Improved Baseline Method to Calculate Lost Construction Productivity

Tong Zhao, Ph.D., P.E., M.ASCE[1]; and J. Mark Dungan[2]

Abstract: The measured mile/baseline method has been widely accepted to quantify labor productivity loss, which is demonstrated by comparing the impacted and unimpacted/lightly impacted portions of the work. Although the distinctions between these portions of a project can at times be identified through a cause and effect analysis, on many projects this distinction is not readily observable. For those projects, researchers and professionals have developed various procedures to implement the measured mile/baseline calculations, but shortcomings in those procedures can result in the failure to objectively identify the baseline. In this paper, a method based on basic statistical techniques is proposed to determine a baseline that represents the contractor's normal operating performance, thus overcoming many of weaknesses in the existing methods. This paper will provide construction professionals and engineers with an objective approach to determine the productivity baseline, thus aiding in the resolution of labor productivity loss claims. Further, this new method avoids the arbitrary baseline sample size and the possibility of multiple competing solutions in existing methods. A numerical example is included to compare the results using different methods and demonstrate the advantages of the proposed method. **DOI: 10.1061/(ASCE)CO.1943-7862.0000800.**

Author keywords: Claims; Productivity; Project management; Quantitative analysis; Quantitative methods.

Introduction

Quantifying lost labor productivity during construction is one of the most difficult and contentious areas in construction claims because it is usually not tracked contemporaneously, making it difficult to establish both causation and entitlement regarding its recovery. Although different methods are available for productivity loss calculation, the measured mile method is by far the most widely accepted in construction industry (Thomas 2007). In this paper, productivity is defined as the ratio of input to output, but as a matter of personal preference, the inverse could also be used.

The original measured mile method compares identical activities in unimpacted and impacted sections of the project in order to quantify the productivity loss resulting from the impact of the events for which the claimant is not responsible. In order to determine the measured mile, Zink (1986) proposed a procedure to identify a reasonable linear or near-linear portion in the labor hours–progress curve to reflect an unimpacted segment that defines the most efficient productivity. Thomas (Thomas and Zavrski 1999; Thomas 2007) introduced the "baseline" concept for those cases when an unimpacted section of the project cannot be found. Since this baseline may include some productivity loss, it is a conservative measurement from the claimant's perspective. However, Thomas's proposed procedure arbitrarily selected 10 percent of the reporting periods with the best production as the baseline subset. Ibbs and Liu (2005) proposed a procedure based on the K-means clustering technique to address Thomas's arbitrary size of the baseline subset and Gulezian and Samelian (2003) proposed a procedure using control charts to determine the baseline. Both methods, however, have flaws that may result in an erroneous conclusion on productivity baseline, as presented later in this paper.

In this paper, an improved baseline method is proposed, combining the advantages of existing methods while avoiding their drawbacks. In this newly proposed method, productivity data is divided into two groups, the unimpacted/lightly impacted and the heavily impacted, using the overall average productivity. Then the baseline is refined from the unimpacted/lightly impacted group by eliminating extreme data points that may not reflect the contractor's normal productivity levels. This new method can not only determine a baseline that represents the attainable and sustained labor productivity, but also minimize the subjectivity in baseline selection.

The improved baseline method presented herein focuses on determining the performance baseline; and this paper does not intend to elaborate on the importance and procedures of data processing, cause and effect analysis, or measuring productivity loss and responsibility allocation.

Improved Baseline Method

When a measured mile or baseline is not obvious, and is difficult to find through only cause and effect analysis, objective methods to establish the baseline are highly appreciated by the industry. Each of the above-mentioned existing methods has advantages, but their shortcomings are also obvious.

In this paper, an improved baseline method is proposed, which will combine the advantages in the above-mentioned methods while diminishing their shortcomings. In the proposed method, the basic principle of labor productivity loss calculation (i.e., comparing the attainable and sustained labor productivity during the unimpacted/lightly impacted sections to the productivity in the impacted sections) is still adopted. The baseline subset is then defined

[1]Manager, Delta Consulting Group, Inc., 4330 Prince William Parkway, Suite 301, Woodbridge, VA 22192 (corresponding author). E-mail: tzhao@delta-cgi.com

[2]Executive Vice President and Co-Founding Partner, Delta Consulting Group, Inc., 4330 Prince William Parkway, Suite 301, Woodbridge, VA 22192. E-mail: mdungan@delta-cgi.com

Note. This manuscript was submitted on March 4, 2013; approved on October 1, 2013; published online on November 25, 2013. Discussion period open until April 25, 2014; separate discussions must be submitted for individual papers. This technical note is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364/06013006(4)/$25.00.

as the section of the work in which the productivity reflects the contractor's normal operating performance, which is attainable and sustained, although not necessarily continuous in time. The proposed approach for determining baseline productivity comprises the following general steps:

- Segregate the data into the good productivity group and the bad productivity group using the overall average productivity; and
- Determine the baseline subset from the good productivity group using statistical techniques, such as a process control chart, and then the baseline productivity is calculated as the average productivity of the baseline subset.

Data Segregation

Productivity can be quantified as the ratio between input and output, and it is affected by disruptions. The causal relationship between disruption and productivity remains at the conceptual level, i.e., the more severe the disruptions, the worse the productivity. The data points with good productivity are normally encountered when no disruptions or light disruptions are experienced. It is reasonable to infer that the productivity observed in the sections of the work without any assignable disruptions or with light disruptions should be better than the overall average productivity, the impacted and unimpacted combined.

Therefore, the first step for the proposed approach is to calculate the overall average productivity, and then assign all the data points of the reporting periods or sections with a productivity that is better than the overall average to the good productivity group (GPG). The data points with a productivity that is worse than the overall average constitute the bad productivity group (BPG). This data segregation process is demonstrated mathematically below:

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m) \tag{1}$$

where $\mathbf{U}$ = the whole data set, and $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_m$ = first, second, and mth data point of set $\mathbf{U}$; and

$$\mathbf{u}_i = (h_i, q_i, p_i) \quad \text{where } i = 1, 2, \ldots, m \tag{2}$$

in which h_i, q_i, and p_i = the input (labor hours), output (production), and productivity of the ith data point in set $\mathbf{U}$, and

$$p_i = \frac{h_i}{q_i} \tag{3}$$

The overall average productivity can be calculated as

$$\bar{p} = \frac{\sum_{i=1}^{m} h_i}{\sum_{i=1}^{m} q_i} \tag{4}$$

Then the GPG can be defined as

$$\mathbf{G} = \{\mathbf{u}_i | p_i \le \bar{p}, \forall i\} \tag{5}$$

Baseline Determination

Although the average productivity for the GPG is a good approximation for the baseline productivity, the GPG may still include extreme data points that do not represent the contractor's normal operating process. Some of them may be impacted by the assignable disruptions, and some others may be caused by clerical errors, data update delays, or the inclusion of dissimilar work on some data points. Thus, a control chart technique is applied to the GPG to eliminate the extreme points. After the extreme points are eliminated, the remaining data points that would represent the contractor's normal or near-normal operating process constitute the baseline subset.

The control chart calculation in the proposed method has two major distinctions from what Gulezian and Samelian (2003) presented. First, the control chart technique is applied only to the GPG rather than the whole data set. Second, the control chart technique is revised to use the average productivity of the GPG as the center line instead of the arithmetic mean of the individual productivity values from different reporting periods/sections, thus addressing the production differences among reporting periods/sections. The calculation process of using the control chart technique to identify productivity baseline is presented below.

1. Select the GPG as the initial data set $\mathbf{D}$ for the control chart.
2. Calculate the average productivity and the standard deviation for set $\mathbf{D}$ as follows:

$$\bar{p}_{\mathbf{D}} = \frac{\sum_{\mathbf{u}_i \in \mathbf{D}} h_i}{\sum_{\mathbf{u}_i \in \mathbf{D}} q_i} \tag{6}$$

$$\sigma_{\mathbf{D}} = \sqrt{\frac{\sum_{\mathbf{u}_i \in \mathbf{D}} q_i (p_i - \bar{p}_{\mathbf{D}})^2}{\sum_{\mathbf{u}_i \in \mathbf{D}} q_i}} \tag{7}$$

3. Calculate set $\mathbf{D}'$ using the following equation to eliminate the out-of-control points:

$$\mathbf{D}' = \{\mathbf{u}_i \in \mathbf{G} | \bar{p}_{\mathbf{D}} - 3\sigma_{\mathbf{D}} \le p_i \le \bar{p}_{\mathbf{D}} + 3\sigma_{\mathbf{D}}, \forall i\} \tag{8}$$

   If $\mathbf{D}' = \mathbf{D}$, or no additional points fall out the control limits, then go to 4; otherwise, rename $\mathbf{D}'$ to $\mathbf{D}$, go to 2.
4. Set $\mathbf{D}$ is the baseline subset, and the baseline productivity equals the average productivity of set $\mathbf{D}$, or $\bar{p}_{\mathbf{D}}$.

Numerical Example and Discussions

An example, which comprises 37 daily productivity data points for HVAC ductwork in a three-story reinforced concrete building, the same project discussed by Ibbs and Liu (2005), is now used to demonstrate the proposed method and compare it to other methods. The data for the numerical example and identified baseline subsets using various methods are summarized in Table 1, and calculation results are compared in Table 2. Note that the results using Zink's procedure, Thomas's method, and Ibbs and Liu's method are cited from Ibbs and Liu (2005).

A measured mile productivity of 3.22 h/m determined using Zink's approach is worse than the overall average productivity. In other words, the identified measured mile may not be able to demonstrate productivity loss. As indicated in Ibbs and Liu (2005) and Thomas (2007), in practice it is often difficult to find a continuous time period free from disruption. According to Gulezian and Samelian (2003), the presence of natural productivity variation is another way to explain why a measured mile, during which there is a linear or near-linear relationship between labor hours and production, is difficult to identify.

It appears that Thomas's method yields a baseline with the best productivity compared to any other methods applied to this example. But the performance of Thomas's method is not consistent. If the input is not uniform or nearly uniform among reporting periods, the reporting periods with both the highest production and worst productivity can be included in the baseline subset using Thomas's method, while excluding the unimpacted or lightly impacted periods from the baseline. As Ibbs and Liu (2005) pointed out, other

Table 1. Labor Productivity Loss Calculation with Identified Baseline Set

Work day	Labor hours	Output (m)	Calculated productivity (h/m)	Baselines[a] (Ibbs and Liu 2005) Zink	Thomas	*K*-means	Alternative[a] *K*-means solution	Improved baseline[a]
1	40	14.3	2.80	—	—	C	D	E
2	40	6.3	6.35	—	—	—	—	—
3	40	8.9	4.49	—	—	—	—	—
4	40	13.9	2.88	—	—	C	D	E
5	40	9.8	4.08	—	—	—	—	—
6	24	10.9	2.20	—	—	C	D	E
7	24	15.1	1.59	—	B	C	D	E
8	24	7.4	3.24	—	—	C	D	—
9	24	5.5	4.36	—	—	—	—	—
10	24	5.7	4.21	—	—	—	—	—
11	24	6.07	3.95	—	—	C	—	—
12	24	6.9	3.48	—	—	C	D	—
13	40	7.3	5.48	—	—	—	—	—
14	40	13.0	3.08	—	—	C	D	E
15	40	10.4	3.85	—	—	C	—	—
16	24	6.3	3.81	—	—	C	—	—
17	24	16.7	1.44	—	B	C	D	E
18	24	5.0	4.80	—	—	—	—	—
19	24	4.3	5.58	—	—	—	—	—
20	24	11.6	2.07	—	—	C	D	E
21	24	9.1	2.64	—	—	C	D	E
22	24	9.7	2.47	—	—	C	D	E
23	40	13.2	3.03	—	—	C	D	E
24	40	22.4	1.79	—	B	C	D	E
25	40	7.1	5.63	—	—	—	—	—
26	40	15.2	2.63	A	B	C	D	E
27	40	8.2	4.88	A	—	—	—	—
28	40	12.8	3.13	A	—	C	D	—
29	32	11.8	2.71	A	—	C	D	E
30	40	12.7	3.15	A	—	C	D	—
31	40	14.1	2.84	A	—	C	D	E
32	40	5.5	7.27	—	—	—	—	—
33	40	19.1	2.09	—	B	C	D	E
34	40	13.9	2.88	—	—	C	D	E
35	40	13.9	2.88	—	—	C	D	E
36	24	12.4	1.94	—	—	C	D	E
37	24	8.0	3.00	—	—	C	D	E
Total	1,216	394.47	3.08	—	—	—	—	—

[a]Letters indicate baselines identified by various methods.

problems with Thomas's method include the arbitrary 10 percent requirement for the baseline sample size, and whether median or average values should be used to calculate baseline productivity.

Multiple solutions using *K*-means clustering based techniques are possible even with the same *K* value (Fielding 2006). Though only one solution was reported in Ibbs and Liu (2005), two solutions with different productivity values can be reached using the *K*-means clustering technique as demonstrated in this paper, because the results yielded by this technique rely on the input information, such as initial conditions. The difference between the two solutions is notable (2.61 h/m versus 2.76 h/m), demonstrating that the baseline selection using the *K*-means clustering technique is not conclusive. Another problem with this method is that it lacks assessment on the data in the selected baseline subset regarding whether they fall in a reasonable range that represents the contractor's normal operating performance.

Using Gulezian and Samelian's method, no data points were found to fall outside the initial control limits for this example.

Table 2. Comparison of Calculation Results Using Different Methods

Method	Baseline sample %	Baseline sample Number	Productivity Baseline (h/m)	Productivity Arithmetic mean (AM)	Productivity Overall average (OA)	Baseline/AM (%)	Baseline/OA (%)	Note
Zink	16.2	6	3.22	3.48	3.08	92.5	104.5	Please see Ibbs and Liu (2005)
Thomas (median)	13.5	5	1.80	3.48	3.08	51.7	58.4	for calculation details.
Thomas (average)	13.5	5	1.90	3.48	3.08	54.6	61.7	
K-means	70.3	26	2.76	3.48	3.08	79.3	89.6	
K-means alternative solution	62.2	23	2.61	3.48	3.08	75.0	84.7	—
Gulezian and Samelian	100	37	3.48	3.48	3.08	100	113.0	—
Improved baseline method	51.4	19	2.42	3.48	3.08	69.5	78.6	—

This result would suggest that all performance during all 37 workdays was normal; therefore no productivity loss could be claimed. This method is not suitable for cases when disruptions are pervasive, because although the majority of the reporting periods are impacted, it is very likely that no data points would fall outside the initial control limits.

The proposed method in this paper combines the advantages of the above-mentioned methods, yet it avoids their drawbacks. When the proposed method is applied to the example, the overall average productivity is calculated to be 3.08 h/m. All the reporting periods (work days) in which the productivity is less than or equal to 3.08 h/m are assigned to the GPG, and the remainder are assigned to the BPG. Then the proposed control chart technique is used to determine the baseline subset in the GPG. The average productivity and standard deviation for the initial control chart data set of the GPG are calculated to be 2.42 h/m and 0.52 h/m, respectively. Correspondingly, the lower control limit is 0.84 h/m while the upper control limit is 3.99 h/m. Since no extreme points fall outside the control limits, the calculation stops. The end result is that there are 19 data points (reporting periods) in the baseline subset, which happens to be the same as the GPG, and the baseline productivity is 2.42 h/m.

Conclusions

This paper has compared various methods to perform the measured mile/baseline calculation, and has identified the advantages and weaknesses of the existing methods. The improved baseline method seeks to overcome those weaknesses while utilizing the advantages. The numerical example shows that the proposed method is indeed an improvement. The proposed method only relies on productivity actually achieved on the project in question, which could be different from the industry average on projects with similar nature.

This proposed method is more objective than other methods because a unique baseline subset is selected using impartial, basic statistical techniques. First, it does not require the baseline set to be continuous in time, which overcomes the limitation in Zink's approach. Second, the sample size of the baseline subset generated is defined by the characteristics of the productivity data themselves, and does not have to rely on an arbitrary percentage as used in Thomas's method. Third, the proposed method avoids the possibility of multiple competing solutions that may exist for using Ibbs and Liu's method. Therefore, this paper has presented a new and unique method that would contribute to the body of knowledge of ASCE by providing construction professionals and engineers an objective approach to determine productivity baseline to resolve labor productivity loss claims.

References

Fielding, A. H. (2006). *Cluster and classification techniques for the biosciences*, Cambridge University Press, Cambridge, U.K.

Gulezian, R., and Samelian, F. (2003). "Baseline determination in construction labor productivity-loss claims." *J. Manage. Eng.*, 10.1061/(ASCE)0742-597X(2003)19:4(160), 160–165.

Ibbs, W., and Liu, M. (2005). "Improved measured mile analysis technique." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2005)131:12(1249), 1249–1256.

Thomas, H. R. (2007). "Quantification of economics losses caused by labor inefficiencies." *Proc., 2007 Construction SuperConf.*, Vendome Group, New York.

Thomas, H. R., and Zavrski, I. (1999). "Construction baseline productivity: Theory and practice." *J. Constr. Eng. Manage.*, 10.1061/(ASCE)0733-9364(1999)125:5(295), 295–303.

Zink, D. A. (1986). "The measured mile: Proving construction inefficiency costs." *Cost Eng.*, 28(4), 19–21.