9]

Steepness of the slope

Some methods produce more accurate results at steeper slopes than others, especially at lower unit numbers. For instance, if the slope is less than 70 percent, the commonly used Asher's approximation can yield significant errors, while the corrected Asher's method does not.

The AGM non-parameterized method provides evidence of more accurate results at slopes around 50-69 percent than the Mean non-parameterized method. The Mean method, however, provides evidence of more accuracy over the AGM method for slopes above 70 percent. Here, we assume that the analyst knows whether the slope is closer to 60 percent or 90 percent.

Cumulative Quantity

Midpoint estimation error is highest for quantities early on the learning curve, especially small quantities, and continually decreases as the cumulative quantity increases. Some non-parameterized techniques recommended using separate rules-of-thumb to account for this large error, but these rules usually provide little improvement.

Required precision of the midpoint estimate

The error associated with estimating midpoints becomes more critical as the average unit cost increases. If precision is important and the learning curve is known, the best answer is provided by the true algebraic midpoint. If only a rough order of magnitude estimate is required, or there is no reason to assume a learning curve, either the non-parameterized Mean or AGM method will provide an accurate enough estimate of the true midpoint (again assuming the analyst has a rough idea of whether the slope is nearer 60 percent or 90 percent).

---

[9] There are many other dangers to guessing a learning curve slope when forecasting costs, chief among which is an incorrect first unit cost.

Spreadsheet tracking and visibility

Practically speaking, sometimes the correct answer might not always be the best answer, especially if the spreadsheet is being validated by a third party or lot numbers are not fully enumerated.

The most accurate answer is given by the true algebraic midpoint, which can be calculated and inserted directly into an Excel spreadsheet by Technomics' Midpoint Calculator. This formula, however, depends on lot numbers being fully enumerated—something that may not be done in practice when creating learning curve spreadsheet models. In spreadsheet models, lots are usually represented by the first and last units of each lot, with the unit numbers in between rarely enumerated. This is because a shortcut to fully enumerating each unit number is to simply calculate the midpoint of the lot. If each lot number was enumerated, each unit cost could be calculated, and units within lots summed; hence, there would be no need for a shortcut midpoint calculation!

In addition, Asher's corrected approximation and the non-parameterized AGM method are both calculated using user-defined Visual Basic for Applications (VBA) functions through the Midpoint Calculator and are inserted into the spreadsheet as answers only. A midpoint estimate in a spreadsheet cell that lacks a formula cannot be validated without an explanation of the VBA functions. Therefore, these two approaches may not be the best methods for estimating midpoints based on the ultimate required visibility of the calculations. An alternative to Asher's corrected approximation is simply Asher's approximation, and the alternative to the AGM method is of course the Mean method. Both of these alternatives require one formula each and no VBA routine.