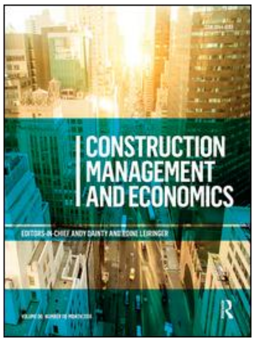

Construction Management & Economics

ISSN: 0144-6193 (Print) 1466-433X (Online) Journal homepage: https://www.tandfonline.com/loi/rcme20

The evidence of complex variability in construction labour productivity

Milan Radosavljević & R. Malcolm W. Horner

Construction Management and Economics (2002) **20**, 3–12

The evidence of complex variability in construction labour productivity

MILAN RADOSAVLJEVIĆ* and R. MALCOLM W. HORNER

Construction Management Research Unit, Department of Civil Engineering, University of Dundee, Dundee DD1 4HN, UK

Received 19 June 2000; accepted 24 September 2001

The complex variability of the 12 construction labour productivity data sets has been examined by analysing the central moments of tendency, and applying the Kolmogorov–Smirnov and Anderson–Darling tests of normality. The results consistently show that the productivity is not normally distributed. In addition, undefined variance causes a failure of the central limit theorem, thus indicating that some basic statistical diagnostics like correlation coefficients and t statistics may give misleading results and are not applicable. A brief comparison with volatility studies in econometrics has revealed surprising similarity with Pareto distributions, which can model undefined or infinite variance. Such distributions are typical of chaotic systems like the logistic equation, whose properties also are described briefly. Therefore, it is suggested that future research should be focused on studying the applicability of chaos theory to construction labour.

Keywords: Construction, labour productivity, normal distribution, tests, chaos

Introduction

Variability has been shown to be a key factor in our study of the behaviour of construction labour productivity (Noor, 1992; Lema and Price, 1996). Construction projects are often turbulent because of the number of variables involved, the labour intensive work, the unique character, and the occurrence of unpredictable events (Choromokos and McKee, 1981; Arditi, 1985; Thomas and Yiakoumis, 1987; Thomas *et al.*, 1990; Horner and Talhouni, 1995; Kaming *et al.*, 1996). These factors are major causes of stagnation in the industry in terms of low profitability and productivity. Manufacturing has increased its productivity by more than 100% in the past 30 years whereas construction is in decline (Choromokos and McKee, 1981; Briscoe, 1988). It is no wonder, then, that most of the research on productivity issues has been concerned mainly with aspects that should lead to improvements (Tavakoli, 1985; Thomas and Yiakoumis, 1987; Proverbs *et al.*, 1999), like poor on-site management (Choromokos and McKee, 1981; Oglesby *et al.*, 1989) and the impact of delays and interruptions (Horner and Talhouni, 1995). Labour intensive work, unique design, the number of factors affecting on-site work and other variables make the construction industry unstable in its performance. These variables create many undesirable events that threaten planned budgets and cause major delays. Thus, often the budget is exceeded in order to meet the client's requirements and complete a project within specified period. Hence, a high net profit margin seems to be the exception rather than the rule. Consequently, improving labour productivity should be a major concern for research and for the industry. Observing and understanding the complex behaviour of productivity is a starting point for seeking

*e-mail: m.radosavljevic@dundee.ac.uk

Construction Management and Economics
ISSN 0144–6193 print/ISSN 1466-433X online
http://www.tandf.co.uk/journals
DOI: 10.1080/01446190110098961

improvement. Appropriate descriptions of complex variability can reveal information to enhance our understanding of the underlying processes.

Objectives

This study demonstrates that construction labour productivity involves dynamic interactions within a complex system of variables. Therefore, the basic assumption is non-normality, which is common to many natural processes (Peters, 1991; Gleick, 1993; Kaye, 1993; Lorenz, 1995). Several reports have shown that economic fluctuations, which are similar to fluctuations in productivity, correspond to some internal non-periodic cyclical order that is not random behaviour (Peters, 1991; Papaioannou and Karytinos, 1995). If non-normality is one of the properties of productivity, then our knowledge and understanding of the processes that generate such behaviour are inadequate. The objective is to reveal evidence that normality (random behaviour) is not a characteristic of productivity. Some recent studies of economic fluctuations show similar non-normal behaviour, and suggest that these may follow the laws of chaos theory. This is also the main reason for suggesting the application of chaos theory in future research on productivity.

The present study is based upon applied statistical techniques. It starts from the core knowledge of some basic statistical properties in order to reveal evidence that classical statistics cannot fully describe productivity. This is further shown to be the major obstacle to achieving a clear understanding of the construction industry's performance and its internal project-related dynamism, which can lead to enormous fluctuations in productivity.

Methodology and data

Five formwork and seven masonry data sets from 11 different sites in the USA and the UK were analysed. Originally they were in the following form.

$$P_i = \frac{\text{Total man hours in day } i}{\text{Unit output in day } i} \quad \cdots \quad \frac{h}{m^2} \tag{1}$$

Data sets were normalized so that they have a zero mean and a standard deviation of 1:

$$P_{ni} = \frac{P_i - \mu_p}{\sigma_p} \tag{2}$$

where μ_p is the sample mean and σ_p is the sample standard deviation.

Normalization was performed due to the soundness of the results; however, the data remained topologically invariant (Peters, 1991). The inverse data were obtained for comparative purposes and for an examination of extreme values. Several different statistical techniques were used to obtain a more accurate representation of the suggested hypothesis of non-normality:

method of moments (basic description of the distribution);
visual inspection (comparison between the properties of normally distributed data and productivity data under inspection);
goodness of fit tests
  Kolmogorov–Smirnov test,
  Anderson–Darling test.

Productivity is analysed in a purely statistical manner. That is why the study does not provide any qualitative information about the construction sites but rather examines common statistical characteristics. The purpose is to demonstrate some general properties common to all data sets irrespective of project nature or geographical origin. In essence, the intention is to show that these common characteristics could be revealed by statistical analysis even when the nature of the projects is not known.

Defining a distribution

In order to analyse the validity of the assumption that labour productivity is normally distributed, we need a thorough insight into a normal distribution. What actually characterizes normality? Often these characteristics are not properly considered, and thus the results may rely on assumptions that are incorrect. Standard practice in the literature on statistics is to summarize the probability density function by its moments (Kottegoda and Rosso, 1997). Despite the fact that the mean and standard deviation (first two moments) fully describe a normal distribution, at least two others should be given attention (skewness and kurtosis) if the distribution is to be sufficiently defined (Hahn and Shapiro, 1994). The first four moments or population measures are the mean (expected value), variance (measure of dispersion), skewness (measure of asymmetry) and kurtosis (measure of peakedness). These four moments should be defined before performing any statistical analyses in order to verify the validity of applied techniques. Nevertheless, this is rarely a common procedure in productivity related research. That is why this study emphasizes the importance of the first four population measures and provides a thorough description of them in the first place. The objective here is to demonstrate the nature of the population measures and, consequently, to show what they tell us about a particular data set.

Let us call μr^* the moment of order r about the point a. It is defined for a discrete variable by [(Kottegoda and Rosso, 1997):

$$\mu_r^* = E\ (X - a)^r = \sum_{all\ x_i}(x_i - a)^r P_X(x_i) \tag{3}$$

where r is a positive integer denoting the order of a moment and $P_X(x_i)$ is a probability density function of a random variable X.

Now let X be a random variable with probability density function $P_X(xi)$. The expected value of X represents the first moment of a random variable X:

$$E[X] = \sum_{all\ x_i} x_i p_X(x_i) \tag{4}$$

If the sum on the right is infinite or does not exist, than the expected value of X also does not exist or is unidentifiable (Conover, 1980). The first moment is the mean of X:

$$\mu_x = E(X)$$

$$\mu_x = \sum_{all\ x_i} x_i p_X(x_i) \tag{5}$$

In the same way we can defined the second moment (it is called the second central moment of tendency because a is actually the mean) or variance by writing:

$$\mu_2 = \sigma^2 = E\ (x - \mu_x)^2$$

$$\mu_2 = \sum_X (X - \mu_x)^2 f(X) = E(X^2) - \mu_x{}^2 \tag{6}$$

The third moment denotes the symmetry of a particular distribution function and is called skewness, defined by:

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2}^3} = \frac{E[(X - E[X])^3]}{\sqrt{\{E[(X - E[X])^2]\}^3}}$$

$$\gamma_1 = \frac{E[X^3] - 3E[X^2] \cdot E[X] + 2E[X])^3}{(E[X^2] - (E[X])^2)^{3/2}} \tag{7}$$

If the probability distribution is symmetrical about the mean, the third moment is equal to zero. If the values of X greater than the mean are more dispersed, then the distribution is positively skewed (dominant tail on the right). If the dominant tail is on the left side, the distribution is negatively skewed.

Kurtosis, or measure of peakedness, is the fourth moment of a distribution. Kurtosis is a condition where extra observations are present in the tails of the distribution and often it is called a fat-tails phenomenon. By using moments of order 4 and 2 we can define it by:

$$\gamma^2 = \frac{\mu_4}{\mu_2^2} = \frac{E[(X - E[X])^4]}{\{E[(X - E[X])^2]\}^2} \tag{8}$$

$$\gamma_2 = \frac{E[X^4] - 4E[X^3]E[X] + A}{\{E[X^2] - (E[X])^2\}^2} + \frac{6E[X^2](E[X])^2 - 3(E[X])^4}{\{E[X^2] - (E[X])^2\}^2}$$

Represented population measures of the random variables from which the sample is realized correspond to sample estimates (mean, standard deviation, coefficient of skewness and coefficient of kurtosis; Kottegoda and Rosso, 1997) when samples are numerically analysed.

1. The sample arithmetic mean (measure of central tendency):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{9}$$

for a sample $x_1, x_2, \ldots, x_n$.

2. The sample standard deviation (measure of dispersion):

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - \bar{x}^2)} \tag{10}$$

3. The sample coefficient of skewness (measure of asymmetry):

$$g_1 = \sum_{i=1}^{n}(x_i - \bar{x}^3)\Big/ns^3 \tag{11}$$

4. The sample coefficient of kurtosis (measure of peakedness):

$$g_2 = \sum_{i=1}^{n}(x_i - \bar{x}^4)\Big/ns^4 \tag{12}$$

Normal distribution

Generally, the objective of this section is to demonstrate the theoretical fundamentals of a normal distribution. This first step is of great importance because it reveals basic properties of normally distributed random numbers that are to be rejected in the case of productivity data. Normal distribution is a continuous probability distribution with the probability density function (Conover, 1980, Kottegoda and Rosso, 1997):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x-\mu)^2/2\sigma^2} \quad \text{for all } x \text{ and } \sigma > 0 \tag{13}$$

It is obvious that, using the method of moments, the mean and standard deviation are defined as:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x-\mu)^2/2\sigma^2}\, dx = \mu \tag{14}$$

and

$$\int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x-\mu)^2/2\sigma^2}\, dx = \sigma^2 \tag{15}$$

The normal distribution with $\mu = 0$ and $\sigma = 1$ is called a standard normal distribution:

$$f_s(x) = \frac{1}{\sqrt{2p\pi}} e^{x^2/2} \tag{16}$$

The important properties of a normal distribution are the following (Kottegoda and Rosso, 1997).

1. A linear transformation $y = a + b{\cdot}X$ or an $N(\mu,\sigma^2)$ variate X makes Y an $N(a + b\mu, b^2\,\sigma^2)$ variate.
2. The moment generating function is given by: $M_X(t) = e^{(\mu t\ +\ \sigma^{2\cdot} t^{2/2})}$
3. Let Y_n be the sum of the n random variables $X_1, X_2,\ldots, X_n$ with mean μ_n and standard deviation μ_n. As n approaches infinity the distribution function of the random variable $(\gamma_n - \mu_n)/\sigma_n$ approaches the standard normal distribution function. If X_i is normally distributed the result holds regardless of sample size. This is the 'central limit theorem'.
4. The mean $\bar{X}$ and variance S^2 of n independent normal variates $X_1, X_2,\ldots, X_n$ are independent.

The above definitions finally give the resulting properties of a normal distribution that can be summed up in the following manner (Martin, 1971).

- According to the central limit theorem mean μ and variance σ^2 must be finite.
- The odd-order moments are zero; therefore skewness equals zero (by virtue of the symmetry).
- The fourth moment (kurtosis) equals three ($g_2 = 3$).

If productivity is truly normally distributed then the following analyses should give the above results. If not, then we cannot use some of the statistics because they are based on independent identically distributed random variables.

Statistical description of original and inverse data

The purpose of the initial analyses is to study the basic statistical properties of the original and inverse data (Tables 1 and 2). The results clearly show enormous differences between the two forms of data representation. If we take into account the form in which productivity originally was presented (total time in hours)/(unit of work done), extreme values seem to be rather a frequent phenomenon. However, the original data exhibit results that are more consistent. Even at first glance we can detect high values of skewness and kurtosis in original data. However, the values are low, erratic, or even negative in the inverse data. Large positive values of skewness in the original data signal that the distribution is fat-tailed to the right. Hence, the distribution is non-symmetrical. Figures 1 and 2 demonstrate a frequency distribution of normalized data in comparison with a frequency distribution of Gaussian random numbers.

High values of kurtosis indicate a leptokurtotic condition. Leptokurtosis is the phenomenon of higher peaks and fatter tails than predicted by a normal distribution (Peters, 1991). The results of this study show that values closely around the mean and extremes are much more frequent than might be expected. Nevertheless, leptokurtosis is not so noticeable in some of the inverse data. The skewness of the inverse data is exhibiting similar behaviour to that of the kurtosis, and in some cases of inverse data the values are even negative. Further investigation of project 12, which was divided into four separate and supposedly independent sets, demonstrates that the inverse data are fluctuating more than the original (Table 3). Such inconsistent behaviour of skewness and kurtosis was the reason that only the original data were analysed further. Nevertheless, these values already signify that construction productivity does not show a normal distribution. However, the last basic property that needs to be validated is the central limit theorem, by testing the stability of mean and variance.

Stability of mean and variance

The two numerical values are basic statistical properties. Normal distribution can be sufficiently defined by these two parameters (Kottegoda and Rosso, 1997). It belongs to the family of stable distributions and one of its properties is a scaling characteristic (Peters, 1991) defined by the central limit theorem (Martin, 1971; Bethea *et al.*, 1995; Kottegoda and Rosso, 1997). If x is a random variable with a density function $f(x)$, then its expected value, $E[x]$, can be represented in the following way (Martin, 1971):

$$E[x] = \int_{-\infty}^{\infty} x\, f(x)\, dx \tag{17}$$

Further, for a function $h(x)$ of x the expected value is:

$$E[x] = \int_{-\infty}^{\infty} dx\, h(x)\, f(x) \tag{18}$$

It follows that the expected value of the sum of functions $h1(x)$ and $h2(x)$ equals the sum of their expected values:

$$E[h_1\,(x) + h_2\,(x)] = E[h_1\,(x)] + E[h_2\,(x)] \tag{19}$$

Table 1 Basic statistical properties of the 12 data sets containing the original data. The letters F and M in parentheses next to the data set number denote formwork and masonry (type of data), respectively

Data set No.	Sample size	Average	Median	Variance	St. Dev.	Min.	Max.	Range	Skewness	Kurtosis
1 (F)	148	0.5946	0.393474	0.434888	0.65946	0.059151	4.319871	4.260721	3.658211	16.67952
2 (F)	263	2.657874	0.132476	141.1113	11.87903	0.04204	98	97.95796	5.940584	37.95501
3 (F)	183	0.54886	0.367	0.48046	0.69315	0.015	4.778	4.763	3.80279	18.7225
4 (F)	86	2.54574	1.753448	6.681485	2.584857	0.453608	16	15.54639	3.073043	11.35453
5 (F)	109	0.11776	0.10336	0.00879	0.09374	0.04054	0.86364	0.8231	6.65567	47.7262
6 (M)	95	0.097	0.088	0.002617	0.51156	0.034	0.415	0.381	3.66674	18.18
7 (M)	110	1.857	0.238	84.86096	9.212001	0.069	64.740	64.671	6.350723	40.51461
8 (M)	112	0.188296	0.158249	0.008483	0.092102	0.080465	0.527163	0.446698	1.700777	3.367414
9 (M)	91	0.153792	0.126617	0.006824	0.082608	0.074413	0.526678	0.452265	2.509496	7.036744
10 (M)	132	0.189676	0.139812	0.055053	0.234634	0.071742	2.199688	2.127946	6.354858	47.20439
11 (M)	108	0.172046	0.1445	0.009504	0.09749	0.076	0.795	0.719	3.368744	16.46324
12 (M)	470	0.12116	0.102898	0.006044	0.077743	0.007832	0.857878	0.850046	3.741651	22.86058

Table 2 Basic statistical properties of the 12 data sets containing the inverse data. The letters F and M in parentheses next to the project number denote formwork and masonry (type of data), respectively

Data set No.	Sample size	Average	Median	Variance	St. Dev.	Min.	Max.	Range	Skewness	Kurtosis
1 (F)	148	3.229492	2.542295	7.287859	2.699603	0.231488	16.906	16.67451	2.143415	6.085611
2 (F)	263	7.641823	7.432292	25.37099	5.036963	0.0102	23.78691	23.77667	0.486976	0.007363
3 (F)	183	5.334074	2.724796	52.58944	7.251857	0.209293	66.66667	66.45737	4.330602	29.04537
4 (F)	86	0.667743	0.570707	0.189596	0.435426	0.0625	2.204545	2.142045	1.242812	1.825096
5 (F)	109	9.847681	9.675	7.442048	2.728012	1.157895	24.66667	23.50877	1.010206	8.344415
6 (M)	95	11.971	11.364	17.61572	4.197109	2.410	29.412	27.002	0.869371	2.503406
7 (M)	110	4.331	4.202	5.278652	2.297532	0.015	14.493	14.477	0.886984	2.900227
8 (M)	112	6.342179	6.319258	5.810775	2.410555	1.896948	12.4278	10.53085	1.185524	−0.766853
9 (M)	91	7.666152	7.89784	6.465819	2.542798	1.898692	13.4385	11.53981	−0.183102	−0.16366
10 (M)	132	7.341782	7.15271	8.093063	2.844831	0.45461	13.93881	13.4842	−0.071637	−0.459271
11 (M)	108	6.938483	6.920498	6.35237	2.520391	1.257862	13.15789	11.90003	0.301018	0.069466
12 (M)	470	10.53615	9.718372	48.37751	6.955394	1.165667	127.6853	126.5197	10.4593	171.9143

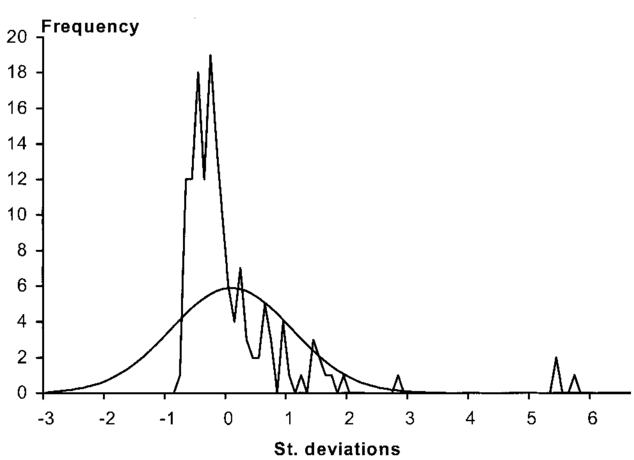

Figure 1 Frequency distribution of data set 1 (formwork) versus normal distribution

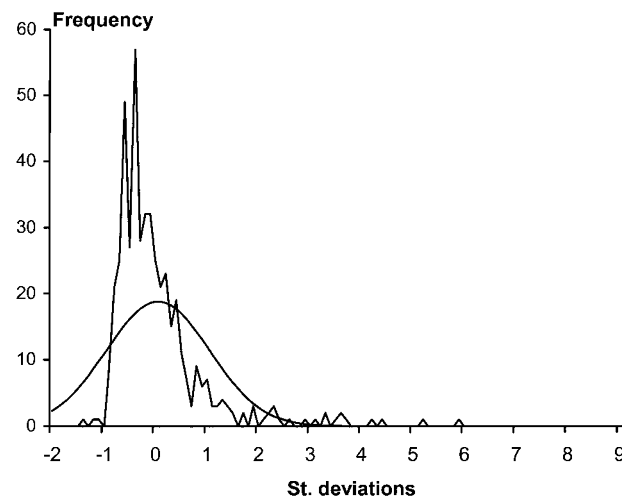

Figure 2 Frequency distribution of data set 12 (masonry) versus normal distribution

Table 3 Summary of four data sets taken successively from project 12 containing 470 data points. The results show that skewness and kurtosis are much more consistent in the case of the original data (O and I in parentheses denote original and inverse data, respectively)

Data set No.	Sample size	Average	Median	Variance	St. Dev.	Min.	Max.	Range	Skewness	Kurtosis
1 (O)	115	0.108392	0.09023	0.005713	0.075587	0.024933	0.577247	0.552314	3.816276	17.98701
2 (O)	115	0.114394	0.089794	0.005447	0.073801	0.048717	0.52277	0.474053	2.745175	9.789552
3 (O)	115	0.124067	0.102693	0.008609	0.092785	0.046717	0.857878	0.811162	4.999092	34.87108
4 (O)	115	0.138313	0.12433	0.004191	0.064741	0.007832	0.446097	0.438265	2.094143	6.46854
1 (I)	115	11.63388	11.0828	27.19834	5.215203	1.73236	40.1079	38.37554	1.774117	7.468842
2 (I)	115	11.03114	11.1366	19.2098	4.382898	1.912889	20.52677	18.61388	0.021639	−0.74924
3 (I)	115	10.18784	9.7378	16.9729	4.119817	1.165667	21.40568	20.24001	0.399806	−0.014
4 (I)	115	9.310598	8.0431	131.7029	11.47619	2.241667	127.6853	125.4437	9.782052	101.5955

By applying this to the random distribution of sums of random samples of size n,

$$S = \sum_{i=1}^{n} x_i \tag{20}$$

then it follows that the sampling distribution of S has the following mean and variance:

$$\mu_S = n\mu, \quad \sigma_s^2 = n\sigma^2 \tag{21}$$

This is the outcome of the central limit theorem. Figures 3–6 reveal that productivity has a stable mean but an unstable (undefined) variance that is definitely not linear in scaling as in Eq. 21.

A comparison between the original data and the normal distribution illustrates similarity in the scaling of the mean. The variance, on the other hand, does not follow the theoretical values that were calculated using Eq. 21 (Figures 7 and 8). Furthermore, variance of normally distributed data converges to a certain value. Such a phenomenon does not occur in the original data, where the variance is growing rapidly and hence clearly showing unstable properties.

We have arrived at the significant assumption that labour productivity is not normally distributed. The above analytical and visual methods disclose that normal distribution cannot be assumed when describing the properties of construction productivity data for three reasons.

- Unstable variance which is growing much faster than in the case of a normal distribution (central limit theorem is not applicable).
- Nonzero skewness (distribution is not symmetrical).
- Leptokurtotic phenomenon (higher peaks than in the case of normal distribution).

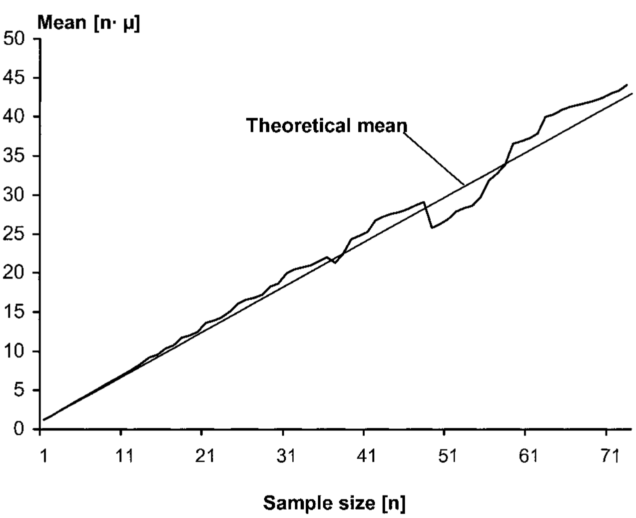

Figure 3 A scaling property for the mean for data set 1 (formwork). Equivalent results were obtained for all data sets

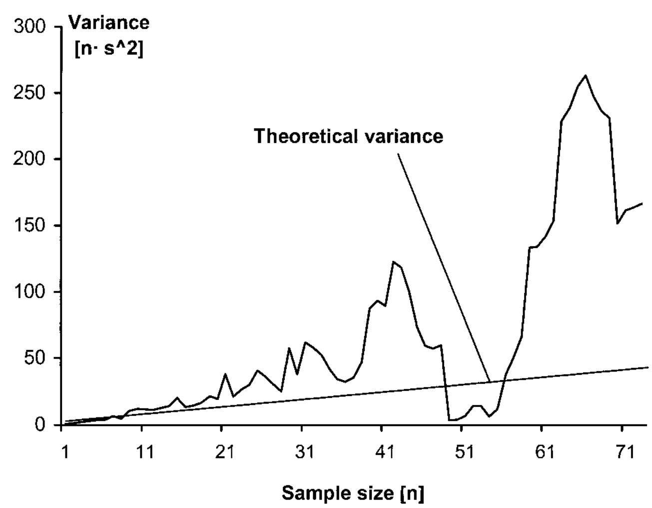

Figure 4 A scaling property of the variance for data set 1 (formwork): clearly this is unstable and not exhibiting linearity in scaling. Equivalent results were obtained for all data sets

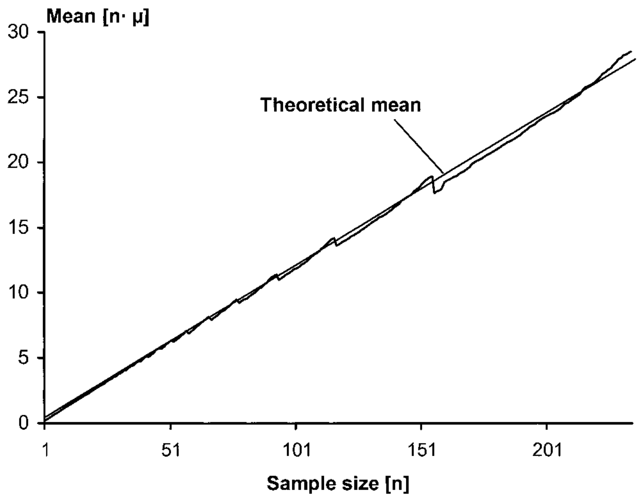

Figure 5 As for Figure 3 but for data set 12 (masonry). Equivalent results were obtained for all data sets

The above indications lead us to reject 'normal distribution' to describe the behaviour of labour productivity. However, before stating the conclusion, we need to perform the tests of normality in order to prove the above assumptions completely.

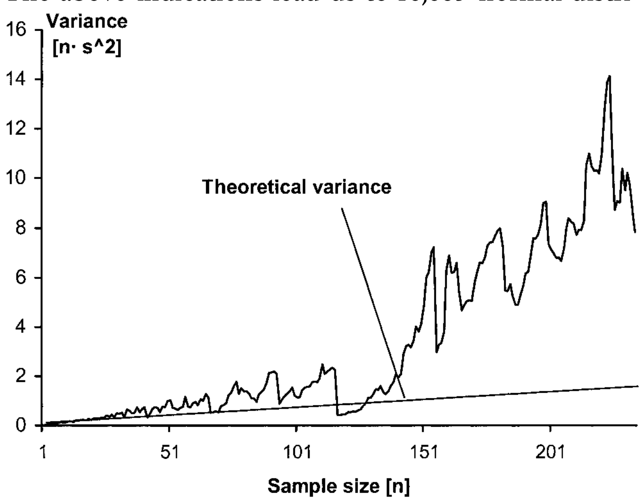

Figure 6 As for Figure 4 but for data set 12 (masonry). Equivalent results were obtained for all data sets.

Tests of normality

There are several tests for normality available in the literature on statistics. Nevertheless, one has to be very careful when applying certain tests for the results to be of any use. First, we have to be certain that sufficiently large data sets are available for testing. Tests differ in their ability to deal with large or small sample sizes (Conover, 1980; Sprent, 1989; Kottegoda and Rosso, 1997). For the purposes of this study, two common goodness-of-fit tests have been chosen: the Kolmogorov–Smirnov test, and the Anderson–Darling test

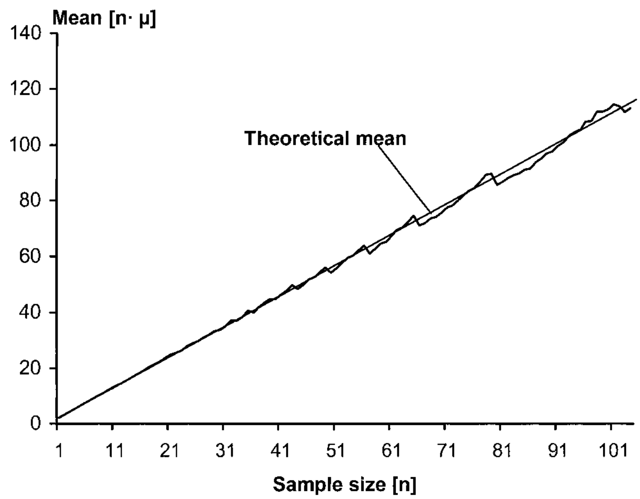

Figure 7 Linear scaling of the mean for normally distributed random numbers similar to the original data and follows the theoretical line

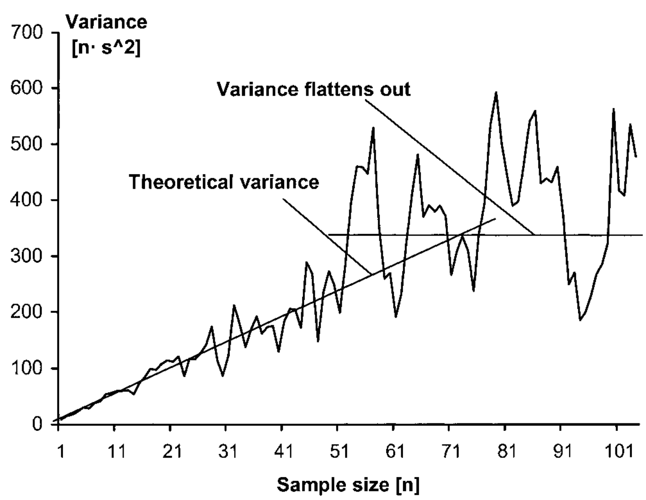

Figure 8 Variance of normal data follows the theoretical scaling (see text), does not grow rapidly and flattens out

The Kolmogorov–Smirnov goodness-of-fit test is a non-parametric test related to the cumulative density function. The limitation of the basic test is that the hypothesized distribution function must be completely specified, i.e., when there are no unknown parameters that should be estimated from the sample (Conover, 1980). However, later modifications now allow testing the composite hypothesis of normality. The null hypothesis states that a particular population belongs to the normal distribution without specifying the mean or variance of a distribution. The test statistic is the maximum vertical distance between empirical $F_n(x)$ and theoretical $F_0(x)$ cumulative density function (Kottegoda and Rosso, 1997).

$$D_n = \sup_x |F_n(x) - F_0(x) \tag{22}$$

The critical values for large samples ($n > 35$) are $1.3581/\sqrt{n}$ and $1.6276/\sqrt{n}$ for probabilities of 0.95 and 0.99, respectively (levels of significance $\alpha = 0.05$ and 0.01).

The second test applied in this study is the Anderson–Darling goodness-of-fit test, which gives heavier weighting to the tails of a distribution (Kottegoda and Rosso, 1997). Therefore it is appropriate for testing the normality of the productivity data due to the fat-tail phenomenon. The test statistic is

$$A^2 = -n - \sum_{i=1}^{n} \frac{(2i-1)\,\{\ln F_0(x_{(i)}) + \ln[1 - F_0(x_{(n-i+1)})]\}}{n} \tag{23}$$

where $x_{(i)}$ are the observations presented in increasing order. For large samples the 5% and 1% values of the statistic are $A^2_{0.05} = 2.492$ and $A^2_{0.01} = 3.857$, respectively. These estimations are good for $n>10$. The summary of the two tests performed on the 12 productivity data sets is presented in Table 4. The results reveal that the both tests are in agreement with the analyses of central moments of tendency, showing that normal distribution is not appropriate as a statistical description of labour productivity.

Discussion of results

The above analyses and tests show clearly that construction labour productivity is not normally distributed. All methods consistently illustrate that departure from normality is present in all data sets, regardless of the site or type of work, in the following manner.

1. The failure of the central limit theorem due to the unstable and immoderate growth of variance.
2. The leptokurtotic condition indicates that more data points are located around the mean and in the tails than predicted by the normal distribution.
3. Both tests show that productivity is not normally distributed (p is infinitesimal in all cases).

Similar results of left skewness and non-normal distribution of productivity have also been detected by Lema and Price (1996); however, they did not investigate the stability of variance while proposing the Johnson S_B distribution. Nevertheless, unstable variance indicates that Johnson S_B distribution is not applicable because it is based on a standard normal variable. Behaviour of that kind in essence resembles the Pareto-Lévy distribution which characteristic function can be presented (Mendelbrot, 1960) in the following form (Pareto-Lévy law):

$$\varphi(\xi) = \exp\left[iM\xi - |\xi|^{\alpha}(u^{\star})^{\alpha} \left\{1 - \left(\frac{i\beta\xi}{|\xi|}\right)\tan\left(\frac{\alpha\pi}{2}\right)\right\} \left|\cos\left(\frac{\alpha\pi}{2}\right)\right| \right] \tag{24}$$

Where α is a measure of 'peakedness' ($0<\alpha<2$), β is a measure of 'skewness' ($-1<\beta< +1$), $u^{\star}$ is a measure of standard deviation, and M is a measure of location (the mean).

A special case of Pareto-Lévy distributions is the normal distribution. Stable laws have no moments of any order if $0<\alpha\leq 1$, they have only one finite moment if $1<\alpha<2$, and they have a finite variance in the Gaussian case (normal distribution), $\alpha = 2$ (Mandelbrot, 1963). Stability of variance is of vital importance in statistical data analyses because most of the standard statistical tools are based on the assumption of stable and finite variance (Mandelbrot, 1960; Peters, 1991). Therefore, normal distribution is only a special case of Pareto-Lévy distributions, which are considered also as fractal because they are statistically self-similar with respect to time (Peters, 1991). Hence, the mean of summed 5-day productivity would be five times the mean of daily productivity (as shown in

Table 4 Summary of the two goodness-of-fit tests, showing strong departure from normality in all cases

Data set No.		1 (F)	2 (F)	3 (F)	4 (F)	5 (F)	6 (M)	7 (M)	8 (M)	9 (M)	10 (M)	11 (M)	12 (M)
Kolmogorov–	n	148	263	183	86	109	95	110	112	91	132	108	470
Smirnov test	D	0.221	0.462	0.221	0.236	0.330	0.194	0.485	0.156	0.235	0.317	0.187	0.176
	$D_{0.05}$	0.112	0.084	0.100	0.146	0.130	0.139	0.132	0.128	0.142	0.118	0.131	0.063
	$D_{0.01}$	0.134	0.100	0.120	0.176	0.156	0.167	0.155	0.154	0.171	0.142	0.157	0.075
	p	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Anderson–	A^2	14.51	83.63	17.00	8.85	21.67	7.37	37.94	4.78	8.06	24.29	7.72	33.55
Darling test	$A^2_{0.05}$	2.492	2.492	2.492	2.492	2.492	2.492	2.492	2.492	2.492	2.492	2.492	2.492
	$A^2_{0.01}$	3.857	3.857	3.857	3.857	3.857	3.857	3.857	3.857	3.857	3.857	3.857	3.857
	p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figures 3, 5 and 7); however, variance would scale only if $\alpha = 2$. As presented earlier, the variance of productivity does not follow scaling and therefore is undefined.

Chaos theory?

The above analyses reveal that productivity is characterized by undefined variance which is actually common to chaotic systems (Peters, 1991). Can we then say that productivity is chaotic? It might well be, because we are dealing with construction projects of a very complex and seemingly uncontrollable nature, and of which the outcome is fluctuating productivity. Similar results have been found in some other scientific disciplines, as reported by authors like Sharpe (1970), Peters (1991), Kaye (1993) and Lorenz (1995). All these systems indicate a complex nature that cannot be explained using standard statistical techniques. Peters (1991) revealed the curious behaviour of some economic indexes, which exhibit strong departure from normality, and his findings were based on unstable variance, skewness and the presence of extreme values. In particular, Peters studied the complex behaviour of the 5-day logarithmic first-differences in price for the S&P 500 index. The high-peak (leptokurtosis) and fat-tail (negative skewness) phenomena found were like those in productivity. In addition, the nonlinear dynamics applied revealed that behaviour of the S&P 500 index seems chaotic. Similar results within econometric studies were detected also by Dechert and Gencay (1993) and by Papaioannou and Karytinos (1995). The typical example of a chaotic signal is presented in Figure 9. The 'bumping' behaviour is similar to productivity. Ruelle (1989) reported that all correlation tests would reveal rapid de-correlation between successive iterations, hence assigning it randomness. The signal is actually an outcome of the simple deterministic difference equation (so-called logistic or May's equation):

$$x_{n+1} = 4\, x_n(1 - x_n) \tag{25}$$

The above equation was first used in the early 1970s for describing population variations when the growth produced population stress, which modified the actual growth curve. Population biologists have developed a graphic method of following the fluctuations predicted by May's equation, described as the 'parabolic cobweb' diagram (Kaye, 1993). Surprisingly, values of k greater than 3.58 generate chaotic behaviour (we have used $k = 4$; Eq. 25). The properties of this simple equation have been applied successfully in population biology and studied by many mathematicians (Kaye, 1993).

Although the objective of the present study is not the application of chaos theory or the above equation, these very briefly interpreted examples from other research disciplines may serve as a guide for future research. However, more profound definitions and descriptions of chaotic behaviour and analyses can be found elsewhere (Ruelle, 1989; Schuster, 1989; Baker and Golub, 1990; Ott, 1991).

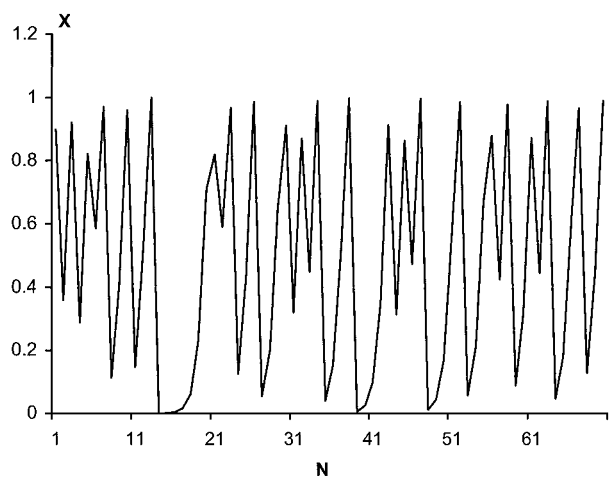

Figure 9 Simple deterministic difference equation exhibiting chaotic behaviour ($k = 4$) and to which all correlation tests would assign randomness

Conclusions

Construction labour productivity is not normally distributed, because of the following factors that were discovered by determining the central moments of tendency and applying two different tests of normality.

1. The failure of the central limit theorem due to an unstable and immoderate growth of variance (undefined variance).
2. The leptokurtotic condition, which indicates that the occurrence of values around the mean and extreme values of several standard deviations from the mean is much more frequent than predicted by the normal distribution.
3. The tests of normality show that productivity is not normally distributed (p is infinitesimal in all cases).

However, some standard characteristics (stable and finite variance, using test statistics that rely on normality, etc.) usually have been taken for granted, and consequently not much could have been done to achieve a better understanding of the ubiquitous complexity. Clearly, characteristics like large positive skewness, leptokurtosis, frequent extreme values and undefined variance demonstrate non-randomness, which means that statistical diagnostics like correlation coefficients and t

statistics and some basic paradigms like the central limit theorem are not applicable. However, based on evidence from other scientific disciplines like econometrics and population biology, there is the promising possibility of applying chaos theory in the further examination of productivity data. Chaotic systems are described by Pareto-Lévy distributions whose characteristic shape is a fat-tailed, high-peak curve, like that for productivity. This may lead to a new paradigm of chaotic productivity in view of some deterministic properties hidden within random-looking productivity, and hopefully future research will focused on the chaotic characteristics of construction labour productivity.

Acknowledgements

I acknowledge a great debt of thanks to Professor Randolph H. Thomas and my supervisor Professor R. Malcolm W. Horner for providing the data for this study.

References

Arditi, D. (1985) Construction productivity improvement. *Journal of Construction Engineering and Management ASCE*, **111**(1), 1–14.

Baker, G. L. and Gollub, J. P. (1990) *Chaotic Dynamics: An Introduction*, Cambridge University Press.

Briscoe, G. (1988) *The Economics of the Construction Industry*, Batsford, London.

Choromokos Jr., J. and McKee, K. E. (1981) Construction productivity improvement. *Journal of the Construction Division ASCE*, **107**(1), 37–47.

Conover, W. J. (1980) *Practical Nonparametric Statistics*, 2nd Edn, Wiley, New York.

Dechert, W. D. and Gencay, R. (1993) Non-linear dynamics chaos and econometrics: using the correlation exponent to decide whether an economic series is chaotic, edited by M. H. Pesaran and S. M. Potter, Wiley, New York.

Gleick, J. (1993) *Chaos*, Abacus, New York.

Hahn, G. J. and Shapiro, S. S. (1994) *Statistical Models in Engineering*, Wiley, New York.

Horner, R. M. W. and Talhouni, B. T. K. (1995) *Effects of Accelerated Working, Delays and Disruption on Labour Productivity*, Chartered Institute of Building, Ascot.

Kaming, F. P., Olomolaiye, P. O., Holt, G. D., Kometa, S. T. and Harris, F. C. (1996) Project managers' perception of production problems: an Indonesian case study. *Building Research and Information*, **24**(5), 302–9.

Kaye, B. H. (1993) *Chaos and Complexity: Discovering the Surprising Patterns of Science and Technology*, VCH, Weinheim.

Kottegoda, T. N. and Rosso, R. (1997) *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York.

Lema, N. M. and Price, A. D. F. (1996) Construction process performance variability: focus on labour productivity. *Building Research and Information*, **24**(6), 339–49.

Lorenz, E. N. (1995) *The Essence of Chaos*, UCL Press.

Mandelbrot, B. (1960) The Pareto-Lévy law and the distribution of income. *International Economic Review*, **1**(2), 79–106.

Mandelbrot, B. (1963) The stable Paretian Income Distribution when the Apparent Exponent is near Two, *International Economic Review*, **4**(1).

Oglesby, C., Parker, H. and Howell, G. (1989) *Productivity Improvement in Construction*. McGraw-Hill, New York.

Ott, E. (1991) *Chaos in Dynamical Systems*, Cambridge University Press.

Papaioannou, G. and Karytinos, A. (1995) Nonlinear time series analysis of the stock market exchange: the case of an emerging market. *International Journal of Bifurcation and Chaos*, **5**(6), 1557–84.

Peters, E. E. (1991) Chaos and Order in the Capital Markets: A New View of Cycles, Prices and Market Volatility, Wiley, New York.

Ruelle, D. (1989) *Chaotic Evolution and Strange Attractors: The Statistical Analysis of Time Series for Deterministic Nonlinear Systems*, Cambridge University Press.

Schuster, H. G. (1989) *Deterministic Chaos: An Introduction*, 2nd Rev. Edn, VCH, Weinheim.

Sharpe, W. F. (1970) *Portfolio Theory and Capital Markets*, McGraw-Hill, New York.

Sprent, P. (1989) *Applied Nonparametric Statistical Methods*, 2nd Edn, Chapman & Hall, London.

Sprent, P. (1998) *Data Driven Statistical Methods*, Chapman & Hall, London.

Tavakoli, A. (1985) Productivity analysis of construction operations. *Journals of Construction Engineering and Management ASCE*, **111**(1), 31–9.

Thomas, H. R. and Yiakoumis, I. (1987) Factor model of construction productivity. *Journal or Construction Engineering and Management ASCE*, **113**(4), 623–39.

Thomas, H. R., Moloney, W. F., Horner, R. M. W., Smith, G. R., Handa, V. K. and Sanders, S. R. (1990) Modelling construction labour productivity. *Journal of Construction Engineering and Management ASCE*, **116**(4), 705–26.