A statistical assessment of the equivalence between creep and hot tensile behaviour of 2.25Cr-1Mo steel

M. Evans[1]

Abstract

It is important to be able to predict the creep life of materials used in power plants. Levi De Oliveira Bueno has suggested an equivalence between creep and high temperature tensile testing for 2.25Cr-1Mo steel. This offers the potential for reducing the development cycle for new materials designed to operate at ever higher temperatures. This paper reviews the literature to identify some suitable statistical tests for this equivalence. When applied to 2.25Cr-1Mo steel, it was found that this equivalence was only valid if the Monkman-Grant exponent equalled -1. However, this constraint was not accepted by the data based on the proposed statistical tests, and when this constraint was relaxed the equivalence between these different types of test data disappeared.

Introduction

Although complex stresses and temperatures are often encountered by materials used in power generation, design decisions are generally based on an "allowable" tensile creep strength. This strength is commonly taken to be 67% of the average stress (up to 1088 K).[1] Currently, expensive testing lasting 12–15 years is required to determine the long-term strengths and lives. A reduction in this 'materials development cycle' was therefore defined as the No.1 priority in the 2007 UK Energy Materials – Strategic Research.[2] With the aim of reducing this development cycle, a number of different approaches have been investigated. First, a new group of parametric creep models[3,4] have been developed in recent years that are characterised through their use of an S shaped curve to describe the relationship between the minimum creep rate and a normalised stress at a fixed temperature. These models have been shown to be stable with respect to time (test duration) and so have been reasonably successful[5–8] in predicting long term creep life from short term test data. Secondly, for materials where the Monkman- Grant[9] relation is stable over all test durations, a lifetime prediction can be made using measured minimum creep rates. As the minimum creep rate is reached well before rupture, this approach also offers the potential to reduce the length and cost of testing programs.

A hypothesis first put forward by Levi De Oliveira Bueno[10] in 2005 also offers the potential to reduce the time and costs of testing programs required for material development. This hypothesis was based around the similarities between a uniaxial creep test and a high temperature tensile test. A creep test involves subjecting a test specimen to a constant stress σ_c (or load) and high temperature (T) until it fails. During this time the specimen will strain at varying rates and the time at which the specimen fails t_{cf} and the minimum strain rate $\dot{\varepsilon}_{mc}$ are measured. A tensile test involves subjecting a test specimen to constant high temperature and a constant strain $\dot{\varepsilon}_{ts}$ until it fails. During this time the specimen will be subjected to a varying stress σ_{ts} (to ensure a constant strain rate) and the time at which the specimen fails t_{ts} and the maximum stress σ_{ts} (i.e., the ultimate tensile strength) are measured. The hypothesis made by Levi De Oliveira Bueno based on their results obtained on 2.25Cr-1Mo steel, was that i. σ_c was equivalent to σ_{ts}, ii. that t_{cf} was equivalent to t_{ts} and that iii. $\dot{\varepsilon}_{mc}$ was equivalent to $\dot{\varepsilon}_{ts}$. In fact, Osgerby and Dyson[11] demonstrated in an earlier paper, that the constant strain rate properties of a material are inextricably linked to its creep properties. Levi De Oliveira Bueno concluded that the Monkman-Grant constant was the same in the creep and tensile data sets which they took as evidence supporting the above hypothesis.

[1]Institute of Structural Materials, Swansea, Wales

Corresponding author:
M. Evans, Institute of Structural Materials, Swansea University Bay Campus, SA1 8EN, Swansea, Wales.
Email: m.evans@swansea.ac.uk

If this hypothesis is true, then a prediction of creep life at a typical operating condition can be quickly obtained. It involves conducting a number of tensile tests at different stain rates to estimate the MG relation in the tensile data set. These will all be cheap and very short-term tests. Then conduct just one creep test at the operating stress and temperature but discontinue it once the minimum creep rate has been observed. This will also be a short-term test in relation to how much time would be required for this specimen to fail. Finally, insert this minimum creep rate into the tensile MG relation for the creep strain to predict the creep failure time. This test program is quicker and cheaper than carrying out many creep tests at accelerated test conditions lasting several months.

But this identified similarity assumed that the Monkman-Grant (MG) exponent was equal to -1 (but as this paper demonstrates this value is larger than this within the creep data set used by Levi De Oliveira Bueno). That said, the Monkman-Grant relation provides an ideal framework within which to study the above hypothesis in more detail. This relation was first identified using data on the time to failure t_{cf} and the minimum creep rate $\dot{\varepsilon}_{mc}$ both measured from uniaxial creep tests

$$t_{cf} = M_c(\dot{\varepsilon}_{mc})^{-1} \quad (1a)$$

where M_c is material dependent. M_c can therefore be interpreted as the total strain experienced by a specimen if it crept at a strain rate of $\dot{\varepsilon}_{mc}$. However, for many materials – especially high chrome steels[12] and Nickel based super alloys[13] – it has been found that the following relation is more suitable

$$t_{cf} = M_c(\dot{\varepsilon}_{mc})^{-\rho_c} \quad (1b)$$

where $0 < p_c < 1$. This deviation (i.e., $p_c < 1$) from the simple MG relation of Eq. (1a) has been explained in a variety of different ways. Dobes and Milicka[14] attributed this deviation to variation in creep failure strain ϵ_{cf}

$$\frac{t_{cf}}{\varepsilon_{cf}} = M_c(\dot{\varepsilon}_{mc})^{-1} \quad (1c)$$

Sklenicka et al.[15] found that for Grade 92 steel, the value of ρ_c in Eq. (1b) was 0.88, but when t_{cf} is replaced with $\frac{t_{cf}}{\varepsilon_{cf}}$ in Eq. (1b) the value for ρ_c increased to 0.96. In contrast to this, Abe[16] attributed this deviation to accelerating creep strain rates during tertiary creep

$$\frac{d\ln(\dot{\varepsilon}_c)}{d\varepsilon_c}t_{cf} = \left(\frac{t_{cf}}{(t_{cf} - t_m)}\right)(\dot{\varepsilon}_{mc})^{-1} \quad (1d)$$

where t_m is the time taken to reach the minimum creep rate, ϵ_c is the creep strain and $\dot{\varepsilon}_c$ the creep strain rate. However, when applied to 9Cr-1 W steel, Abe obtained a value for ρ_c that was less than 1.

The Levi De Oliveira Bueno hypothesis can be tested within the MG framework, by assuming the MG relation of Eq. (1b) applies to the tensile test results as well

$$t_{ts} = M_{ts}\dot{\varepsilon}_{ts}^{-\rho_{ts}} \quad (1e)$$

where t_{ts} is the time it takes for the varying stress in a tensile test to reach its maximum (i.e., the tensile stress) and $\dot{\varepsilon}_{ts}$ is the constant strain under which this tensile test was carried out. If parts ii and iii of the Levi De Oliveira Bueno hypothesis are true, then this will only be so if $M_c = M_{ts}$ and if $\rho_c = \rho_{ts}$. When this is the case, data pairings of $(t_{ts}, \dot{\varepsilon}_{ts})$ and $(t_c, \dot{\varepsilon}_c)$ will all fall on the same curve implying that t_{cf} and t_{ts} are equivalent and that $\dot{\varepsilon}_{mc}$ and $\dot{\varepsilon}_{ts}$ are equivalent. The hypothesis made by Levi De Oliveira Bueno can therefore be expressed in terms of parameter restrictions within the MG relation. This enables some standard statistical tests to be used.

The Levi De Oliveira Bueno hypothesis is an interesting hypothesis, because if true, it would enable a reliable creep life prediction at design conditions to be made from very short term (minutes) high temperature tensile tests. This would remove the need for even accelerated creep testing. The aim of this paper it to apply statistical test to this hypothesis within the MG framework using the data of Levi De Oliveira Bueno[10] on 2.25Cr- 1Mo steel. To achieve this aim, the paper is structured as follows. The following section describes this data set in a bit more detail. This is followed by a section outlining some statistical tests for this hypothesis (with much of the statistical theory confined to an appendix). The penultimate section presents the results from applying these statistical tests and finally a conclusion section outlines some proposed areas for future work.

The data

High temperature tensile and creep properties for 2.25Cr-1Mo steel have been published by Bueno & Sobrinho.[10] The steel they used for testing was supplied in plate form with 25.4 mm thickness according to ASTM A 387, Grade 22 and in the normalized and tempered condition with the following chemical composition: Fe – 2.09Cr – 1.08Mo – 0.097C – 0.32Si – 0.50Mn – 0.007P – 0.002S – 0.03Ni – 0.01Cu – 0.05Al. Metallographic analysis carried out by these researchers indicated the presence of 30% bainite and 70% ferrite. Test specimens were extracted from the rolling direction and had a gauge length of 25 mm and an initial diameter of 6.25 mm. High temperature tensile testing was carried out on a servo-hydraulic 8802 model INSTRON machine using five different temperatures ranging from 873 K to 973 K and at five different constant strain rates ranging from 6.67×10^{-6} s^{-1} up to 0.01 s^{-1}. Figure 1 plots the constant strain rates against the time taken to reach the ultimate tensile strength in all these high temperature tensile tests (open squares).

The creep tests carried out by Bueno & Sobrinho were at constant load, according to ASTM E139,[17] using a set of 10 creep machines model STM-MF 1000. (Further information about this equipment and testing techniques appears in.[18] The elongation of the specimens was followed with creep extensometers having Transtek LVDT transducers - model DCDT 0243-000. The readings from the transducers were collected by a Fluke Data Logger, model Hydra 2635A series II, using scan rates varying from 6 readings/min to 6 readings/h. The creep tests were carried out at
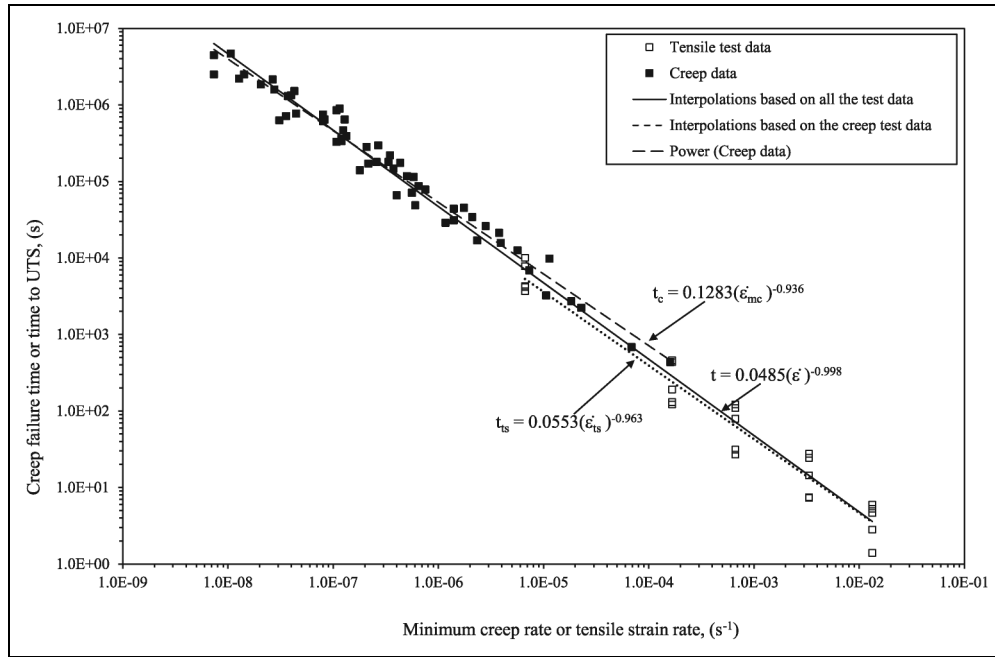
Figure 1. Variation of the minimum creep rate with the rupture time in creep test data, plotted together with the strain rate and time to occurrence of the ultimate tensile stress (UTS) in the constant strain rate tensile tests. Straight lines refer to the least squares estimates of the MG relation in the creep (long dashed), tensile (short dashed) and combined (sold line) test data sets.

nine temperatures levels, ranging from 773 K to 973 K, covering nineteen levels of applied load, varying from 34 MPa to 448 MPa. The resulting minimum creep rates and rupture lives are also shown in Figure 1 with failure times ranging from 0.12 to about 1300 h (closed squares). It can be seen that there seems to be more variability in the high temperature tensile tests data set.

Methodology

In Figure 1, the MG relation is applied separately to the creep test data and the tensile test datasets, together with the fit to the pooled data set. A difference is observed in the values for the MG constant and exponents in each data set, but what is required is a test that assesses whether these observed differences in the parameters of the MG relation are statistically significant, i.e., that these observed differences have not occurred by chance.

Testing single parameter restrictions when the Monkman-Grant exponent is 1

If the Monkman-Grant exponent term is unity, it can be expressed as follows for both the creep and tensile test results

$$t_{cf}(\dot{\varepsilon}_{mc}) = M_c \quad (2a)$$

$$t_{ts}(\dot{\varepsilon}_{ts}) = M_{ts} \quad (2b)$$

If parts ii and iii of the Levi De Oliveira Bueno hypotheses are true, then this will only be so if $M_c = M_{ts}$. So, one approach to testing this parameter restriction is to test

whether the mean values for M_c and M_{ts} are the same. Existing tests for differences in population means are heavily reliant on use of the normal distribution or the central limit theorem, yet it is well established that creep failure times and minimum creep rates have long tailed distributions. Taking natural logs can transform such variables into normal variates – for example if t_{cf} has a long-tailed log normal distribution, $\ln(t_{cf})$ will be normally distributed. There is therefore good reason for expressing the above MG equations in natural logs

$$\ln(t_{cf}) + \ln(\dot{\varepsilon}_{mc}) = \ln(M_c) \quad (3a)$$

$$\ln(t_{ts}) + \ln(\dot{\varepsilon}_{ts}) = \ln(M_{ts}) \quad (3b)$$

Let μ_c be the population or true mean for $\ln(M_c)$ and μ_{ts} be the population or true mean for $\ln(M_{ts})$. A test of the null hypothesis H_o: $\mu_c - \mu_{ts} = 0$ is then one way to test the Levi De Oliveira Bueno hypothesis. To conduct such a test, two samples of data are required. The first is the tensile test data set to be made up of $i = 1$ to n_1 measurements on $\ln(t_{ts})$ and $\ln(\dot{\varepsilon}_{ts})$ and the second is a creep test data set to be made up of $i = n_1 + 1$ to n measurements on $\ln(t_{cf})$ and $\ln(\dot{\varepsilon}_{mc})$ (such that $n > n_1$). Then let the variables $y_i = \ln(t_{cf})_i$ and $x_{1i} = \ln(\dot{\varepsilon}_{ts})_i$ when $n \leq i > n_1$, but $y_i = \ln(t_{ts})$ and $x_{1i} = \ln(\dot{\varepsilon}_{ts})_i$ when $i \leq n_1$. Eqs. (3) can then be written as

$$w_{ts,i} = y_i + x_i = \ln(M_{ts})_i \qquad i = 1 \text{ to } n_1 \quad (4a)$$

$$w_{c,i} = y_i + x_i = \ln(M_c)_i \qquad i = 1 + n_1 \text{ to } n \quad (4b)$$

Section A of the appendix shows that if the variance in each test data set are the same, $\sigma_c^2 = \sigma_{ts}^2 = \sigma_w^2$, and if $w_{ts,i}$ and $w_{c,i}$ are both normally distributed (if not then the central limit

theorem will ensure $\bar{w}_c$, $\bar{w}_{ts}$ will still be normally if calculated from a large sample) then the following t statistic can be used to test the null hypothesis H_o: $\mu_c - \mu_{ts} = 0$

$$t^a = \frac{[\bar{w}_c - \bar{w}_{ts}] - [\mu_c - \mu_{ts}]}{\sqrt{\frac{S_w^2}{n_1} + \frac{S_w^2}{(n - n_1)}}} = \frac{[\bar{w}_c - \bar{w}_{ts}] - 0}{\sqrt{\frac{S_w^2}{n_1} + \frac{S_w^2}{(n - n_1)}}} \quad (5)$$

because if the null is true, t^a has a student t distribution with n – 2 degrees of freedom. $\bar{w}_c$, $\bar{w}_{ts}$ are the sample means for w_c and w_{ts} respectively, and s_w^2 is the sample estimate of the common variance σ_w^2 (Appendix A gives the formulas for $\bar{w}_c$, $\bar{w}_{ts}$ and s_w^2).

Values for the probability of observing various values for t^a have been tabulated. For example, there is a 5% chance of observing a t^a value outside the range −2.23 to +2.23 when the degrees of freedom equals 10. If a calculated value for $|t^a|$ from Eq. (5) then exceeds 2.23 it can be said that the chances of getting such a value is less than 5%. As this $|t^a|$ value was calculated under the assumption that that $\mu_c - \mu_{ts} = 0$, it follows that there is also less than a 5% chance that the true value for $\mu_c - \mu_{ts} = 0$. The decision rule is to reject the null hypothesis whenever $|t^a| > |t_{\alpha/2,n-2}|$ as this runs only a α% chance of being wrong in that rejection decision. $|t_{\alpha/2,n-2}|$ is called the critical value for t. Alternatively, the p-value is the probability of observing $|t^a|$ or more, and so is equivalent to the probability of the null hypothesis being true. The null hypothesis is then rejected when p-value $< \alpha$%.

The normality assumption required to hold for the above test is almost guaranteed by the central limit theorem that states that the distribution for a sample mean tends to a normal distribution as the sample size from which it's calculated tends to infinity and from the fact that linear combinations of normal variables $[\bar{w}_c - \bar{w}_{ts}]$ are also normal. The approximation to normality is excellent in samples of size 30 or more. The assumption that $\sigma_{wc}^2 = \sigma_{wts}^2 = \sigma_w^2$ can be tested using the ratio of the sample estimates of these variances

$$F^V = \frac{S_{w_{ts}}^2}{S_{w_c}^2} \quad (6)$$

which under the null hypothesis that $\sigma_c^2 = \sigma_{ts}^2$ has an F distribution with n - n_1 - 1 and n_1 −1 degrees of freedom (as the variances are chi squared distributed and the ratio of two chi square variables has an F distribution). Consequently, if the value for F^V in Eq. (6) exceeds F_α, then the observed value for F^V has less than a α% chance of being observed. If such a value is then actually observed, this can only then be explained by F^V not having an F distribution – which will only be so if the null hypothesis is not true. Here F_α is the value for F^V such that there is only a probability of α that this value or more will be observed. So, the null hypothesis is rejected only if $F^V > F_\alpha$, i.e., when the chances of the null hypothesis being true have dropped so low that it cannot be entertained as a realistic possibility.

Another reason for testing differences between log means, is that taking logs linearises non-linear equations like the MG relation. This then enables the technique of linear least squares to be used to test the Levi De Oliveira Bueno when the MG exponent is not unity, i.e., when the above means difference test is no longer suitable. If the variable $I_i = 0$ when $i \leq n_1$ but $I_i = 1$ when $i > n_1$, then the value for I_i distinguishes results from the creep and tensile tests. Given this, it is possible to combine Eqs. (4) into a single equation where $\beta_0 = \ln(M_{ts})$ and $\beta_0 + \beta_1 = \ln(M_c)$ in

$$y_i + x_i = \beta_o + \beta_1 I_i + v_i \quad \text{for} \quad i = 1 \text{ to } n \quad (7a)$$

Consequently, testing the Levi De Oliveira Bueno hypothesis boils down to testing the hypothesis that $\beta_1 = 0$. The variable v_i in Eq. (7a) represents the scatter that is measured around the regression line and reflects the stochastic nature of creep and tensile testing. The least squares procedure is one that choses values for β_o and β_1 that minimises the Σv_i^2, where the sum is over all n observations. The solution to this optimisation problem is

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n_1} (y_i + x_i)}{n_1} = \bar{w}_{ts} \text{ and}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1+n_1}^{n} (y_i + x_i)}{n - n_1} - \hat{\beta}_0 = \bar{w}_c - \bar{w}_{ts} \quad (7b)$$

where the hat denotes that these are estimates of the true population values β_o and β_1. If the v_i are assumed to be normally distributed and if the variance of the residuals v_i is a constant (i.e., the same value in the creep and tensile data sets) equal to σ_v^2, then

$$t^b = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s_v^2 \left[\frac{1}{n_1} + \frac{1}{(n - n_1)}\right]}} \qquad j = 1 \text{ to } 2 \quad (8a)$$

has a student t distribution with n - 2 degrees of freedom. s_v^2 is the following unbiased estimate of $\sigma^2 v$

$$s_v^2 = \frac{\sum_{i=1}^{n} \hat{v}_i^2}{(n - 2)} \quad (8b)$$

where the $\hat{v}_i^2$ are calculated from Eqs. (7a) using the parameter estimates given by Eq. (7b). Any hypothesised value for β_j can be tested simply by substituting into Eq. (8a) the hypothesised value for β_j. For example, part of the Levi De Oliveira Bueno hypothesis can be tested by substituting $\beta_1 = 0$ into Eq. (8a) for the null hypothesis H_o: $\beta_1 = 0$

$$t^b = \frac{\hat{\beta}_1 - 0}{\sqrt{s_v^2 \left[\frac{1}{n_1} + \frac{1}{(n - n_1)}\right]}} \quad (9)$$

Interestingly, and as discussed in Appendix B, from a statistical perspective, this t test is equivalent to the above t-test for testing for a difference between two means assuming constant variance. That is, t^a in Eq. (5) is equivalent to t^b in Eq. (8a).

Testing single parameter restrictions when the Monkman-Grant exponent is less than 1

When the MG exponent differs from -1, Eq. (7a) needs to be generalised to account for this

$$y_i = \beta_o + \beta_1 I_i + \beta_2 x_i + \beta_3 I_i x_i + e_i \text{ for } i = 1 \text{ to } n$$

(10a)

where $\beta_0 = \ln(M_{ts})$, $\beta_0 + \beta_1 = \ln(M_c)$, $\beta_2 = \rho_{ts}$ and $\beta_2 + \beta_3 = \rho_c$. Consequently, testing the Levi De Oliveira Bueno hypothesis boils down to testing the dual hypothesis that $\beta_1 = 0$ and $\beta_3 = 0$. The variable e_i in Eq. (10a) represents the scatter that is measured around the regression line and reflects the stochastic nature of creep and tensile testing. In matrix form, Eq. (10a) can be written as

$$\mathbf{y} = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{e}$$

where

$$\mathbf{y} = \begin{pmatrix} \ln(t_{ts,1}) \\ \vdots \\ \ln(t_{ts,n1}) \\ \ln(t_{c,n1+1}) \\ \vdots \\ \ln(t_{c,n}) \end{pmatrix}; \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}; \mathbf{x}$$

$$= \begin{pmatrix} 1 & 0 & \ln(\dot{\varepsilon}_{ts,1}) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \ln(\dot{\varepsilon}_{ts,n1}) & 0 \\ 1 & 1 & \ln(\dot{\varepsilon}_{c,n1+1}) & \ln(\dot{\varepsilon}_{c,n1+1}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \ln(\dot{\varepsilon}_{c,n}) & \ln(\dot{\varepsilon}_{c,n}) \end{pmatrix}; \mathbf{e}$$

$$= \begin{pmatrix} e_1 \\ \vdots \\ e_{n1} \\ e_{n1+1} \\ \vdots \\ e_n \end{pmatrix}$$

(10b)

The ordinary least squares procedure estimates the parameters β_0 to β_3 as those values which minimise Σe_i^2 ($i = 1$ to n). The solution to this least squares problem is

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} \quad (11a)$$

where $\hat{\boldsymbol{\beta}}$ is the least squares estimator of $\boldsymbol{\beta}$, $\mathbf{x}^T$ is the transpose of $\mathbf{x}$ and $()^{-1}$ is the inverse of the matrix in the brackets. These least squares estimators are called linear estimators as they are a linear combination of the y_i. They are sample estimates as collecting a different creep and tensile sample of the same sizes will result in different values for $\hat{\beta}_o$ to $\hat{\beta}_3$. Next assume that the values for x_i are fixed numbers (or alternatively are uncorrelated with the e_i), and that the mean of all the e_i values equals zero. Then, it can be shown that if all possible samples of size n_1 and $n-n_1$ of observations on y_i and x_i are taken, and for

each sample the estimates $\hat{\beta}_o$ to $\hat{\beta}_3$ are computed, then a large number of estimates for each will be obtained and the means (called the expected value E) of these values will equal the true (i.e., population) parameter values. For example, $E[\hat{\beta}_o] = \beta_o$ and $E[\hat{\beta}_3] = \beta_3$. So, the least squares estimates are said to be unbiased.

If the residuals e_i from both the creep and tensile data have the same constant variance equal to σ_e^2, and if the e_i are all independent of each other, then the covariance matrix for $\hat{\boldsymbol{\beta}}$ is given by

$$Var(\hat{\boldsymbol{\beta}}) = \sigma_e^2 (\mathbf{x}^T \mathbf{x})^{-1} \quad (11b)$$

and so $Var(\hat{\boldsymbol{\beta}})$ is a 4×4 square matrix. The diagonal of this matrix contains the variances of $\hat{\beta}_o$ to $\hat{\beta}_3$ (for example, $Var(\hat{\beta}_o)$ is in position (1,1) of the matrix and $Var(\hat{\beta}_3)$ is in position (4,4) of the matrix). The covariances between these parameter estimates are then in the of-diagonal positions. The variance of each least squares estimate - given by Eq. (11b) - also has the smallest variance amongst the class of all possible linear estimators. This in combination with their unbiased nature ensures that the chances of obtaining an estimate for each β that is different from their true values is minimised. It is this result that makes the least squares estimation procedure so popular.

If in addition to the above assumptions, the residuals e_i are assumed to be normally distributed, then the y_i and each $\hat{\beta}_j$ variate will also be normally distributed. Each $\hat{\beta}_j$ will then have a normal distribution with a mean value equal to their true value and with minimised variances

$$\hat{\beta}_j \sim N[\beta_j, Var(\hat{\beta}_j)] \quad (12)$$

Unfortunately, σ_e^2 is rarely known in Eq. (11b) and needs to be estimated from the sample of data. It can be shown that an unbiased estimate of this residual variance is

$$s_e^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{(n-4)} \quad (13)$$

where the $\hat{e}_i^2$ are calculated from Eqs. (10a) using the parameter estimates given by Eq. (11a). If in Eq. (11b), σ_e^2 is replaced with s_e^2, then

$$t^g = \frac{\hat{\beta}_j - \beta_j}{\sqrt{Var(\hat{\beta}_j)]}} \quad (14)$$

has a student t distribution with $n - 4$ degrees of freedom. Any hypothesised value for β_j can be tested by simply substituting into Eq. (14) the hypothesised value for β_j. For example, consider the null hypothesis that the MG exponent is the same in the creep and tensile data sets. This can now be tested by expressing the null hypothesis as $\beta_3 = 0$. Substituting this value into Eq. (14)

$$t^g = \frac{\hat{\beta}_3 - 0}{\sqrt{Var(\beta_3)}} \quad (15)$$

where $\hat{\beta}_3$ is located in the 4th row of $\hat{\boldsymbol{\beta}}$ in Eq. (11a) and $Var(\hat{\beta}_3)$ is located in position (4,4) of the matrix $Var(\hat{\boldsymbol{\beta}})$ in

Eq. (11b). If the $|t^g|$ value is less than $|t_{\alpha/2, n-4}|$ the null hypothesis is accepted. The Levi De Oliveira Bueno hypothesis, at least in part, can now be tested by expressing the null hypothesis as $H_o: \beta_3 = 0$. To sum up, the t test of a single parameter restriction requires the following assumptions to hold: i. The e_i are normally distributed, ii the e_i over all n data points (i.e., in the creep and tensile data sets) have the same constant variance (homoscedastic residuals) and iii. the covariance between x_{1i} and e_i is zero).

White[19] suggested the following test for homoscedastic residuals. First, estimate values for the squared residuals

$$\hat{e}_i^2 = \{y_i - [\widehat{\beta}_0 + \widehat{\beta}_1 I_i + \widehat{\beta}_2 x_i + \widehat{\beta}_3 \, I_i x_i]\}^2 \; i = 1 \text{ to } n \quad (16a)$$

where the $\widehat{\beta}_j$ and found from Eq. (11a). Then carry out the regression

$$\hat{e}_i^2 = \lambda_o + \lambda_1 I_i + \lambda_2 x_{1i} + \lambda_3 I_i x_i + e_i \text{ for } i = 1 \text{ to } n \quad (16b)$$

$\hat{e}_i^2$ is an estimate of the residual variance (as the residuals are assumed to have zero mean) when the log strain rate equals x_i. So constant residual variance requires $\hat{e}_i^2$ to be independent of x_{1i}, i.e., requires $\lambda_1 = \lambda_2 = \lambda_3 = 0$. This will only be the case if the R^2 value associated with Eq. (16b) is close to zero. So, if the null hypothesis of homoscedastic residuals is true, the statistic

$$\chi^w = nR^2 \quad (16c)$$

has a chi square distribution with n-1 degrees of freedom. Bowman and Shenton[20] and Jarque and Bera[21] proposed the following test for normally distributed variables. It is based on sample measures of Skewness and Kurtosis and is given by

$$\chi^N = \frac{n\left(\frac{m_3}{m_2^{3/2}}\right)^2}{6} + \frac{n\left(\frac{m_4}{m_2^2} - 3\right)^2}{24} \quad \text{where} \quad m_j = \frac{\sum_{i=1}^n [\hat{e}_i]^j}{n} \quad (17)$$

The first term measures the degree of skewness in the residual distribution which equals zero when the distribution is symmetric. The second term measures the flatness of the distribution that again equals zero when it corresponds to the degree of flatness associated with the normal distribution. Under the null hypothesis of normality, this test statistic should be close to zero and follows a chi distribution with 2 degrees of freedom. So, if the value for χ^N exceeds the selected critical value for χ, the null hypothesis of normality is rejected.

Unfortunately, this test has poor small sample properties and a version of this test with a sample size correction can be found in Doornick and Hansen.[22] This modification has a chi square distribution with 2 degrees of freedom under the null hypothesis - even in small samples.

Testing multiple parameter restrictions when the Monkman-Grant exponent is less than 1

The above t tests can only be used for testing a single parameter restriction, but the Levi De Oliveira Bueno hypothesis involves the two restrictions $\beta_1 = \beta_3 = 0$ when the MG exponent is not known to equal unity. The F test for such multiple parameter restrictions requires the same assumptions as with the above t tests, and is written as

$$F^R = \frac{[R_U^2 - R_R^2] / k}{(1 - R_U^2) / (n - 4)} \quad (18a)$$

where k equals the number of restrictions. R_U^2 is the coefficient of determination associated with the unrestricted regression given by Eq. (10a)

$$R_U^2 = 1 - \frac{\sum_{i=1}^n \hat{e}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18b)$$

R_R^2 is the coefficient of determination associated with the restricted regression (which is determined by the nature of the null hypothesis). For example, to test the Levi De Oliveira Bueno hypothesis, the null hypothesis is $\beta_1 = \beta_3 = 0$ and so k = 2 within Eq. (18a), and R_R^2 is the coefficient of determination associated with the regression under this null hypothesis, which is

$$y_i = \beta_o + \beta_2 x_i + g_i \quad (18c)$$

and so

$$R_R^2 = 1 - \frac{\sum_{i=1}^n \hat{g}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18d)$$

where $\hat{g}_i^2$ are calculated from Eqs. (18c) using the least squares estimates of the parameters β_0 and β_2. A well-known statistical theorem states that the square of a standard normal variate has a chi square distribution with 1 degree of freedom and the square of two standard normal variates has a chi square distribution with 2 degrees of freedom and so on. Further, the ratio of two chi square variables has an F distribution. So, if the residuals e_i and g_i are normally distributed, then given Eq. (18b), R_U^2 involves the sum of n squared normal variates and so irrespective of the null hypothesis, the variable in the denominator of F^R has a chi square distribution with n - 4 degrees of freedom (4 degrees of freedom are lost from n in estimating 4 parameters). The numerator of F^R also has a chi square distribution but only if the null hypothesis is true - with the degrees of freedom equal to the number of restrictions under the null hypothesis (if the null hypothesis is not true the numerator has a non-central chi square distribution as it will then involve the sum of the squared residuals whose mean is not zero).

If, and only if, the null hypothesis is true, F^R will have an F distribution with k and n – 4 degrees of freedom. Consequently, if the value for F^R in Eq. (18a) exceeds F_α, then the observed value for F^R has less than a α% chance of being observed. The fact that it is observed can only then be explained by F^R not having an F distribution –

which will only be so if the null hypothesis is not true. Here F_α is the value for F^R such that there is only a probability of α that this value or more will be observed. So, the null hypothesis is rejected only if $F^R > F_\alpha$, i.e., when the chances of the null hypothesis being true have dropped so low that it cannot be entertained as a realistic possibility.

At a more intuitive level, if the null hypothesis is true, allowing the values for β_0 and β_2 to be different values in the tensile data compared to the creep data, should not explain any more of the variation in all the observed failure times. That is, under the null hypothesis $R_U^2 - R_R^2 \cong 0$ and F^R in Eq. (18a) will be small in magnitude. The larger it is, the more likely it is therefore that the null hypothesis is false and when its value exceeds F_α, the chances of the null hypothesis being true is at least as as low as α%. Given this low probability, the null hypothesis would then be rejected.

These F tests are uniformly most powerful tests for parameter restrictions under the above three assumptions. The reason for requiring homoscedastic residuals is that if the residual variance is larger in the tensile data set, it will deflate the value for R_U^2 and so inflate the value for F^R even if the null hypothesis is true. These so-called heteroscedastic residuals will therefore tend to lead to researchers rejecting the null hypothesis even when it is true. The heteroscedasticity problem can be overcome by adjusting the standard errors of the parameter estimates in Eqs. (11b) to adjust for this heteroscedasticity. As discussed above, White[19] suggested the variance in e_i associated with each value of x_i can be estimated using the estimated squared residuals. If the estimated squared residuals calculated using Eq. (16b) are placed down the diagonal of square matrix $\mathbf{\Omega}$

$$\hat{\mathbf{\Omega}} = \begin{pmatrix} \hat{e}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{e}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \cdots & \hat{e}_n^2 \end{pmatrix} \quad (19a)$$

then White showed that the standard errors for the parameters $\hat{\beta}_j$ in Eq. (11a), corrected for any heteroscedasticity in the residuals, is given by

$$Var(\hat{\mathbf{\beta}}) = \frac{n}{n-4}(\mathbf{x}^T\mathbf{x})^{-1}(\mathbf{x}^T\hat{\mathbf{\Omega}}\,\mathbf{x})(\mathbf{x}^T\mathbf{x})^{-1} \quad (19b)$$

The square root of the values on the diagonal of this matrix are often termed heteroscedastic consistent standard errors, or HCSE for short. Appendix C shows that Eq. (19b) is a consistent or asymptotic estimator of $Var(\hat{\mathbf{\beta}})$, which implies that the estimator is valid, strictly speaking, only for very large data sets. MacKinnon and White[23] showed that in small samples Eq. (19c) produced a better estimate of the true variance for $Var(\hat{\mathbf{\beta}})$ (see Appendix C)

$$Var(\hat{\mathbf{\beta}}) = \frac{n}{n-4}(\mathbf{x}^T\mathbf{x})^{-1}\left[\mathbf{x}^T\hat{\mathbf{\Omega}}\,\mathbf{x} - \frac{1}{m}\mathbf{x}^T\hat{e}\hat{e}^T\mathbf{x}\right](\mathbf{x}^T\mathbf{x})^{-1}$$
$$(19c)$$

where $\hat{e}$ is a vector of the $\hat{e}_i^2$. White and MacKinnon[23] showed that in small samples Eq. (19c) produced a better estimate of

the true variance for $Var(\hat{\mathbf{\beta}})$. The square root of the values on the diagonal of this matrix are often termed Jackknife heteroscedastic consistent standard errors, or JHCSE for short. The test statistic t^g can now be corrected for any heteroscedasticity by simply using Eq. (19c) for the denominator in Eq. (14).

Multiple constraints corrected for any heteroscedasticity can now be tested for by writing multiple constraints in matrix form

$$\mathbf{p} = \mathbf{R}\hat{\mathbf{\beta}} = \mathbf{q} \quad (20a)$$

where vector $\hat{\mathbf{\beta}}$ is a sample estimate of $\mathbf{\beta}$ and each row of R corresponds to a single parameter constraint and $\mathbf{q}$ contains the constrained parameter value. For example, Levi De Oliveira Bueno informally looked at whether $\beta_3 = 0$, when the restrictions $\beta_1 = -1$ is assumed true. This dual set of restrictions can be written as

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{q} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad (20b)$$

Appendix C demonstrates that the matrix $\mathbf{p}$ will follow a normal distribution provided $\hat{\mathbf{\beta}}$ is normally distributed (as it is a linear combination of them) which in turn requires the residuals to be normally distributed with a mean vector of zero. Given that $\hat{\mathbf{\beta}}$ is unbiased, the mean value for p is zero

$$\mathbf{p} \sim N[\mathbf{0}, Var(\mathbf{p})]$$

with the covariance matrix of $\mathbf{p}$ given by

$$Var(\mathbf{p}) = \mathbf{R}\,Var[\hat{\mathbf{\beta}}]\mathbf{R}^T \quad (21a)$$

$Var[\hat{\mathbf{\beta}}]$ is given by Eq. (19b) for the asymptotic test or Eq. (19c) for the finite sample version of this test. So if the null hypotheses $H_0: R\beta = q$ is true, the Wald statistic

$$W_k = \mathbf{p}^T[Var(\mathbf{p})]^{-1}\mathbf{p} \quad (21b)$$

has a chi square distribution with k degrees of freedom. k corresponds to the number of constraints being tested and so equals the number of rows contained in $\mathbf{R}$ and $\mathbf{q}$. This is because the square of a standard normal variate has a chi squares distribution (with k degrees of freedom as $\mathbf{p}$ is a kx1 vector containing k standardised variates). There are two versions of this test. W_k^W is the value for W_k in Eq. (21b) when $Var(\hat{\mathbf{\beta}})$ is given by Eq. (19b), whilst W_k^{JK} is the value for W_k in Eq. (21b) when $Var(\hat{\mathbf{\beta}})$ is given by Eq. (19c). In either case it is a test for multiple restrictions adjusted for any heteroscedasticity in the residuals and so is more flexible than the F^R test in Eq. (18a) as that is invalid for non-constant variance in the residuals. An F variant of this test is found by dividing through by the number of restrictions being tests. If the null hypothesis is true, $F^W = \frac{W_k^W}{k}$ and $F^{JK} = \frac{W_k^{JK}}{k}$ follow a F distributions with k and n-4 degrees of freedom.

Predictions

Another approach to assessing the Levi De Oliveira Bueno hypothesis is to estimate the MG relation using just the tensile test data. If the hypothesis is true, then the resulting estimated relation should produce unbiased forecasts of the failure times associated with the creep data set and the actual creep failure times should then be within a multiple of the variability of the predictions given by

$$
s_{\hat{y}_i}^2 = s_e^2 \left[1 + \frac{1}{n - n_1} + \frac{\left[x_i - \frac{\sum_{i=1+n_1}^{n} x_i}{n - n_1} \right]^2}{\sum_{1+n_1}^{n} \left[x_i - \frac{\sum_{i=1+n_1}^{n} x_i}{n - n_1} \right]^2} \right]
$$

(22a)

Eq. (22a) picks up variability due to the residuals e_i and uncertainty about the parameter estimates. An approximate 95% confidence interval for any prediction of y_i is then given by

$$
\hat{y}_i = (\widehat{\beta}_0 + \widehat{\beta}_2 x_{1i}) \mp 2(s_{\hat{y}_i}) \quad (22b)
$$

where $\widehat{\beta}_0$, $\widehat{\beta}_2$ are estimated from

$$
y_i = \beta_o + \beta_2 x_i + e_i \text{ for } i = 1 \text{ to } n_1
$$

using only the tensile data

Evans[24] suggested the randomness of the prediction errors is best measured using the mean percentage absolute error (MPAE). Following Robeson and Cort,[25] the MPAE is formally defined as

$$
MPAE \approx \frac{1}{n - n_1} \sum_{i=1+n_1}^{n} |e_i| \quad (23a)
$$

where e_i is the difference between the actual log time to failure at creep test condition i and the corresponding log time to failure predicted by the MG relation estimated from the tensile data set. As this prediction error is expressed in terms of (natural) log differences, this represents an approximate percentage error. Robeson and Cort showed that the MPAE can be decomposed into three separate parts. The first component of $|e_i|$ is the absolute bias error $|BE| = |(\overline{y_i} - \overline{\hat{y}_i})|$, where the bar indicates the average values for the actual and predicted value for y). Robeson and Cort then remove this component by adding it onto the predictions, $\hat{y}_{pi} = \hat{y}_i + |BE|$. $|BE|$ is clearly a systematic prediction error. They then use least squares to determine the values for δ_0 and δ_1 in the regression

$$
y_i = \delta_0 + \delta_1 \hat{y}_{pi} + \varepsilon_i \quad (23b)
$$

The second part is also a systematic prediction error and is equal to $|f_i| = |\hat{y}_{pi} - (\delta_0 + \delta_1 \hat{y}_{pi})|$. This quantity only exists if δ_1 is different from 1 and δ_0 is different from 0. Robeson and Willmott call this component the proportionality error. The third component is a random prediction error and

equals $|\epsilon_i|$. On this basis, the decomposition of the MPAE is

$$
\text{MPAE}_M = \frac{1}{n - n_1} \sum_{i=1+n_1}^{n} \frac{|BE|}{|BE| + |f_i| + |\epsilon_i|} |e_i| \quad (24a)
$$

$$
\text{MPAE}_R = \frac{1}{n - n_1} \sum_{i=1+n_1}^{n} \frac{|f_i|}{|BE| + |f_i| + |\epsilon_i|} |e_i| \quad (24b)
$$

$$
\text{MPAE}_D = \frac{1}{n - n_1} \sum_{i=1+n_1}^{n} \frac{|\epsilon_i|}{|BE| + |f_i| + |\epsilon_i|} |e_i| \quad (24c)
$$

These can be expressed as proportions of MPAE

$$
|U_M| = \text{MPAE}_M / \text{MPAE} ; |U_R|
$$
$$
= \text{MPAE}_R / \text{MPAE} ; |U_D|
$$
$$
= \text{MPAE}_D / \text{MPAE} \quad (24d)
$$

If the Levi De Oliveira Bueno hypothesis is true, the expectation is that $|U_D| \approx 1$ or 100%, so that the MG relation in the tensile test data predicts creep failure times with an average error of zero - as the $0 = \text{MPAE}_M = \text{MPAE}_R$.

Results

What Levi De Oliveira Bueno found

Levi De Oliveira Bueno assessed their hypothesis by fitting the MG relation to the creep data set and seeing how well it extrapolated to the lower times (to UTS) associated with the tensile test data assuming $\rho_c = 1$ in Eq. (1b). Using the digitised data from the paper by Levi De Oliveira Bueno, $\ln(M_c)$ was estimated as

$$
\bar{w}_c = -3.0025
$$

Consequently, the estimate for M_c was found to be $\exp(-3.0025) = 0.050$. This is very close the value reported by Levi De Oliveira Bueno[10] in Figure 2(a) in Ref [10], namely, 0.051. This small difference is most likely due to digitising the results from that paper.

The solid line in Figure 2 plots $t_{cf} = 0.05(\dot{\varepsilon}_{mc})^{-1}$ and the dashed line is the extrapolation of this relation to the tensile test results. The extrapolation to the tensile test data in Figure 2 looks reasonable and Levi De Oliveira Bueno stated "The agreement of the CSR tensile data with the creep results, according to the Monkman-Grant fit is evident. Therefore, the scatter observed on the creep data in this kind of plots could be of the same nature as observed for the CSR results". In this quote, CSR refers to the tensile test data.

This paper expands on this analysis by considering the following questions:

Does the extrapolation work in the opposite direction?

Given the benefit of the Levi De Oliveira Bueno hypothesis being valid, is to predict creep failure times associated with
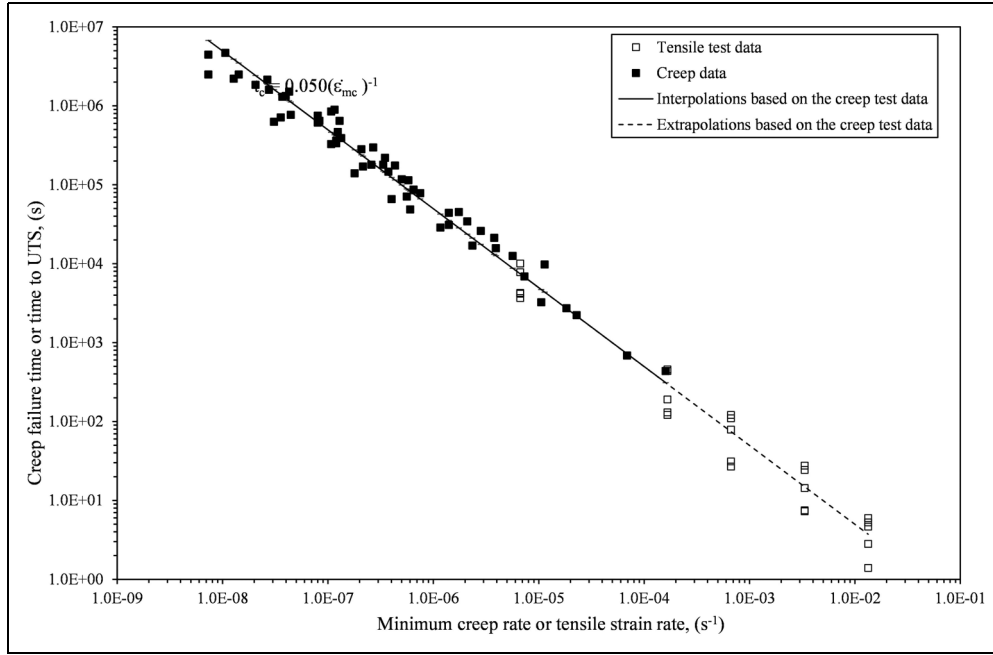
Figure 2. Extrapolation of the MG relation fitted to the tensile test data when forcing $\rho_{ts} = 1$.

typical operating conditions from the very short test times associated with tensile data, it seems reasonable to fit the MG relation to the tensile data and see how well it extrapolates to the creep failure times, assuming $\rho_{ts} = 1$. The results are shown in Figures 3. Using the digitised data from the paper by Levi De Oliveira Bueno, and based on Eq. (4a), $\ln(M_{ts})$ is estimated as

$$\bar{w}_{ts} = -3.1773$$

Consequently, the estimate for M_{ts} was found to be $\exp(-3.1773) = 0.0417$. The solid line in Figure (3a) plots $t_{ts} = 0.0417(\dot{\varepsilon}_{ts})^{-1}$ with the dashed segment of this line showing the extrapolation of this relationship to the creep failure times. The predicted creep failure times along this extrapolated lines look very good. This is confirmed in Fig. (3b) where these creep failure time predictions are plotted against the actual creep failure times. The MPAE in the creep failure time predictions is some 31.11%. The best fit line to the data points in this figure is slightly flatter the ideal 45° line – given by the dashed line in Figure (3b). 24.33% of the MPAE is explained by this phenomenon ($=|U_D|$). The remaining part of the MPAE is random in nature – some 75.67% ($=|U_R|$).

Despite the reasonable fit shown in Figures (3), are the MG constants M_c and M_{ts} really equal if $\rho_c = \rho_{ts} = 1$?

As can be seen from Figure 3 and Figure 2, the values for M_c and M_{ts} are quite different, and despite this, the predictions made from the tensile test data are quite good. So, are these differences statistically significant? When $\rho_c = \rho_{ts} = 1$,

then in Eq. (10a) $\beta_2 = -1$ and $\beta_3 = 0$ giving

$$y_i + x_{1i} = \beta_o + \beta_1 I_i + v_i \text{ for } i = 1 \text{ to } n$$

where the n = 83 observations are from the creep and tensile data sets combined. Here $\beta_0 = \ln(M_{ts})$ and $\beta_0 + \beta_1 = \ln(M_c)$. So, given the assumption that $\rho_c = \rho_{ts} = 1$, $M_c = M_{ts}$, only when $\beta_1 = 0$. Estimated values for M_c and M_{ts} where given in the previous sub section, giving the estimates $\hat{\beta}_0 = \bar{w}_{ts}$ and $\hat{\beta}_1 = \bar{w}_c - \bar{w}_{ts}$ (see Eq. (7b)). So for this sample of data, $\ln(M_c)$ exceeds $\ln(M_{ts})$ by 0.1749. From Eq.(8b), s_v^2 was found to be 0.2046, which is equal to the pooled variance s_w^2 in Eq. (A3). Then from Eq. (B1)

$$Var(\hat{\beta}_o) = \frac{s_v^2}{(n - n_1)} = \frac{0.2046}{25} = 0.00819$$

$$Var(\hat{\beta}_1) = s_v^2 \left[\frac{1}{n_1} + \frac{1}{(n - n_1)} \right] = 0.2046(0.0572)$$
$$= 0.0117$$

The null hypothesis that $\beta_1 = 0$ can be tested using the t statistic of Eq. (8a)

$$t^b = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{s_v^2 \left[\frac{1}{n_1} + \frac{1}{(n - n_1)} \right]}} = \frac{0.1749 - 0}{\sqrt{0.0117}} = 1.62$$

At the 5% significance level the critical value for t is $|t_{\alpha/2,n-2}| = |t_{0.05/2,83-2}| = 1.99$ and so the null hypothesis of $\beta_1 = 0$ is accepted. Put differently, the p-value of 0.11 suggests that the probability of the null hypothesis being true is 11% which is too high to be able to reject the null hypothesis (with a 5% significance level, the chances of wrongly rejecting the null hypothesis must be lower than 5%).
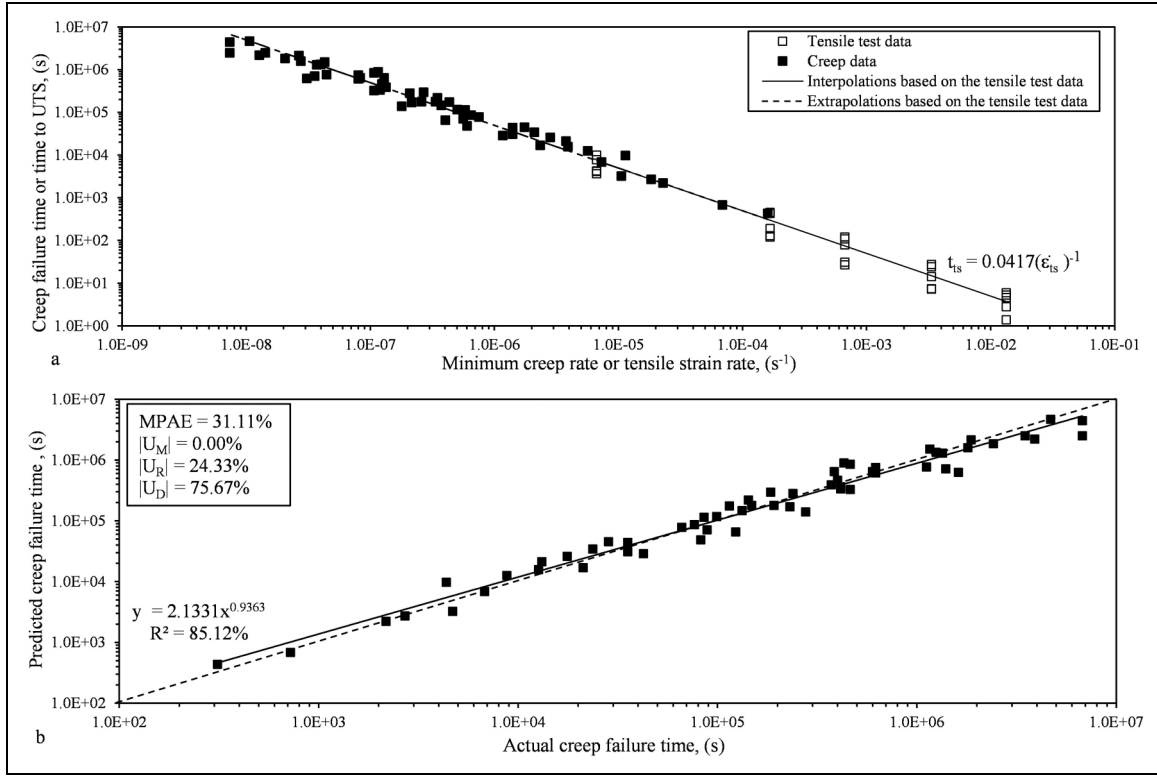
Figure 3. (a) Variation of the minimum creep rate with the rupture time in creep test data, plotted together with the strain rate and time to occurrence of the ultimate tensile stress in the constant strain rate tensile tests. Straight lines refer to the least squares estimates of the MG relation in the creep data (solid line) with extrapolation to the tensile data (dashed line). (b) Creep failure time predictions given by the dashed line in (a) are plotted against the actual creep failure times.

This test suggests that if $\rho_c = \rho_{ts} = 1$, the Monkman Grant relation is the same in the creep and tensile data sets, i.e., under this condition the Levi De Oliveira Bueno cannot be rejected with the required degree of confidence. This supports the conclusion drawn in the previous sub section. But how reliable is this test?

First it requires the v_i to have constant variance or equivalently for $\sigma^2_{wc} = \sigma^2_{wts} = \sigma^2_w$. Using Eq. (A2), the F^V test of Eq. (6) comes out at

$$F^V = \frac{S^2_{w_{ts}}}{S^2_{w_c}} = \frac{0.3214}{0.1555} = 2.07$$

At the 5% significance level, the critical value for F^V is $|F_{\alpha, n1, n-n1}| = |t_{0.05,24,57}| = 1.71$. and so the null hypothesis $\sigma^2_{wc} = \sigma^2_{wts}$ is rejected at this significance level, i.e., can be rejected with less that a 5% chance of being wrong. But at the 1% significance level the critical value for $F^V = 2.13$ and so only at this lower significance level can the null hypothesis of $\sigma^2_{wc} = \sigma^2_{wts}$ be accepted. Given this, it seems sensible to use the HCSE (of Eq. (19c) in the t^b test

$$t^b = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{0.1749 - 0}{\sqrt{0.0159}} = 1.39$$

Irrespective of whether there is any heteroscedasticity, the p-value now increases to 0.17 which suggests that the probability of the null hypothesis being true is 17% which is too high to be able to reject the null hypothesis. So provided the v_i are normally distributed, this test is valid. The normality

test given by Eq. (17) comes out at $\chi^N = 4.56$. At the 5% significance level, normality requires this value to be less than 5.99, and so the assumption of normality of the v_i is accepted at this significance level. With a p-value of 0.102 it is also accepted at the 10% significance level.

Taken all together, the fact that the (heteroscedastic robust) p-value is well above the 10% significance level and that the normality of v_i cannot be rejected at this significance level – the hypothesis that the population values for $\ln(M_{ts})$ and $\ln(M_c)$ are the same, cannot be rejected. The MG constants are not statistically significantly different in value from each other. This is consistent with the good creep failure time predictions made from extrapolating the short time tensile test results.

Are the Monkman-Grant exponents ρ_c and ρ_{ts} in Eqs. (1b,1e) really equal to 1?

The analysis in the previous sub section, is conditional on $\rho_c = \rho_{ts} = 1$ being true. So, this subsection delves into whether this assumption is even true. To test the restriction imposed on the data by Levi De Oliveira Bueno, this restriction is relaxed by using the general model given by Eq.(10a). Table 1 shows the results from estimating the parameters of this model.

For $\rho_c = \rho_{ts} = 1$ the parameter β_2 must equal -1 (which makes $\rho_{ts} = 1$) and the parameter β_3 must equal 0 (which makes ρ_c equal to ρ_{ts}, which in turn equals 1 when $\beta_2 = -1$). The t-value for the null hypothesis H_o: $\beta_3 = 0$ is 0.58 from

Table 1, giving a probability that this null hypothesis is true of 56.32%. Thus, this hypothesis easily accepted at the 5% significance level, implying that the Monkman-Grant exponent is the same in the tensile and creep data sets. The t-value for testing the null hypothesis H_o: $\beta_2 = -1$ is 0.92 ($-0.9629 - -1$)/0.0403 so the probability of this constraint being true is 36%. Again, this hypothesis on its own can be easily accepted, i.e., that that $\rho_c = \rho_{ts} = 1$. But can they both be accepted jointly? Well, one thing to note in Table 1, is the parameter standard errors adjusted for heteroscedasticity (column 4) are quite different from the unadjusted standard errors (column 3). Given this, the assumption of constant variance

Table 1. Least squares estimate of the general model given by Eq. (10a).

Parameter	Estimate	Standard Error[a]	JHSCE[b]	t -value	p-value
$\hat{\beta}_o$	−2.8954	0.2681	0.3422	−8.44	0.0000
$\hat{\beta}_1$	0.8416	0.4643	0.4678	1.80	0.0758
$\hat{\beta}_2$	−0.9629	0.0334	0.0403	−23.90	0.0000
$\hat{\beta}_3$	0.0265	0.0418	0.0457	0.58	0.5632

[a]The numbers is this column are given by the square root of the diagonal elements of matrix $Var(\hat{\beta})$ in Eq. (11b) – and so make no adjustment for any heteroscedasticity present in the residuals e_i. [b]The numbers is this column are given by the square root of the diagonal elements of the matrix $Var(\hat{\beta})$ in Eq. (19c) – and so are the MacKinnon and White[23] Jackknife standard errors that adjuste for the presence of any heteroscedasticity in the residuals e_i. t-value is the t test statistic for the null hypothesis that the population value for the shown parameters is zero, based on the jackknife standard errors (=Estimate/JHSCE). p-value is the probability of this null hypothesis being true based on the shown t-values.

in the e_i looks unlikely to be true and so there is little point in forming the F^R test of Eq. (18a) for this joint restriction as it requires this assumption to be true. So, the Wald test given by Eq. (D11b, D12) that adjusts for heteroscedasticity in the errors is used and comes out as $F_2^{JK} = 4.81$ with a p-value of 0.011. The chances of $\rho_c = \rho_{ts}$ and of $\rho_c = \rho_{ts} = 1$ being true, is only 1.1%. This hypothesis is clearly rejected at the 5% significance level. The normality of the e_i required for this test to be valid is accepted at the 5% significance level as $\chi^N = 4.13$ (with a 5% critical value of 5.99). The best conclusion to draw from these results is that the Monkman-Grant exponent has the same value in the creep and tensile data sets, but that this value is less than 1. Consequently, the result shown above that $M_c = M_{ts}$, *cannot be accepted as a valid result as it was conditional on* $\rho_c = \rho_{ts} = 1$ *being true.*

Is the dual constraint ($M_c = M_{ts}$ and $\rho_c = \rho_{ts}$) needed for the *Levi De Oliveira Bueno hypothesis to be true accepted by the data?*

The equivalence of M_c and M_{ts} therefore needs to be looked at again with a test that is not conditional on $\rho_c = \rho_{ts} = 1$ being true. For the the Levi De Oliveira Bueno hypothesis to be true now requires that $M_c = M_{ts}$, (or $\beta_1 = 0$) and $\rho_c = \rho_{ts}$ (or $\beta_3 = 0$). The Wald statistic for testing this joint restriction comes out at $F_2^{JK} = 5.72$ with a p-value of 0.005. So the chances of $\rho_c = \rho_{ts}$ and $M_c = M_{ts}$ being true is only 0.5% and so this hypothesis is clearly rejected even at the 1% significance level. This suggests that the creep failure times predicted from an extrapolation of the MG relation
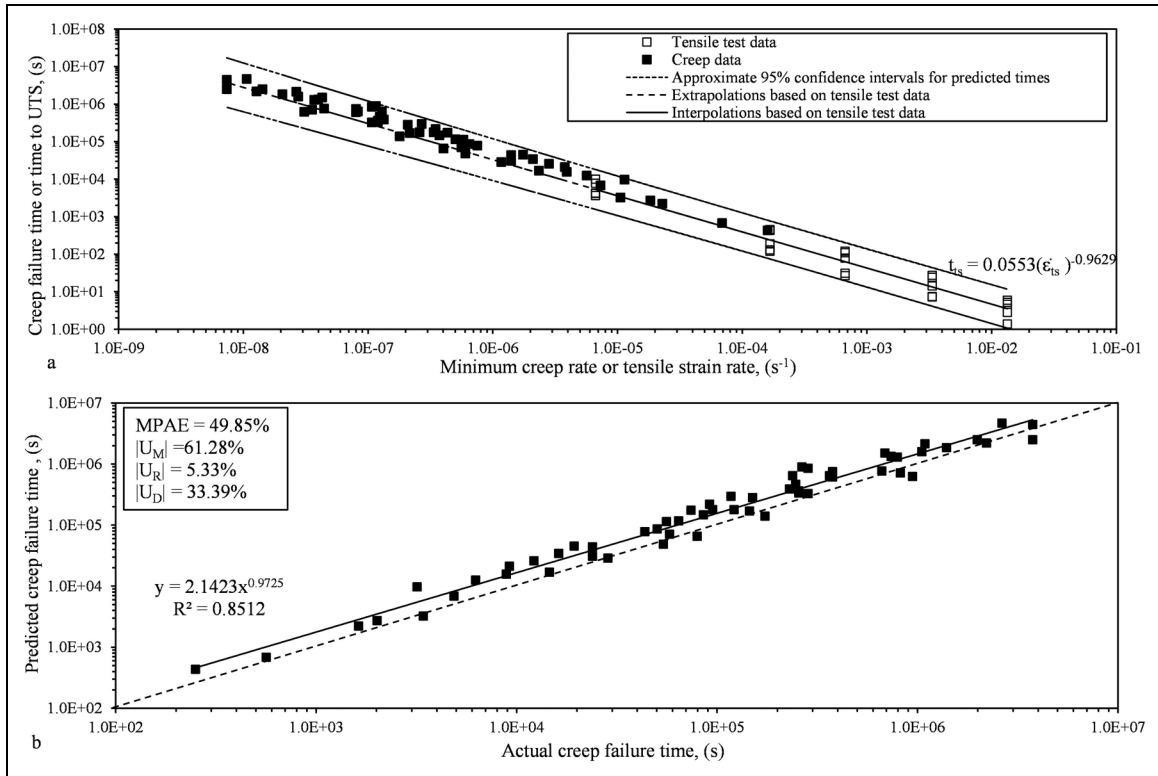

Figure 4. (a) Extrapolation of the MG relation fitted to the tensile test data allowing $\rho_{ts} < 1$. (b) Actual v predicted creep failure times.

estimated from just the tensile data will consistently over estimate the actual creep failure times as $\ln(M_c) > (M_{ts})$ by 0.8416 when allowing for the MG exponent to differ from -1.

How random are the errors made in predicting of creep failure times using only tensile test results?

The section looks at whether such predictions are indeed biased. The solid line in Figure 4(a), shows a fit of Eq. (1e) to the tensile test data. This long-dashed line then extrapolates this relation to the creep data set. The resulting creep failure time predictions are towards the lower end of the range of actual creep failure times, suggesting a degree of average bias. The short dashed lines show the approximate 95% confidence interval for these predictions based on Eq. (22a). Whilst all the data points are within these limits, the limits are quite wide and the actual creep failure times tend to cluster more towards the upper rather than the lower limit.

In Figure 4(b) the MPAE is almost 50% and is some 20 percentage points higher than when the $\rho_{ts} = 1$ constraint is imposed on the tensile data set. What's more, only a third of this MPAE is random in nature – down from 75% when the $\rho_{ts} = 1$ constraint is imposed on the tensile data set. As a result, the average prediction error is 61%, whereas as when the $\rho_{ts} = 1$ constraint is imposed on the tensile data set, this average error did not exist. So, whilst the $\rho_{ts} = 1$ constraint is not the best representation of the tensile data, it produces substantially better creep failure time predictions. The problem is that if tensile data is to be used to predict unknown creep failure times at operating conditions, these times will not be known to assess whether imposing $\rho_{ts} = 1$ will produce sensible operating lives.

Conclusion

The Levi De Oliveira Bueno hypothesis offers the possibility of reducing the cost and length of the development cycle for new materials operating at high temperatures. This paper outlined a number of statistical tests for this hypothesis, highlighting clearly the assumptions behind these tests, together with tests for these assumptions being true. Whilst it was found that the MG relation fitted to tensile test data could provide very accurate long term creep life predictions when this exponent of this relation was forced to equal 1, this additional constraint was not accepted by the data using the tests outlines in this paper. When the constraint on this exponent was relaxed, the MG relation fitted to tensile data was no longer able to predict creep life without bias, with the mean absolute percentage error rising from around 30% to nearly 50%. In fact, the average prediction error made up over 60% of the mean percentage absolute error in predicting creep failure times.

One area of future work would include the application of the statistical tests outlined in this paper to other materials to assess whether the findings of this paper are material specific or apply everywhere. Of particular interest would be an analysis of more ductile materials such as Nickel

based super alloys where the value for ρ_c is considerably different from -1 (and so the kinds of equivalence seen in this paper are less likely to holds). In doing this, some consideration should be given to the nature of the tensile testing. It should include testing up to the high strain rates that result in the tensile strength becoming independent of further increases in the strain rate. In addition to this, tests should be carried out down to a temperature where the tensile strain becomes strain invariant. Finally, the sample size should be selected so as to achieve a confidence interval of the required size - $n \approx (1.96 s_e / E)^2$, where E is the required width of the 95% confidence interval for failure time predictions.

Whilst there is quite a lot of creep data in the public domain on such ductile materials (for example, the NIMS creep data bases), assessing the reliability of extrapolating from tensile test data would require additional tensile testing that takes into account the above considerations on the test matrix. If it can be demonstrated that such extrapolations produce creep life predictions with a MPAE whose decomposition has small values for U_M and U_R then this could provide the confidence in the technique required for designers to adopt the approach to developing new materials. A theoretical basis for the equivalence between creep and tensile test data would also help in this respect – indeed this would constitute an important area of future research. A good starting point would be within the theta creep methodology where Evans[26] derived the shape of a uniaxial creep curve in terms on hardening, softening and damage mechanisms. Recently, Harrison et al.[27] showed that this approach to accounting for preexisting damage can explain the shape of a uniaxial creep curve obtained when the stress is varied during a uniaxial creep test. Given that a tensile test can be viewed as a creep test in which the stress is continually changed (to maintain a constant strain rate), this evidence suggests some sort of equivalence between a creep and a tensile test.

References

1. *Boiler and pressure vessel code*. New York: ASME, 2004, 450 2.
2. Allen D and Garwood S. *Energy materials-strategic research Agenda*, http://www.matuk.co.uk/docs/1_ StrategicResearch Agenda%20FINAL.pdf (2007, accessed 28 May 2019).
3. Yang M, Wang Q, Song XL, et al. On the prediction of long term creep strength of creep resistant steels. *Int J Mater Res* 2016; 107: 133–138.
4. Wilshire B and Battenbough AJ. Creep and creep fracture of polycrystalline copper. *Mater Sci Eng A* 2007; 443: 156–166.
5. Wilshire B and Scharning PJ. Prediction of long term creep data for forged 1Cr-1Mo-0.25 V steel. *Mater Sci Technol* 2008; 24: 1–9.

6. Wilshire B and Whittaker M. Long term creep life prediction for grade 22 (2.25Cr-1Mo) steels. *Mater Sci Technol* 2001; 27: 642–647.

7. Wilshire B and Scharning PJ. A new methodology for analysis of creep and creep fracture data for 9–12% chromium steels. *Int Mater Rev* 2008; 53: 91–104.

8. Whittaker MT, Evans M and Wilshire B. Long-term creep data prediction for type 316H stainless steel. *Mater Sci Eng A* 2012; 552: 145–150.

9. Monkman FC and Grant NJ. An empirical relationship between rupture life and minimum creep rate in creep-rupture tests. *Proc Am Soc Test Mater* 1956; 56: 593–620.

10. Bueno LdO. Creep behaviour of 2.25 Cr-1Mo steel-an equivalence between hot tensile and creep testing data. In: Proceedings of the ECCC Creep Conference on Creep & Fracture in High Temperature Components–Design & Life Assessment, London, 12–14 September 2005, pp.969–980.

11. Osgerby S and Dyson BF. Proceedings of the 5th.international conference on creep and fracture of engineering materials and structures, swansea. In: Wilshire B and Evans RW (eds). London: The Institute of Metals, 1993, pp.53–61.

12. NIMS Creep Data Sheet No.14B. *Data sheets on the elevated-temperature properties of 18Cr-12Ni-Mo stainless steel plates for reactor vessels (316HP)*. Tokyo, Japan: National Research Institute for Metals, 1988.

13. Evans M. Incorporating the Wilshire equations for time to failure and the minimum creep rate into a continuum damage mechanics for the creep strain of Waspaloy. *Mater High Temp* 2022; 39: 113–148.

14. Dobes F and Milicka K. The relation between minimum creep rate and time to fracture. *Metal Sci* 1976; 10: 382–384.

15. Sklenicka V, Kucharova K, Kral P, et al. Applicability of empirical formulas and fractography for assessment of creep life and creep fracture modes of tempered martensitic 9%Cr steel. *Kov Mater* 2017; 55: 69–80.

16. Abe F. Effect of quenching, tempering, and cold rolling on creep deformation behavior of a tempered martensitic 9Cr-1 W steel. *Metall Mater Trans* 2003; 34A: 913–925.

17. ASTM E 139-11. American Society for testing and materials. In: *Annual book of ASTM standards*, vol. 03. West Conshohocken, PA, USA: ASTM International, 2011, pp.309–319.

18. BUENO LO. Máquinas-protótipos para ensaios de fluência em metais a altas temperaturas. Parte 1: Detalhes de construção e montagem do equipamento. In: *Anais do II ETUAN -Encontro de Tecnologia e Utilização dos Aços Nacionais*. Maio: Rio de Janeiro (RJ), 1987, pp.916–934.

19. White H. A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980; 48: 817–838.

20. Bowman KO and Shenton LR. Omnibus test contours for departures from normal ity based on $\sqrt{b_1}$ and b_2. *Biometrika* 1975; 62: 243–250.

21. Jarque CM and Bera AK. A test for normality of observations and regression residuals. *Int Stat Rev* 1987; 55: 163–172.

22. Doornik JA and Hansen H. An omnibus test for univariate and multivariate normal ity. *Oxford Bull Econ Stat* 2008; 70: 927–939.

23. MacKinnon JG and White H. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J Econom* 1985; 29: 305–325.

24. Evans M. Assessing the predictive performance of creep models using absolute rather than 635 squared prediction errors: An application to 2.25Cr-1Mo steel. *Mater High Temp* 2023; 40: 457–468.

25. Robeson SM and Cort JW. Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *Plos one* 2023; 18.

26. Evans RW. A constitutive model for the high-temperature creep of particle-hardened alloys based on the θ projection method. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2000; 456: 835–868.

27. Harrison WJ, Whittaker MT and Deen C. Creep behaviour of Waspaloy under non-constant stress and temperature. *Mater Res Innov* 2013; 17: 323–326.

28. Hinkley DV. Jackknifing in unbalanced situations. *Technometrics* 1977; 19: 285–292.

29. Horn SD, Horn RA and Duncan DB. Estimating heteroskedastic variances in linear models. *J Am Stat Assoc* 1975; 70: 380–385.

30. Efron B. The Jackknife, the bootstrap and other resampling plans. In: *Society for industrial and applied mathematics*. Philadelphia: PA, 1982.

Appendix

A. Testing differences between two population means

The sample means provide unbiased estimates of the population means μ_c and μ_{ts}

$$\bar{w}_{ts} = \frac{\sum_{i=1}^{n_1} (y_i + x_i)}{n_1} \quad \text{and} \quad \bar{w}_c = \frac{\sum_{i=1+n_1}^{n} (y_i + x_i)}{n - n_1} \tag{A1}$$

These are sample estimates of μ_c and μ_{ts}, and collecting a different creep and tensile sample of the same size will result in different values for $\bar{w}_c$ and $\bar{w}_{ts}$. It can be shown that if all possible samples of size n_1 and $n - n_1$ of observations on y_i and x_i are taken, and for each sample the estimates $\bar{w}_c$ and $\bar{w}_{ts}$ are computed, then a large number of estimates for each of these will be obtained, and it is well known that the means (called the expected value E) of these values will equal the true (i.e., population) mean values. That is, $E(\bar{w}_c) = \mu_c$ and $E(\bar{w}_{ts}) = \mu_{ts}$. The sample means are said to be unbiased estimates. It is also well-known that the variance in all these means is given by $\sigma^2_{wc}/(n-n_1)$ and σ^2_{wts}/n_1, where σ^2_{wc} and σ^2_{wts} are the population variances for w_c and w_{ts} respectively. Again, these can be estimated unbiasedly from samples of data

$$S^2_{w_{ts}} = \frac{\sum_{i=1}^{n_1} [\, w_{ts,i} - \bar{w}_{ts} \,]^2}{n_1 - 1} \quad \text{and}$$

$$S^2_{w_c} = \frac{\sum_{i=1+n_1}^{n} [\, w_{c,i} - \bar{w}_c \,]^2}{(n - n_1) - 1} \tag{A2}$$

with $E[S^2_{w_{ts}}] = \sigma^2_{wts}$ and $E[S^2_{w_c}] = \sigma^2_{wc}$. The t-test for differences in population means requires the further assumption that the variances is each test data set are the same $\sigma^2_{wc} = \sigma^2_{wts} = \sigma^2_w$. If this assumption is true, then an unbiased estimate of this common variance is

$$S^2_w = \frac{(n_1 - 1)S^2_{w_{ts}} + (n - n_1 - 1)S^2_{w_c}}{(n - 2)} \tag{A3}$$

The central limit theorem states that as the sample size of each dataset increases, the variables $\bar{w}_c$ and $\bar{w}_{ts}$ tend towards having a normal distribution, irrespective of how the data points making up each mean are distributed. The approximation to a normal distribution is excellent in samples sizes of 30 or more. Further, linear combinations of normal variables will also have a normal distribution. Thus

$$\bar{w}_c - \bar{w}_{ts} \sim N\left[\mu_c - \mu_{ts}, \frac{\sigma_w^2}{n_1} + \frac{\sigma_w^2}{(n-n_1)}\right] \quad (A4)$$

given that variances are always additive in nature. Eq. (A4) states that the variable $\bar{w}_c - \bar{w}_{ts}$ is normally distributed (more so in large samples) with a population mean value of $\mu_c - \mu_{ts}$ and a variance of $\frac{\sigma_w^2}{n_1} + \frac{\sigma_w^2}{n_1}$. It follows from this that the following Z variate has a standard normal distribution

$$Z = \frac{[\bar{w}_c - \bar{w}_{ts}] - E[\bar{w}_c - \bar{w}_{ts}]}{\sqrt{\frac{\sigma_w^2}{n_1} + \frac{\sigma_w^2}{(n-n_1)}}}$$

$$= \frac{[\bar{w}_c - \bar{w}_{ts}] - (\mu_c - \mu_{ts})}{\sqrt{\frac{\sigma_w^2}{n_1} + \frac{\sigma_w^2}{(n-n_1)}}} \sim N(0, 1) \quad (A5)$$

Eq. (A5) is the starting point for testing various hypothesis. For example, the Levi De Oliveira Bueno hypothesis corresponds to the null hypothesis: $\mu_c - \mu_{ts} = 0$. Then if this hypothesis is true, the Z statistic

$$Z = \frac{[\bar{w}_c - \bar{w}_{ts}] - 0}{\sqrt{\frac{\sigma_w^2}{n_1} + \frac{\sigma_w^2}{(n-n_1)}}} \sim N(0, 1) \quad (A6)$$

has a standard normal distribution. Values for the probability of observing various values for Z from the standard normal distribution have been calculated and tabulated. For example, there is a 5% chance of observing a Z value outside the range -1.96 to $+1.96$. If the calculated value for $|Z|$ from Eq. (6e) then exceeds 1.96 it can be said that the chances of getting such a value is less than 5%. As these Z values were calculated under the assumption that $\mu_c - \mu_{ts} = 0$, it follows that there is also less than a 5% chance that the true value for $\mu_c - \mu_{ts}$ equals zero. In such a situation the null hypothesis would be rejected. 1.96 is an example of a critical value for Z (called Z_α) and $\alpha = 5\%$ is an example of a significance level. Researchers typical work with a small value for α (1, 5 or 10%) so that if the null hypothesis is rejected the chances of the null hypothesis actual being true is very small. The decision rule then is reject the null hypothesis whenever $|Z| > |Z_{\alpha/2}|$ and run at most an $\alpha\%$ chance of being wrong in a decision to reject the null hypothesis. Alternatively, the p-value is the probability of observing $|Z|$ or more, and so is equivalent to the probability of the null hypothesis being true. The null hypothesis is then rejected whenever the p-value $< \alpha\%$.

Unfortunately, σ_w is rarely known and needs to be estimated from the sample of data. Eq. (A3) is an unbiased estimate of this variance under the equal variance assumption, and if σ_w^2 is replaced with s_w^2 in Eqs. (A6), then

$$t^a = \frac{[\bar{w}_c - \bar{w}_{ts}] - 0}{\sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{(n-n_1)}}} \quad (A7)$$

has a student t distribution with n – 2 degrees of freedom.

B. Least squares version of test for mean differences

$\hat{\beta}_0$ and $\hat{\beta}_1$ in Eq. (7b) are unbiased estimates of β_o and β_1 in Eq. (7a) provided the variables I and v are independent of each other. It can again be shown that the variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$Var(\hat{\beta}_o) = \frac{\sigma_v^2}{(n-n_1)} \text{ and } Var(\hat{\beta}_1)$$

$$= \sigma_v^2 \left[\frac{1}{n_1} + \frac{1}{(n-n_1)}\right] \quad (B1)$$

where σ_v^2 is the variance and the residuals v_i in Eq. (7a). If the v_i are further assumed to be normally distributed and if the variance of the residuals v_i is a constant (i.e., the same value in the creep and tensile data sets) then

$$\hat{\beta}_j \sim N[\beta_{1j}, Var(\hat{\beta}_j)] \quad j = 1 \text{ to } 2$$

and so

$$Z = \frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma_v^2\left[\frac{1}{n_1} + \frac{1}{(n-n_1)}\right]}}$$

$$\sim N(0, 1) \qquad (B2)$$

Unfortunately, σ_v^2 is unknown, but an unbiased estimate of it is

$$s_v^2 = \frac{\sum_{i=1}^n \hat{v}_i^2}{(n-2)} \quad (B3)$$

where the $\hat{v}_i^2$ are calculated from Eqs. (7a) using the parameter estimates given by Eq. (7b). If σ_v^2 is replaced by s_v^2 in Eq. (B1), then

$$t^b = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s_v^2\left[\frac{1}{n_1} + \frac{1}{(n-n_1)}\right]}} \quad (B4)$$

has a student t distribution with n-2 degrees of freedom. The test statistic t^a in Eq. (A7) is equivalent to t^b in Eq. (B4). This is because $s_w^2 = s_v^2$ and so

$$Var(\hat{\beta}_1) = s_v^2\left[\frac{1}{n_1} + \frac{1}{(m-n_1)}\right]$$

$$= s_w^2\left[\frac{1}{n_1} + \frac{1}{(n-n_1)}\right] \quad (B5)$$

and so the denominator of Eq. (A7) and Eq. (B4) are the same. Also, from Eq. (7b), $\hat{\beta}_1$ is the difference between

the average value for (y + x) in the tensile and creep data sets. Thus, the numerator of Eq. (A7) and Eq. (B4) are also the same. Thus, the t test given by Eqs. (A7) and Eq. (B4) are the same test. The former test assumes that the mean values for (y + x) in both the creep and tensile data sets are normally distributed and that the variance for (y + x) in both the creep and tensile data sets are constant equal to S_w^2. The latter test assumes, equivalently, that v is normally distributed with a constant variance equal to s_v^2.

C. Estimating the variance of the least squares parameter estimates

The covariance matrices for $\widehat{\boldsymbol{\beta}}$ when no assumptions are made about whether the residuals e_i are independent or homoscedastic is given by

$$Var(\hat{\boldsymbol{\beta}}) = (\mathbf{x}^\mathsf{T}\mathbf{x})^{-1}\mathbf{x}^\mathsf{T}\boldsymbol{\Omega}\mathbf{x}(\mathbf{x}^\mathsf{T}\mathbf{x})^{-1} \quad (C1)$$

where $E(\mathbf{e}\mathbf{e}^\mathsf{T}) = \boldsymbol{\Omega}$. To become operational, an estimate of $\boldsymbol{\Omega}$ is required. Conventionally, it is assumed that the residuals have a constant variance given by σ_e^2

$$E(\mathbf{e}\mathbf{e}^T) = \boldsymbol{\Omega} = \sigma_e^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \cdots & 1 \end{pmatrix} \quad (C2)$$

When the residuals are independent but heteroscedastic, the covariance matrix of the residuals takes the form

$$E(\mathbf{e}\mathbf{e}^\mathsf{T}) = \boldsymbol{\Omega} = \begin{pmatrix} \sigma_{e1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{e2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \cdots & \sigma_{en}^2 \end{pmatrix} \quad (C3)$$

White[19] has proposed estimating $\boldsymbol{\Omega}$ in Eq. (C3) using the squared residuals calculated from Eqs. (10a) using the parameter estimates given by Eq. (11a)

$$\hat{\boldsymbol{\Omega}} = \begin{pmatrix} \hat{e}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{e}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \cdots & \hat{e}_n^2 \end{pmatrix} \quad (C4)$$

Under this procedure, Eq. (C1) becomes

$$Var(\hat{\boldsymbol{\beta}}) = \frac{n}{n-4}(\mathbf{x}^T\mathbf{x})^{-1}(\mathbf{x}^\mathsf{T}\widehat{\boldsymbol{\Omega}}\,\mathbf{x})(\mathbf{x}^T\mathbf{x})^{-1} \quad (C5)$$

The term n/(n-4) is a degrees of freedom correction suggested by Hinkley,[28] that is made to the estimate proposed by White, to take into account the fact that the least squares residuals tend to be too small in relation to their population values even in the absence of heteroscedasticity. White[19] and Hinkley[28] have shown that when Var($\hat{\boldsymbol{\beta}}$) in Eq. (C5) are estimated using larger and larger samples (i.e., as n → ∞), the probability

that the difference between Var($\hat{\boldsymbol{\beta}}$) and the true or population value will become arbitrarily small is close to 1. Thus Eqs. (C5) is a consistent or asymptotic estimator of Var($\hat{\boldsymbol{\beta}}$) which implies that the estimator is valid, strictly speaking, only for very large data sets. The square root of the diagonal elements of Var($\hat{\boldsymbol{\beta}}$) in Eq. (C5) are termed heteroscedastic consistent standard errors, or HCSE for short. These standard errors are robust to the presence of heteroscedasticity in the residuals e_i.

Horn et al.[29] have suggested another consistent estimator of Var($\hat{\boldsymbol{\beta}}$) using the following estimate of $\boldsymbol{\Omega}$

$$\tilde{\boldsymbol{\Omega}} = \begin{pmatrix} \tilde{e}_1^2 & 0 & \cdots & 0 \\ 0 & \tilde{e}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \cdots & \tilde{e}_n^2 \end{pmatrix} \quad (C6a)$$

where

$$\tilde{e}_i = \frac{\hat{e}_i}{1 - l_{ii}} \quad (C6b)$$

where l_{ii} is the diagonal element of the matrix $\mathbf{x}(\mathbf{x}^\mathsf{T}\mathbf{x})^{-1}\mathbf{x}^\mathsf{T}$. These authors show that $\tilde{e}_i^2$ is an "almost" unbiased estimator of the σ_e^2.

All of these estimators are related to the jackknife procedure. Efron[30] notes that the White estimator of Var($\hat{\boldsymbol{\beta}}$) can also be obtained by the infinitesimal jackknife method. The Hinkley[19] estimator of Var($\hat{\boldsymbol{\beta}}$) can be obtained using the weighted jackknife. All of this suggests that the jackknife method can provide another modified HCSE estimator. The jackknife technique involves recomputing the estimates of the parameters in Eq. (11a) n times, each time dropping one of the observations. The sample variability of these computed estimates is then an estimate of the variability of the original estimator. That is

$$\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} - (\mathbf{x}^\mathsf{T}\mathbf{x})^{-1}\mathbf{x}_i^\mathsf{T}\tilde{e}_i \quad (C7)$$

where $\hat{\boldsymbol{\beta}}_i$ denotes the estimate of $\boldsymbol{\beta}$ based on all observations except the ith one, and $\mathbf{x}_i$ denotes the ith row of $\mathbf{x}$. Essentially then the jackknife is a form of cross validation. The Jackknife estimate of Var($\hat{\boldsymbol{\beta}}$) is the sample variance of all the $\boldsymbol{\beta}_{(j)}$

$$\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\beta}}_i - \bar{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_i - \bar{\boldsymbol{\beta}})^\mathsf{T} \quad (C8)$$

where $\bar{\beta}$ is the sample mean of all the n values for $\hat{\boldsymbol{\beta}}_i$. White and MacKinnon[23] showed that that the values given by Eq. (C8) can be obtained without the iterations necessary in implementing this equation using

$$Var(\hat{\boldsymbol{\beta}}) = \frac{n-1}{n}(\mathbf{x}^T\mathbf{x})^{-1}\left[\mathbf{x}^T\,\tilde{\boldsymbol{\Omega}}\,x - \frac{1}{\boldsymbol{n}}\mathbf{x}^T\tilde{e}\tilde{e}^T\mathbf{x}\right](\mathbf{x}^T\mathbf{x})^{-1}$$

$$(C9)$$

where $\tilde{e}$ is a vector of the $\tilde{e}_i$. White and MacKinnon[23] showed that in small samples Eq. (C9) produced a better estimate of the true variance for $Var(\hat{\beta})$.

D. Testing multiple restrictions

Various parameter restrictions can be written in the following general way

$$\mathbf{p} = \mathbf{R}\hat{\beta} = \mathbf{q} \quad (D1)$$

For example, for the null hypothesis H_o: $\beta_2 = \beta_3 = 0$ in Eq. (10a), $\mathbf{R}$ and $\mathbf{q}$ take the form

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{q} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (D2a)$$

and for the null hypothesis H_o: $\beta_1 = -1$ and $\beta_3 = 0$

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{q} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad (D2b)$$

The vector $\hat{\beta}$ is a sample estimate of β. That is, this is a random vectors that has a sampling distribution (like the sampling distribution of the sample mean). Consequently, the vector $\mathbf{p}$ also has a sampling distribution. To progress further, the residuals e_i in Eq. (10a) must be assumed to have a normal distribution, so that the vectors $\hat{\beta}$ and $\mathbf{p}$ are then normally distributed (as any linear combination of normal variates will also have a normal distribution). The mean of the sampling distribution of $\mathbf{p}$ ($E(p)$) is found using the rules associated with the expectations operator E, and the additional assumption that x_i are uncorrelated with e_i (which ensures that $E[\hat{\beta} = \beta]$)

$$E[\mathbf{p}] = RE[\hat{\beta}] - E[q] = R\beta - q \quad (D3)$$

as the expected value of a constant is the constant itself. It follows that $E[\mathbf{p}] = 0$ only if $R\beta = q$, i.e only if the null hypothesis is true (with $\mathbf{q} = 0$ under the null, this requires $R\beta = 0$).

The variance of $\mathbf{p}$ is found by further assuming that the residuals e_i have constant variance given by σ_e^2. It can be shown that under these assumption, the following formula provides an efficient (i.e., no other linear estimator has a smaller variance or level of uncertainty) estimate for the variance of $\hat{\beta}$

$$Var(\hat{\beta}) = \sigma_e^2 (\mathbf{x}^T\mathbf{x})^{-1} \quad (D4)$$

and so the variance of $\mathbf{p}$ is

$$Var[\mathbf{p}] = Var[\mathbf{R}\hat{\beta} - \mathbf{q}] = Var[\mathbf{R}\hat{\beta}]$$
$$= \mathbf{R}Var[\hat{\beta}]\mathbf{R}^T = \sigma_e^2 \mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T \quad (D5)$$

So under the above-mentioned assumptions, and if the null hypothesis $R\beta = q$ is true, the random vector $\mathbf{p}$ has a normal distribution with mean $\mathbf{0}$ and variance given by Eq. [D5]

$$\mathbf{p} \sim N[\mathbf{0}, \sigma_e^2 \mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T] \quad (D6)$$

So if the null hypotheses H_0: $R\beta = q$ is true, the Wald statistic

$$W_k = \mathbf{p}^T [Var(\mathbf{p})]^{-1}\mathbf{p}$$
$$= \frac{1}{\sigma_e^2}(\mathbf{R}\hat{\beta} - \mathbf{q})^T [\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T]^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q}) \quad (D7)$$

has chi square distributions with degrees of freedom equal to the number of rows in $\mathbf{q}$ – the number of parameter restrictions, k. This is because the square of a standard normal variate has a chi squares distribution (with 2 degrees of freedom if $\mathbf{p}$ is a 2 × 1 vector containing 2 standardised variates).

When σ_e^2 is not known, it can be replaced by any consistent estimator of it, and then Eq. (D7) become valid asymptotically. Under the assumption that the residuals e_i have constant variance (i.e., are homoscedastic), one such consistent estimator is

$$s_e^2 = \frac{\sum_{i=1}^{n}\hat{e}_i^2}{n-4} \quad (D8)$$

where $\hat{e}_i^2$ is calculated from Eqs. (10b) using the parameter estimates given by Eq. (11a). Substituting this estimator of σ_e^2 into Eq. (D7) and dividing through by the number of restrictions k gives the test statistic

$$F_k = \frac{\{(\mathbf{R}\hat{\beta} - \mathbf{q})^T [\mathbf{R}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^{T-}]^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})\}/k}{s_e^2}$$
$$(D9)$$

If the null hypothesis is true (and the above assumptions are valid), this F-statistic will follow an F-distribution with k and n − 4 degrees of freedom. The null hypothesis is then rejected when F exceeds its critical value at some chosen level of significance (typically a 5% chance of the null hypothesis being true).

When the residuals e_i are heteroscedastic, a lot of the desirable properties of least squares estimators break down. The estimator given by Eq. (11a) remains unbiased, but the estimator given by Eqs. (11b) becomes biased. When the variance of e_i increase with increasing values for x_i, the mean values of the estimated variances of $\hat{\beta}$ and $\mathbf{p}$ given by Eqs. (11a,D5) under-estimate their true variances. This will also result in the test statistic W_k rejecting the null hypothesis more frequently than it should. So, when the residuals are heteroscedastic there are two possible estimators for $Var[\mathbf{p}]$

$$Var[\mathbf{p}] = Var[\mathbf{R}\hat{\beta} - \mathbf{q}] = Var[\mathbf{R}\hat{\beta}]$$
$$= \mathbf{R}Var[\hat{\beta}]\mathbf{R}^T$$
$$= \frac{n}{n-4}R(\mathbf{x}^T\mathbf{x})^{-1}(\mathbf{x}^T\hat{\Omega}\mathbf{x})(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T \quad (D10a)$$

which is based on White [19] estimator of $Var(\hat{\beta})$ and

$$Var[\mathbf{p}] = \frac{n-1}{n}R(\mathbf{x}^T\mathbf{x})^{-1}\left[\mathbf{x}^T\tilde{\Omega}\mathbf{x} - \frac{1}{n}\mathbf{x}^T\tilde{e}\tilde{e}^T\mathbf{x}\right](\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R}^T$$
$$(D10b)$$

which is the White and Mackinnon[23] Jackknife estimator. This gives two versions of the Wald statistic for testing k parameter restrictions

$$W_k^W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T \left[\frac{n}{n-4} R(\mathbf{x}^T\mathbf{x})^{-1}(\mathbf{x}^T\hat{\boldsymbol{\Omega}}\,\mathbf{x})(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R^T} \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})$$

(D11a)

which has a chi square distribution with k degrees of freedom (χ_k^2) under the null hypothesis when the sample size is very large, and

which is much closer to a chi square distribution with k degrees of freedom (χ_k^2) under the null hypothesis than

W_k^W when the sample size is finite. The F version of these tests is obtained by simply dividing through by the number of restrictions being tested

$$F_k^W = \frac{W_k^W}{k} \quad \text{and} \quad F_k^{JK} = \frac{W_k^{JK}}{k} \quad \text{(D12)}$$

which has an F distribution with k and n - 4 degrees of freedom under the null hypothesis (the former only asymptotically so).

$$W_k^{JK} = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^T \left[\frac{n}{n-1} R(\mathbf{x}^T\mathbf{x})^{-1} \left[\mathbf{x}^T \ddot{\mathbf{A}}\boldsymbol{\Omega}\,\mathbf{x} - \frac{1}{n}\mathbf{x}^T\tilde{e}\tilde{e}^T\mathbf{x} \right] (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{R^T} \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \quad \text{(D11b)}$$