Improving the Accuracy of Time-Driven Activity-Based Costing by Stochastic Modelling

Lea Vermeire[1,4], Sophie Hoozée[2] and Werner Bruggeman[3]
[1]KU Leuven Kulak, Kortrijk, BELGIUM
[2]Erasmus University, Rotterdam, THE NETHERLANDS
[3]Ghent University, Gent, BELGIUM
[4]Corresponding author: Lea Vermeire, e-mail: lea.vermeire@kuleuven-kulak.be

Abstract

A recent technique in cost modelling for an activity in a company, such as the delivery of a product or a service, is time-driven activity-based costing. The time for an activity is modelled in a time equation as the cumulative time effect of a number of included drivers (the partial time) and an error term for omitted drivers (the residual time). The estimated time for a transaction – that is a particular occurrence of an activity – is the estimated partial time, where the time per driver unit is obtained from interviews with personnel, and the driver volume or number of units in a driver is reported by the company computer system. An analysis of the error in the estimated partial time as predictor of the true transaction time is given, as well as the impact on this error when the time equation is refined by adding a time-driver. On this base, a design protocol to obtain an optimal time equation is suggested, and a prediction interval for the true transaction time from the estimated partial time is given.

Keywords: cost accounting model, costing errors, error index, mean squared error, prediction interval, time equation.

1. Introduction

In cost accounting, heuristics are used to estimate the cost of manufacturing a product or delivering a service. One of these heuristics, the method of activity-based costing (ABC) (Kaplan & Bruns, 1987, Cooper & Kaplan, 1988), showed some shortcomings in accuracy (Datar & Gupta, 1994), which stimulated the development of time-driven activity-based costing (TDABC) (Kaplan & Anderson, 2004). In TDABC, cost estimation is linked to time, i.p. to time units for an event's resources. For example, a specific delivery at a warehouse consists of 20 manual units and 3 layers of equal units, all on a pallet that needs to be wrapped up, plus 2 full pallets. Its accounting time is computed from a time equation: $\textit{time for picking the delivery} = 5.0 + 0.1 \times (\textit{number of units to be picked manually}) + 0.2 \times (\textit{number of layers}) + 1.0 \times (\textit{number of full pallets}) + (3.0 \times (\textit{number of pallets to wrap up}) = 5.0 + 0.1 \times 20 + 0.2 \times 3 + 1.0 \times 2 + 3.0 \times 1 = 12.6$, where 5.0, 0.1, etc. are unit time estimates, while some time drivers may be omitted (e.g., the time effect of checking each layer on the pallet). Thus, this accounting time is an approximation of the true time for picking the delivery. We present a mathematical approach to capture the stochastic aspects – measurement error and structural variance – of TDABC, in order to improve the accuracy of the system. The paper is a synthesis of our former study (Hoozée e.a., 2012), and includes a warning on the prediction problem.

2. Problem and model

Activity and transaction. In general, an *activity* (or *cost object*) is an economic operation, like picking a delivery, and is a succession of non-overlapping subactivities, like manual picking, layer picking, wrap-up, etc. A *transaction* is a particular occurrence of an activity, like the above specific delivery, and is similarly a succession of subtransactions, like manually picking 20 units, etc. A *time equation* is a mathematical equation that expresses the time required to perform an activity as a function of several time drivers.

True time. The *true time* of a random transaction is the cumulative time of a basic time and the times of the subtransactions, expressed by the *time equation*

$$t = \beta_0 + \beta_1\xi_1 + \ldots + \beta_k\xi_k + \varepsilon_k = t_k + \varepsilon_k, \quad t_k = \beta_0 + \beta_1\xi_1 + \ldots + \beta_k\xi_k \ , \tag{1}$$

with

t : *true* time for a particular transaction;

β_0 : basic time for the activity, treated as a within activity constant, i.e. constant over the transactions, and independent of k;

β_i : true *unit time* for one unit of driver i, treated as constant within the activity;

ξ_i : *driver volume* or number of units of driver i in the transaction, a within activity random variable, say with mean and variance $\mathrm{E}(\xi_i) = x_i, \ \mathrm{V}(\xi_i) = \kappa_i^2$;

$\tau_i = \beta_i\xi_i$: true time contribution of driver i in the transaction;

t_k : approximation of the true transaction time t based on the inclusion of k observable drivers ($k \geq 1$, a fixed integer), a *partial time* for the transaction;

ε_k : *residual time effect* of the omitted drivers, as some drivers may be unknown or too costly to observe in practice.

For obvious reasons all $\beta_j, \ \xi_i, \ \varepsilon_k$ are non-negative; $\varepsilon_k = 0$ if all drivers are included. The ξ_i and hence $t, \ t_k, \ \varepsilon_k$ are constants for a particular transaction, but are within activity random variables, as they change stochastically from transaction to transaction (e.g. the number of units to be picked manually may be 20 for one transaction and 25 for another transaction). In practice, both, independent and correlated drivers are frequently met (e.g. it is likely that the number of order lines ξ_1, and the number of pickings in the warehouse ξ_2, are positively correlated, while ξ_1 and the number of reusable containers ξ_3, are independent drivers).

Estimated time. The *estimated transaction time* based on k drivers is

$$\hat{t}_k = B_0 + B_1X_1 + \ldots + B_kX_k \ , \tag{2}$$

with

B_0 : estimated basic time, given by the personnel;

$B_i = \beta_i + b_i$: estimation for one unit time β_i of driver i, given by the personnel, a within activity constant; thus its *bias* b_i is also a within activity constant;

X_i : estimate of the driver volume ξ_i, as reported by the computer system, a random variable subject to random error. Within a transaction, which is summarized as $\xi = (\xi_1, \ldots, \xi_k, \varepsilon_k)$, the error has mean 0 and variance ν_i^2, thus $\mathrm{E}(X_i|\xi) = \xi_i, \ \ \mathrm{V}(X_i|\xi) = \mathrm{V}(X_i - \xi_i|\xi) = \nu_i^2,$ where ν_i is a within activity constant for driver i. Given the errors on the counts are random, then the $X_i|\xi$ are independent random variables; also each pair $X_i - \xi_i, \xi_i$ is independent.

$T_i = B_iX_i$: estimated time contribution of driver i.

Defining $X_0 \sim \xi_0 \sim U\{1\}$, the above time equations for the true time of a transaction and the estimated partial time can be written in a homogeneous structure, now with drivers $i = 0, 1, \ldots, k$,

$$t = \sum_{i=0}^{k} \beta_i \xi_i + \varepsilon_k, \qquad \hat{t}_k = \sum_{i=0}^{k} B_i X_i \ . \tag{3}$$

Error structure and variance-covariance structure. The estimator $\hat{t}_k$ for t has error and variance. A basic error decomposition

$$t = t_k + \varepsilon_k = \hat{t}_k - (\hat{t}_k - t_k) - (t_k - t) \tag{4}$$

explicites the error as an aggregation of two types of error: *identification error (model error)*, $t_k - t = -\varepsilon_k$, due to the omission of drivers, and *estimation error (measurement error)*, $\hat{t}_k - t_k$, caused by estimation of the time parameters and the driver volumes in the partial time; these two errors are independent. The variance in the estimator $\hat{t}_k$ covers variability from two natures: *error variance* due to estimation error in the driver volumes, and *structural variance* due to within activity stochastic variation of the driver volumes ξ_i. Special cases of interest are:

the *unbiased case*: the case of accurate unit time estimates, i.e. all $b_i = 0$;
the *accurate driver volumes case*: all $X_i = \xi_i$, i.e. $\nu_i = 0$;
the *independence case*: the case of independent drivers, all drivers are independent in their time effects, thus all random variables $\xi_1, \ldots, \xi_k$ and ε_k are independent; then the X_i are independent as well;
the *inclusion-residual uncorrelated case*: in $t = t_k + \varepsilon_k$ the pool of k included drivers and the pool of residual drivers are uncorrelated, i.e. t_k and ε_k are uncorrelated; it implies that correlated drivers are pooled for inclusion/exclusion in the time equation, which should be realistic in practice;
the *inclusion-residual independent case*: t_k and ε_k are independent;
the *addition-residual uncorrelated case*: in $t = t_{k-1} + \tau_k + \varepsilon_k$ the added driver is assumed not to be correlated with the residual pool of omitted drivers, i.e. $\tau_k = \beta_k \xi_k$ and ε_k are uncorrelated.

Problem. Design a protocol to select an optimal set of drivers, and give a prediction interval for the true transaction time t from the estimated partial time $\hat{t}_k$. In a preparative study, analyse the error of this estimated partial time for the true transaction time, and the effect of refinement on this error by adding a time driver.

Methods. *MSE accuracy*: as measure of accuracy for the estimator $\hat{t}_k$ for t, we use the *mean squared error*, $\mathrm{MSE}(\hat{t}_k; t) = \mathrm{E}(\hat{t}_k - t)^2$. The variance and the bias or mean error or denoted $\mathrm{V}(\hat{t}_k)$, $B(\hat{t}_k; t) = \mathrm{E}(\hat{t}_k - t) = E\hat{t}_k - Et$.

Computational techniques: standard properties for the MSE; conditional computing for the transition from within transaction accuracy to within activity accuracy; suitable sources are a.o. Mittelhammer (2013), Rice (2007), Wackerly e.a. (2008).

Assumptions: whenever moments of random variables are used in the paper – typically moments up to order 2 – it is assumed that they exist and are finite.

3. Results for the error analysis

This section lists mathematical expressions, successively for the identification accuracy and for the estimation accuracy, which are the base of the design protocol and the prediction. Proofs and extensive interpretations are in Hoozée e.a. (2012).

Total error. The accuracy of the estimated transaction time for the true transaction time, using (4), is expressed in:

$$\begin{aligned}\mathrm{MSE}(\hat{t}_k;t) &= \mathrm{MSE}(t_k;t) + \mathrm{MSE}(\hat{t}_k;t_k) + 2\ B(t_k;t)\ B(\hat{t}_k;t_k) \\ &= \mathrm{V}(t_k - t) + \mathrm{V}(\hat{t}_k - t_k) + \left[\mathrm{B}(t_k;t) + \mathrm{B}(\hat{t}_k;t_k)\right]^2 . \qquad (5)\end{aligned}$$

Identification accuracy. For the partial transaction time t_k as approximation of the true transaction time t, the accuracy and the effect of refinement by adding driver k (or the transition $t_{k-1} \mapsto t_k = t_{k-1} + \tau_k,\ \ \tau_k = \beta_k \xi_k$), can respectively be written:

$$\mathrm{MSE}(t_k;t) = \mathrm{V}(t) - \mathrm{V}(t_k) + (Et - Et_k)^2 - 2\ \mathrm{C}(t_k, \varepsilon_k)\ ; \qquad (6)$$

$$\begin{aligned}\Delta\mathrm{MSE}(t_k, t_{k-1};t) &:= \mathrm{MSE}(t_{k-1};t) - \mathrm{MSE}(t_k;t) \\ &= \mathrm{E}(\tau_k^2) + 2\ \mathrm{E}(\tau_k\,\varepsilon_k) \\ &= \mathrm{V}(\tau_k) - (E\varepsilon_{k-1} - E\tau_k)^2 + (E\varepsilon_{k-1})^2 + 2\,\mathrm{C}(\tau_k, \varepsilon_k)\ . \qquad (7)\end{aligned}$$

A *sufficiency index (accuracy index)* of the pool of included drivers is defined as:

$$\kappa = 1 - \left[\frac{\mathrm{MSE}(t_k;t)}{\mathrm{E}(t^2)}\right]^{1/2} . \qquad (8)$$

In concept this sufficiency index is a relative measure of accuracy for the approximation t_k for t, namely the complement of the *error index*, which is the root relative MSE w.r.t. the maximum MSE. It satisfies $0 \le \kappa \le 1$, and is closer to 1 as the included pool is MSE better. A pool of included drivers may be considered sufficient as soon as $\kappa \ge \kappa_\star$ for some chosen *accuracy level* $\kappa_\star$, $0 < \kappa_\star < 1$ (e.g. $\kappa_\star = 90\%$).

Accuracy of the estimated partial transaction time. For the estimator $\hat{t}_k$ for t_k, the within transaction accuracy, the within activity accuracy, and the effect of refinement by adding driver k (the change in MSE for the transitions $t_{k-1} \mapsto t_k = t_{k-1} + \tau_k,\ \ \hat{t}_{k-1} \mapsto \hat{t}_k = \hat{t}_{k-1} + T_k$, with $\tau_k = \beta_k \xi_k,\ \ T_k = B_k X_k$), can respectively be written:

$$\mathrm{MSE}\left[\ (\hat{t}_k;t_k)|\xi\ \right] = b^2 + v^2\ , \qquad (9)$$

$$\mathrm{MSE}(\hat{t}_k;t_k) = \mathrm{E}(b^2) + v^2\ , \qquad (10)$$

$$\begin{aligned}\Delta\mathrm{MSE}(\hat{t}_{k-1}, t_{k-1};\hat{t}_k, t_k) &:= \mathrm{MSE}(\hat{t}_k;t_k) - \mathrm{MSE}(\hat{t}_{k-1};t_{k-1}) \\ &= \mathrm{MSE}(T_k;\tau_k) + 2\,\mathrm{E}(\hat{t}_{k-1} - t_{k-1})(T_k - \tau_k) \\ &= B_k^2\ \mathrm{V}(X_k|\xi) + b_k^2\ \mathrm{E}(\xi_k^2) + 2\,b_k \sum_{\ell=0}^{k-1} b_\ell\ \mathrm{E}(\xi_\ell \xi_k)\ , \qquad (11)\end{aligned}$$

with $b = \sum b_i \xi_i\ ,\quad v^2 = \sum\ B_i^2\ \nu_i^2 = \sum\ (\beta_i + b_i)^2\ \nu_i^2$.

Prediction of the true transaction time from the partial time. Given a chosen *confidence level* $1-\alpha$, $0 \leq \alpha < 1$. Then a $(1-\alpha)$ *prediction interval* (PI) for the true transaction time t, given the estimated partial time $\hat{t}_k$, is a numerical interval $[L, R]$ such that

$$\mathrm{P}\big[L \leq (t|\hat{t}_k) \leq R\big] \geq 1-\alpha\ . \tag{12}$$

4. A protocol for an optimal time equation

Given, an accuracy level $\kappa_\star$ for the time equation, how can we obtain an optimal time equation, i.e. how can we select a sufficient set of drivers, with as few drivers as possible? The accuracy analysis above and a cost efficient purposive collection of time data lead to the following solution.

(1) *Start with a set of relevant drivers.* Guidelines suggested by close interpretation of the error expressions (6) and (7):
pool highly correlated drivers and intend to include/exclude them jointly;
try to absorb most of the mean ànd the variance of t;
for equivalent sets of drivers, i.e. sets which absorb the same mean and the same variance, prefer the high correlation (t_k, t);
prefer drivers with accurate unit times and accurate driver volumes or with accurate estimates affordable by a pilot study.

(2) *Check if this set of k drivers is sufficient.* Do a pilot study (PS1): N accurate observations of (t, t_k), and hence of the error $\varepsilon_k = t - t_k$, provide data $t_j,\ t_{k,j},\ \varepsilon_{k,j} = t_j - t_{k,j}$. Compute the empirical estimates $\mathrm{MSE}(t_k;t) \mathrel{\mathring{=}} \frac{\sum_j \varepsilon_{k,j}^2}{N}$, $\mathrm{E}(t^2) \mathrel{\mathring{=}} \frac{\sum_j t_j^2}{N}$, $\kappa = 1 - \left[\frac{\mathrm{MSE}(t_k;t)}{\mathrm{E}(t^2)}\right]^{1/2}$. The set of drivers is sufficient if $\kappa \geq \kappa_\star$; otherwise add more drivers.

5. Construction of a prediction interval

Given a confidence level $1-\alpha$, find a prediction interval for the true time t of a transaction in an activity, from a particular estimated partial time $\hat{t}_k$ provided by a – preferably optimal – time equation.

Case 1: Accurate partial time, and inclusion-residual independence. Thus $t_k \approx \hat{t}_k$, t_k and ε_k are independent. Then $(t|\hat{t}_k) = (t|t_k) = t_k + (\varepsilon_k|t_k) \sim t_k + \varepsilon_k$. A generic PI follows from definition (12):

$$[L;R] = [\,t_k + \varepsilon_{k,p}\ ;\ t_k + \varepsilon_{k,q}\,], \quad q = p + (1-\alpha), \tag{13}$$

where quantiles $\varepsilon_{k,\gamma}$ (with left probability γ) should be estimated. Empirical estimates or model-based estimates can be used, as shown in the following examples.

(a) No information on the residual time distribution is available: empirical quantiles from the above pilot PS1 give a solution, $\varepsilon_{k,p} \mathrel{\mathring{=}} \hat{\varepsilon}_{k,p} = \varepsilon_{k,<\lfloor (N+1)p \rceil>}$.

(b) Normal residual time $\varepsilon_k \sim N[\,\mathrm{E}(\varepsilon_k), \mathrm{V}(\varepsilon_k)\,]$.
Estimate the model parameters $\mathrm{E}(\varepsilon_k)$, $\sigma(\varepsilon_k) = \sqrt{\mathrm{V}(\varepsilon_k)}$ from the pilot PS1 data.
Empirical check of the normality: $\mathrm{E}(\varepsilon_k) \gtrsim 2\,\sigma(\varepsilon_k)$, QQ-plot.
Prediction interval: $t = t_k + \mathrm{E}(\varepsilon_k) \pm z_{1-\alpha/2}\,\sigma(\varepsilon_k)$.

(c) Halfnormal residual time $\varepsilon_k \sim HN(0, \lambda^2)$, $\mathrm{E}(\varepsilon_k) = \sqrt{2/\pi}\,\lambda$.
Empirical check, on the PS1 data: $\sigma(\varepsilon_k) \approx (3/4)\,\mathrm{E}(\varepsilon_k)$.
Prediction interval: $[L, R] = \big[\ t_k\ ;\ t_k + 1.25\ z_{1-\alpha/2}\ \mathrm{E}(\varepsilon_k)\ \big]$.

Case 2: Acurate driver volumes, and inclusion-residual independence. The only errors are the unit time biases b_i. We recommend a correction for the unit time biases, by accurate estimation of the true unit times β_i. Either internal estimation: as empirical estimates from a detailed pilot study (PS2) of accurate observations $\beta_{i,j}$. Either external estimation: as Cardinaels and Labro (2008) arrived empirically at an overestimation of 37% in the reported unit times, the correction $\beta_i = (1/1.37)\, B_i$ may be proposed. Then return to case 1.

Case 3: General case. General measurement errors and inclusion-residual dependence is possible. Writing $t = \hat{t}_k + \hat{\varepsilon}_k$, the term $\hat{\varepsilon}_k$ is the total error variable. The definition equation (12) leads to the generic prediction interval

$$[L, R] = \left[\, \hat{t}_k + (\hat{\varepsilon}_{k,p} | \hat{t}_k) \ ; \ t_k + (\hat{\varepsilon}_{k,q} | \hat{t}_k) \,\right], \quad q = p + (1 - \alpha), \tag{14}$$

which needs estimates of conditional quantiles $(\hat{\varepsilon}_{k,\gamma} | \hat{t}_k)$. We should be aware that $\hat{\varepsilon}_k$ and $\hat{t}_k$ are not stochastically independent; even in the case of accurate unit times and inclusion-residual independence, there is covariance $\mathrm{C}(\hat{\varepsilon}_k, \hat{t}_k) = -\mathrm{V}(\delta_k),\ t_k = \hat{t}_k + \delta_k$. A pilot study of accurate data for $(t, t_k, \hat{t}_k)$ provides data for $(\hat{t}_k, \hat{\varepsilon}_k),\ \hat{\varepsilon}_k = t - \hat{t}_k$, and should allow empirical estimates for the conditional quantiles.

Conclusion

The paper contributes to the improvement of a time-driven activity-based costing system in four aspects: a mathematical description to elucidate the stochastics of the time equation, an accuracy analysis, a design protocol to arrive at a time equation with prescribed accuracy level, an outline to construct a prediction interval for the true time cost based on the estimated time from the time equation. To grow into maturity, this theoretical study may be further explored, and complemented by a simulation study and a field experiment, and by a software implementation.

References

Cardinaels, E. and Labro, E. (2008) "On the Determinants of Measurement Error in Time-Driven Costing", *The Accounting Review*, 83(3), 735-756.

Cooper, R. and Kaplan, R. (1988) "Measure Costs Right: Make the Right Decisions", *Harvard Business Review*, 66(5), 96-103.

Datar, S. and Gupta, M., (1994) "Aggregation, Specification and Measurement Errors in Product Costing", *The Accounting Review*, 69(4), 567-591.

Hoozée, S., Vermeire, L. and Bruggeman, W. (2012) "The Impact of Refinement on the Accuracy of Time-Driven ABC." *Abacus*, 48(4), 439-472.

Kaplan, R. and Bruns, W. (1987) *Accounting and Management: A Field Study Perspective.* Harvard Business School Press.

Kaplan, R., Anderson, S. (2004) "Time-Driven Activity-Based Costing," *Harvard Business Review*, 82(11), 131-138.

Mittelhammer, R. (2013) *Mathematical Statistics for Economics and Business.* 2nd ed., Springer, New York.

Rice, J. (2007) *Mathematical Statistics and Data Analysis*, 3rd ed., Thomson / Duxbury, Belmont (Ca).

Wackerly, D., Mendenhall III, W. and Scheaffer, R. (2008) *Mathematical Statistics with Applications*, 7th ed., Brooks/Cole, Belmont (Ca).