A structured model for estimation of automotive paint labour times

X. Luo[1]   P. V. Shevchenko[2]   B. Sayer[3]
W. Blackhall[4]   C. Coelho[5]

Abstract

As the old system known in the smash repairs industry as "funny time, funny money" is being replaced by fairer systems based on empirical evidence, there is lack of rigorous analysis based on observed data and sound statistical methods. This paper proposes and calibrates a statistical model for estimating paint labour times. The model considers terms proportional to paint area such as sanding and spray time, and terms proportional to the perimeter of the panel representing panel masking and other panel boundary related work. Fine details such as 'flash off' time, the number of paint layers and the number of coats per layer are also included into the model. The proposed model can be extended and calibrated for real commercial applications.

Contents

1 Introduction

In 2005, the Australian Government engaged the independent Productivity Commission to investigate the appropriateness and transparency of criteria used by insurance companies to determine the rates paid for smash repair work [1]. Included within the scope of investigating the rates paid, the Productivity Commission also learned of a quoting methodology known as "funny time, funny money". This methodology was based on the premise that industry labour rates remained stagnant over a period of (some) 14 years, whilst costs had increased. Thus, in practice, smash repairers provided inflated repair times (funny time) at the stagnant hourly rate (funny money) to achieve the desired total revenue.

A number of Australian insurers were party to the Productivity Commission's investigation, including Insurance Australia Group (IAG), Suncorp Metway

Limited (Suncorp) and Allianz Australia Insurance Limited (Allianz). Following the Productivity Commission's recommendation that *"Insurers and Repairers should abandon funny time, funny money"* and instead adopt *"Realistic Times and Rates"* [1], IAG developed the "new times and rates" (NTAR) system in collaboration with the Smash Repair industry as an alternative methodology to *"funny time, funny money"*. This methodology includes a calculation of labour time for the paint refinishing process.[1]

This article develops and calibrates a statistical model that estimates paint labour times using experimental data. No vigorous statistical modelling and analysis on paint labour times have previously been reported. After describing the automotive paint process and experimental data, Section 2, the statistical model and its calibration through linear regression are presented in Section 3. This is followed by model fitting results, Section 4, including quantification of uncertainty and comparison of model prediction against the current commercial system NTAR developed by IAG. Section 5 concludes.

2   Paint process and data

2.1   Paint process

This study considers the following painting applications with 'solvent-based' paints:

- *Single-layer paint*: one layer of paint colour with 'hardener';

- *Two-layer paint*: one layer of paint 'colour' and one layer of 'clear';

- *Three-layer paint*: one layer of paint 'colour', one layer of 'effect' (also referred to as 'pearls'), and one layer of 'clear'.

---

[1]See http://ntar.com.au/

Each layer may require several (about two to four) coats. Following industry terminology we refer to the required drying time between paint coats/layers as 'flash off'.

Overall, the paint labour time includes the following major processes: initial setup; sanding; masking; application of 'primer'; application of paint layers; and unmasking and polishing.

2.2   Experimental data

Fourteen experiments covering a total of **123** panels were conducted between 2007 and 2010 by IAG for different car models and panel types. A summary of results for these experiments is shown in Table 1. The details of these data are not in the public domain. To limit any potential bias in data collection, the experiments were performed in twelve different smash repairer premises. A total of **87** hours of film footage was produced, with a 'time clock' overlaid to enable a continuous measurement of time for each experiment. Field notes, transcripts, and observational notes were also produced to complement the film footage, which form the combined body of evidence used in determining appropriate paint labour times.

In general, preparation time (masking, putting on stand, sanding, application of 'primer') is the most time consuming component of the above process in comparison with the actual application of 'colour', 'effect' and 'clear' coats.

Table 1: Experimental data for the paint labour time in hours. J is the number of car panels for each experiment, N_l is the number of paint layers done in each experiment. The unit dm is 10 centimeters.

#	J	N_l	Panel areas (dm^2)	Paint time (h)
1	3	2	211, 41, 41	2.51
2	3	2	156, 52, 52	2.68
3	1	2	416	1.70
4	1	1	144	0.88
5	1	1	144	0.42
6	3	2	48, 184, 94	2.75
7	4	3	170, 51, 51, 128	3.95
8	15	2	168, 45, 45, 149, 89, 89, 79, 79, 64, 64, 167, 58, 58, 114, 173	18.25
9	15	2	144, 48, 48, 158, 106, 106, 80, 80, 61, 61, 168, 31, 31, 106, 180	13.00
10	15	2	84, 84, 109, 109, 320, 29, 6, 6, 68, 68, 9, 9, 2, 2, 2	6.63
11	12	2	48, 90, 38, 81, 33, 72, 22, 34, 20, 20, 19, 20	6.82
12	4	2	120, 58, 9, 81	2.25
13	5	2	168, 164, 75, 45, 45	2.72
14	41	4	160, 61, 9, 210, 98, 61, 9, 110, 48, 3, 3, 76, 110, 48, 3, 3, 76, 13, 13, 26,26, 41, 41, 5, 5, 1, 1, 16, 16, 172, 58, 110, 49, 58, 26, 26, 5, 1, 1, 9, 9	23.06

3   Paint labour time model

3.1   Model structure

Consider the following structured model for paint labour time T of a single panel with the area A:

$$T = \alpha + \beta\sqrt{A} + \gamma \times A + \tau \times A \times N_c + \omega_c \times (N_c - N_l) + \omega_l \times (N_l - 1) + A \times \widetilde{\varepsilon}, \ (1)$$

with the following notation.

- α is the time component independent of panel size (for example, check job sheet, check and setup tools).

- $\beta\sqrt{A}$ is the time required for masking and work on panel edges. Shapes with the same area can have different perimeters, therefore it would be more accurate to use the perimeter instead of $\sqrt{A}$. However, in practice the perimeter is never measured, while the area is conveniently measured by overlaying a transparent sheet over the panel. This term can be removed (after a proper model is fitted) by increasing the gradient for linear terms if a conservative (higher) estimate of paint labour time is acceptable and model simplicity is preferred.

- $\gamma \times A$ is the time required for sanding and application of primer; it is proportional to the area and the same for different paint types.

- $\tau \times A \times N_c$ is the time required to spray all coats (across all layers); here N_c is a total number of coats across all layers. Thus it is dependent on the number of layers. Typically in practice two coats are applied for each layer, that is, $N_c = 2$ for one layer paint; $N_c = 4$ for two layer paint; $N_c = 6$ for three layer paint. Also, τ is a time to spray paint per unit of area for a simple coat that can be estimated separately from other experiments. It might be logical to split this term further into the times for 'colour' coats, 'clear' coats and 'effect' coats if their application times are materially different.

- $\omega_c \times (N_c - N_l)$ is the total flash off time between the coats; here ω_c is flash off time between the coats within a layer, N_l is the number of layers required: $N_l = 1, 2$ and 3 for single-layer, two-layer and three-layer paints respectively.

- $\omega_l \times (N_l - 1)$ is the total flash off time between layers (excluding the flash off after the last layer); ω_l is flash off time between layers.

- $A \times \widetilde{\varepsilon}$ is zero mean random error due to human factors, panel quality and other factors not included in the other formula terms. The error is proportional to the area. We assume that zero mean random variables $\widetilde{\varepsilon}$ are independent for different panels and different experiments, and have the same variance, $\mathrm{var}[\widetilde{\varepsilon}] = \sigma^2$.

Such a model has several advantages. The model formally includes flash off times that can be excluded by removing $\omega_c \times (N_c - N_l)$ and $\omega_l(N_l - 1)$ terms. The structural formula (1) not only puts more clarity in the model but also helps greatly with accurate statistical fitting of the model using experimental data. For example, we can estimate τ separately from experiments, introducing a priori difference in labour times for paint applications with different number of coats; then other model parameters can be estimated from total observed paint labour times. Further refinements can be done by adding extra terms that depend on panel quality/condition. Using (1), the model for time (excluding flash off) to paint J panels with areas $A_1, \ldots, A_J$ is

$$T_{\text{tot}} = \alpha + \beta \sum_{j=1}^{J} \sqrt{A_j} + \gamma \sum_{j=1}^{J} A_j + \tau \times N_c \sum_{j=1}^{J} A_j + \sqrt{\sum_{j=1}^{J} A_j^2} \times \varepsilon, \qquad (2)$$

where ε is zero mean random error independent for different experiments and with the same variance $\mathrm{var}[\varepsilon] = \sigma^2$. That is, the last term in (2) is a random variable representing the sum of independent errors across J panels $A_1 \widetilde{\varepsilon}_1 + \cdots + A_J \widetilde{\varepsilon}_J$. The flash off time was not included into observed paint labour times in Table 1 and thus excluded in (2). Note that α is not multiplied by J, because in general the preparation job (such as checking equipment and setting up tools) is considered to be independent of the number of panels.

Since parameter τ is a time to spray a paint per unit of area for one coat, we estimate it separately using data extracted from experiments. Using observed times (from recorded experiments for painting different panels), we estimate parameters α, β and γ using statistical techniques such as ordinary least square or maximum likelihood methods. The analysis of residuals ε allows assessment of the model appropriateness and statistical statements on predicted values.

3.2 Model fit

We estimate τ, the time to spray paint per unit of area, separately using data extracted from experiments. On average, its estimate $\widehat{\tau} = 0.5\,\mathrm{sec/dm^2}$ with standard deviation $0.2\,\mathrm{sec/dm^2}$. This estimate was based on using a standard gravity feed spray-gun for the application of a solvent-based paint.

Assume K experiments were conducted ($K = 14$ in our case). The total number of panels in the kth experiment is denoted as J_k, with panel areas $A_1^{(k)}, \ldots, A_{J_k}^{(k)}$. Then the total paint labour time for the kth experiment is

$$T_{\mathrm{tot}}^{(k)} = \alpha + \beta R^{(k)} + (\gamma + \tau \times N_c)A^{(k)} + B^{(k)} \times \varepsilon_k \equiv T_{\mathrm{pred}}^{(k)} + B^{(k)} \times \varepsilon_k, \quad (3)$$

where

$$R^{(k)} = \sum_{j=1}^{J_k} \sqrt{A_j^{(k)}}, \quad A^{(k)} = \sum_{j=1}^{J_k} A_j^{(k)}, \quad B^{(k)} = \sqrt{\sum_{j=1}^{J_k} \left(A_j^{(k)}\right)^2},$$

$$T_{\mathrm{pred}}^{(k)} = \alpha + \beta R^{(k)} + (\gamma + \tau \times N_c)A^{(k)},$$

and $\varepsilon_1, \ldots, \varepsilon_K$ are independent with the same variance $\mathrm{var}[\varepsilon_k] = \sigma^2$. Given experimentally observed times $T_{\mathrm{tot}}^{(1)}, \ldots, T_{\mathrm{tot}}^{(K)}$, we formulate a general weighted linear least squares problem with the following chi-square merit function

$$\chi^2(\alpha, \beta, \gamma) = \frac{1}{\sigma^2} \sum_{k=1}^{K} \left(\frac{T_{\mathrm{tot}}^{(k)} - T_{\mathrm{pred}}^{(k)}}{B^{(k)}}\right)^2. \quad (4)$$

Minimising (4) with respect to (α, β, γ) gives their estimates $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$, which are also maximum likelihood estimates if a normal distribution for the random errors ε_k is assumed; Bevington [2] or von Mises [3] detail linear regression methodology. The variance of errors $\mathrm{var}[\varepsilon_k] = \sigma^2$ is estimated as a sample variance of $\varepsilon_1, \dots, \varepsilon_K$ calculated from (3) using $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ estimates. A commonly used statistic in association with weighted linear regression is the coefficient of determination

$$R^2 = 1 - \frac{\sum_{k=1}^{K} w_k (y_k - f_k)^2}{\sum_{k=1}^{K} w_k (y_k - \bar{y})^2} , \tag{5}$$

where f_k is the model prediction, y_k is the observation, $\bar{y} = \frac{1}{K}(y_1 + \cdots + y_K)$ and $1/w_k$ is the variance in the kth experiment. For our model, these are $f_k = T_{\mathrm{pred}}^{(k)}$, $y_k = T_{\mathrm{tot}}^{(k)}$ and $1/w_k = (\sigma B^{(k)})^2$. The coefficient of determination R^2 gives some indication about the goodness of fit of a model. In regression, R^2 is a statistical measure of how well the regression model approximates the real data points. It is related to the unexplained variance, since the second term in (5) compares the unexplained variance (variance of the model's errors) with the total variance of the data. A larger R^2 indicates a better fit.

The adjusted R^2, denoted as $\bar{R}^2$, is a modification of R^2 that adjusts for the number of explanatory terms in a model. Unlike R^2, $\bar{R}^2$ increases only if the new term improves the model more than would be expected by chance. It is formally defined as

$$\bar{R}^2 = 1 - (1 - R^2)\frac{K - 1}{K - p - 1}, \tag{6}$$

where p is the total number of regressors in the linear model and K is the sample size; in the studied case $p = 3$ and $K = 14$.
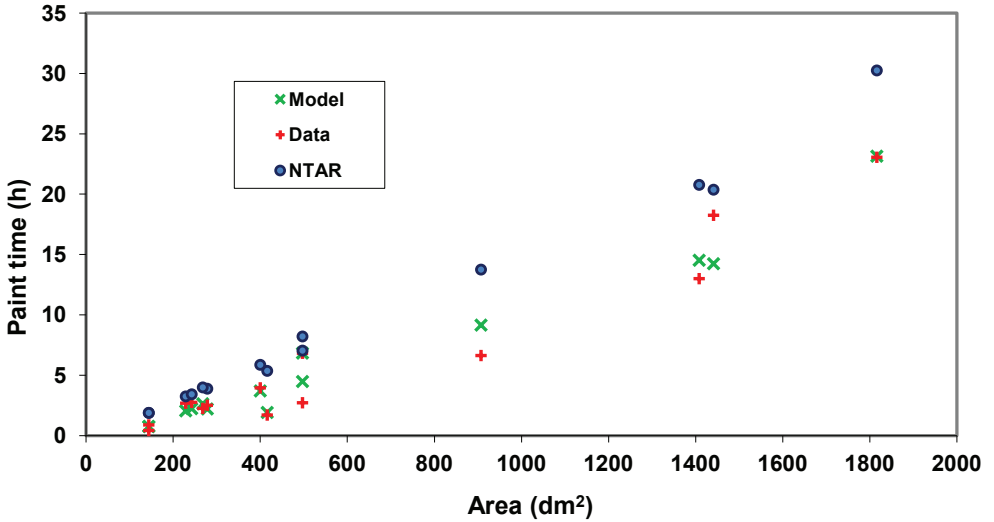
Figure 1: Comparison of model predictions, observed times and NTAR system labour times for the 14 experiments.

4 Results

This section shows model fitting results and compares them with the paint labour times from the NTAR system.[2]

The input data for panel areas and paint labour times required for the linear regression problem are summarised in Table 1. For the number of coats N_c, we assume three coats per layer. That is, three coats for a one-layer paint, six coats for a two-layer paint and nine coats for a three layer paint. With these data, a general linear least squares problem (4) was solved and the parameter estimates found to be $\widehat{\alpha} = -0.448$, $\widehat{\beta} = 0.0896$, $\widehat{\gamma} = 0.000453$ and $\widehat{\sigma} = 0.0043$.
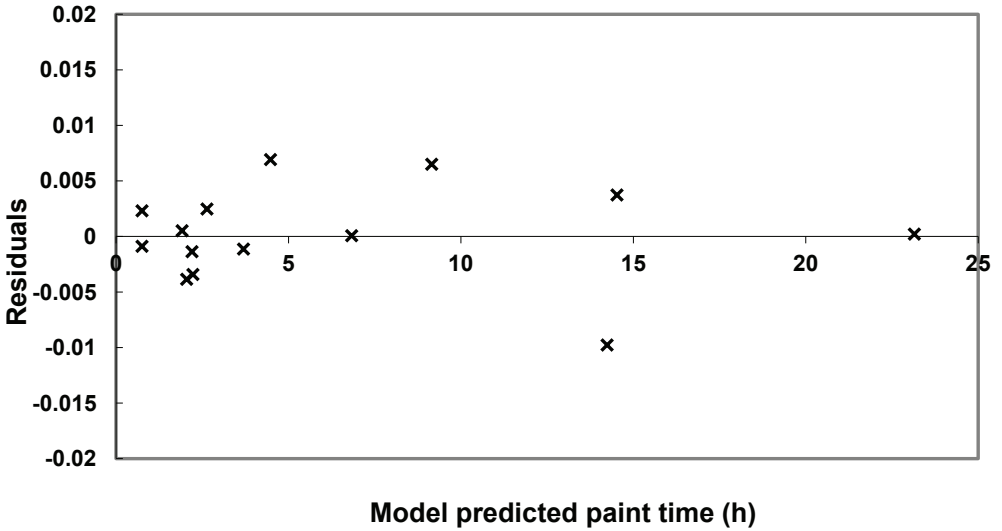
---

[2]See http://ntar.com.au/

Figure 2: Plot of the residuals ε_k versus model predicted paint labour time $\mathsf{T}_{\mathrm{pred}}^{(\mathrm{k})}$.

Figure 1 compares model predictions with observed times and labour times from the NTAR system, corresponding to the panel configurations of the experiments (given in Table 1). The R^2 and $\bar{\mathsf{R}}^2$ values for the model fit are 0.968 and 0.959 respectively, indicating a good fit. Figure 2 plots the model residuals ε_k versus the model predicted paint labour time $\mathsf{T}_{\mathrm{pred}}^{(\mathrm{k})}$ that appear to be scattered randomly around zero indicating a good fit too.

Figure 3 compares the model prediction with the labour times from the NTAR system for a single panel. In this case paint time is a unique function of the panel area. The prediction by a reduced model (it is discussed below) is also shown in Figure 3. The reduced model has fewer parameters but can still fit the data well.

The negative α does not make physical sense, but gives a good fit between model and data in this case, as shown in Figure 1. The negative α value might
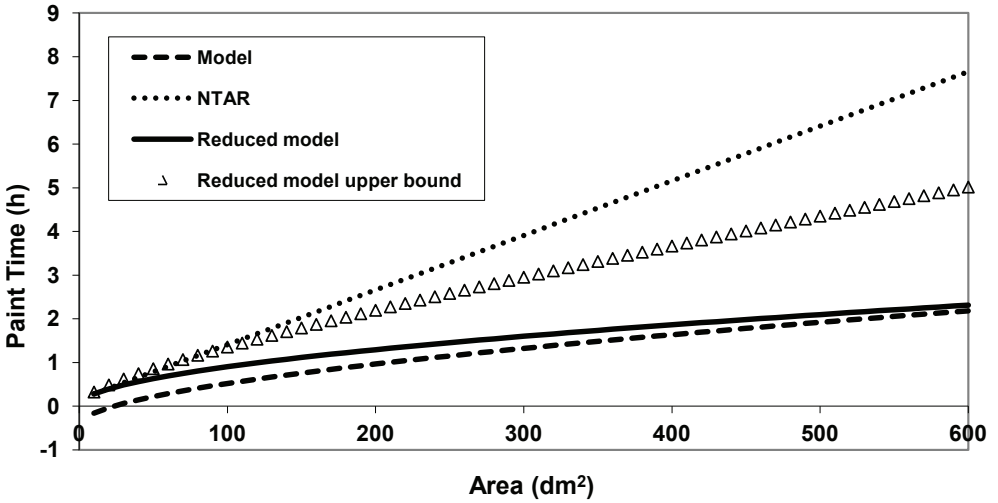
Figure 3: Comparison of the model predictions for a single panel. "Model" is the full model (3); "NTAR" is NTAR system results; "Reduced model" is the model (3) with $\alpha = \gamma = 0$; and "Reduced model upper bound" is labour time prediction from the reduced model plus one standard deviation of the random term, that is, $T_{\text{pred}}^{(k)} + B^{(k)}\sigma$.

be due to that none of the experimental data have total areas close to zero, thus α cannot be determined accurately. We can overcome this drawback by minimizing with the constraint $\alpha \geqslant 0$. However, if we are interested in a conservative (higher) estimate, we simply set $\alpha = 0$ after a negative value is obtained from regression. The comparison in Figure 3 shows model predictions both with and without setting $\alpha = 0$.

On the other hand, we can take a formal parameter selection approach to determine the most appropriate set of parameters in the model. For example, if we simply drop parameters α and γ (because corresponding terms are small) and fit the reduced model to the data, we find $\widehat{\beta} = 0.0876$ and $\widehat{\sigma} = 0.0045$. The R^2 and $\bar{R}^2$ values for the reduced model fit are 0.964 and 0.961 respectively.

The adjusted R^2 value has slightly increased after dropping two parameters α and γ. Surprisingly we can drop γ without materially affecting the goodness of fit for the reduced model, since the γ term accounts for paint labour times proportional to panel areas. However, the term $\tau \times N_c A^{(k)}$ in (3) is also proportional to panel area and it is the dominant factor in comparison to the γ term. This is apparent when the value of $\tau \times N_c$ is compared with γ value (fitted for the full model): $\tau \times N_c = 0.00083$ and $\widehat{\gamma} = 0.000453$. As shown in Figure 3, the reduced model is slightly more conservative than the full model for a single panel, based on the model fit with the current set of 14 experiments.

Since both α and γ can be dropped without material impact on the model prediction, it is apparent the dominate term is the β term. The standard error for $\widehat{\beta}$ estimated from the linear regression is $\widehat{\sigma}_\beta = 0.0047$, which is approximately 5% of the estimated β value. The labour time curve formed by replacing β with $\widehat{\beta} + 2\widehat{\sigma}_\beta$ is visually the same as the curve for the labour time predicted from the reduced model in Figure 3. The impact of parameter uncertainty is significantly less than the impact of the process uncertainty. The latter comes from the model error terms $B^{(k)}\varepsilon$. To demonstrate the magnitude of the process uncertainty, "reduced model upper bound" is calculated as the labour time prediction from the reduced model plus one standard deviation of the random term, that is, $T_{\text{pred}}^{(k)} + B^{(k)}\sigma$, and presented in Figure 3. The curve formed by adding two standard deviations $T_{\text{pred}}^{(k)} + 2B^{(k)}\sigma$ is almost the same as the "NTAR" curve in Figure 3.

5   Conclusions

This article proposes a structured model to predict paint labour times based on observational data recorded from 14 case studies totalling 123 paint applications. This statistical model considers times proportional to the panel area, such as sanding and paint spray, and times proportional to the perimeter, such as masking and un-masking. Fine details such as flash off time, the

number of paint layers and the number of coats per layer are also included in the model. In addition, the model has an uncertainty term built-in which we quantify to cater for conservatism typically required by the insurer in order to have a robust margin for a fair payment system. All the model parameters have clear and intuitive interpretations and the model is parsimonious. We can extend the model to account for waterborne paint or flash off times. Results show that the NTAR system labour times are conservative when compared to the actual observations or compared to the model predictions.

References

[1] Australian Government. Smash repairers and insurers. *Productivity Commission Inquiry Report No. 34*, 17 March, 2005. C423, C424

[2] P. R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill. Chapter 1–4. 1969. C430

[3] R. Von Mises. *Mathematical Theory of Probability and Statistics*. New York: Academic Press, 1964. C430

Author addresses

1. **X. Luo**, CSIRO Mathematics, Informatics and Statistics, Sydney, Locked Bag 17, North Ryde, NSW 1670, AUSTRALIA.
   mailto:Xiaolin.Luo@csiro.au

2. **P. V. Shevchenko**, CSIRO Mathematics, Informatics and Statistics, Sydney, Locked Bag 17, North Ryde, NSW 1670, AUSTRALIA.
   mailto:Pavel.Shevchenko@csiro.au

3. **B. Sayer**, Insurance Australia Group, 388 George Street, Sydney NSW 2000, AUSTRALIA.

4. **W. Blackhall**, Insurance Australia Group, 388 George Street, Sydney NSW 2000, Australia.

5. **C. Coelho**, Insurance Australia Group, 388 George Street, Sydney NSW 2000, Australia.