A method for forecasting owner monthly construction project expenditure flow

Skitmore, Martin

Published in:
International Journal of Forecasting

DOI:
10.1016/S0169-2070(97)00042-3

A METHOD FOR FORECASTING OWNER MONTHLY CONSTRUCTION PROJECT EXPENDITURE FLOW

Professor Martin Skitmore
School of Construction Management
Queensland University of Technology
Gardens Point
Brisbane Q4001
Australia

Tel: (07) 3864 2234
Fax: (07) 3864 1170
E-mail: RM.Skitmore@qut.edu.au

Biographical sketch

Martin Skitmore MSc PhD FRICS MCIOB FAIB AAIQS is Professor and Head, School of Construction Management and Property, Queensland University of Technology. His research interests include estimating, bidding, cost modelling and the application of information technology to economic decision making. He is author of many papers on construction economics and forecasting including the book *Contract Bidding in Construction: Strategic Management and Modelling* (Longman) and *The Accuracy of Construction Price Forecasts: a Study of Quantity Surveyors' Performance in Early Stage Estimating* (Salford University). He is Co-editor-in-Chief of *The International Journal of Construction Information Technology* and a member of several journal editorial boards. He is also a member of the International Council for Building Research, Studies and Documentation Working Commission W-55 Building Economics and W-65 Construction Management and a former member of both the Royal Institution of Chartered Surveyors and Chartered Institute of Builders Research Committees.

A METHOD FOR FORECASTING OWNER MONTHLY CONSTRUCTION PROJECT EXPENDITURE FLOW

Abstract: Under the normal conditions of construction contracts, the client is obliged to pay the contractor in monthly instalments. The amount of each instalment is based on the value of construction work actually produced in the previous month and forecasts are needed in advance of the likely value of these payments. A database of previously completed contracts and payments made is available.

A method for forecasting the value of these instalments is described. This method utilises three approaches, termed (1) analytic, (2) synthetic, (3) hybrid, in combination with six alternative models comprising (1) Hudson, (2) Kenley-Wilson, (3) Berny-Howes, (4) cumulative logistic, (5) cumulative normal, and (6) cumulative lognormal. The forecasts produced by each of these are then subject to a cross-validation analysis to determine the best approach/model combination for the available database and hence forecasts for future expenditure flows.

An example is provided for an actual 27 construction project database.

Key Words: Construction contracts, expenditure flow models, forecasting system, regression model, cross validation, time series.

1. Introduction

Under the normal conditions of construction contracts, the client is obliged to pay the contractor in monthly instalments. The amount of each instalment is based on the value of construction work actually produced in the previous month and forecasts are needed in advance of the likely value of these payments. A database is available of previously completed contracts and payments made, by the clients, to the contractors involved. These forecasts may be needed to be made prior to the commencement of the construction work or after some instalments have already been paid in order to provide sufficient indication to the client of his likely future expenditure commitment over the whole or remainder of the contract.

Previous research suggests that the cumulative frequency distribution of the payments over time is sigmoidal and several alternative mathematical models have been proposed (e.g. Hudson, 1978; Berny & Howes, 1982; Tucker, 1986, 1988; Kenley & Wilson, 1989; Miskawi, 1989; Khosrowshahi, 1991). Comparisons between some of these have been made in a recent work by Skitmore (1992) which tests their accuracy in situations where forecasts are needed prior to the commencement of any construction work.

In this paper we consider the case of a contract where some instalments have already been paid, and estimates of the value of future payments are needed to provide an indication of the series of client expenditures. Hudson's (1978) approach to this is to use the values of the two most recent payments, after smoothing by polynomial regression, to estimate the parameter values of Hudson's model. Here, we propose a method which utilises three techniques, termed here analytic, synthetic and hybrid. Each technique is applied to six models comprising (1) Hudson, (2) Kenley-Wilson, (3) Berny-Howes, (4) cumulative logistic, (5) cumulative normal, and (6) cumulative lognormal. The forecasts produced by each of these are then subject to a cross-validation analysis to determine the best approach/model combination for the available database and hence forecasts for future expenditure flows.

An example is provided for an actual 27 construction project database.

2. Problem definition

Assume that over the agreed-in-advance **contract duration** of a construction project there have been n valuations of work completed (instalments), each at time point T=1,...,n. Consider an arbitrary intermediate point in time, say t. There will have been some valuations done before t and some valuations done after. Letting p be the number of valuations done before t, then the number of valuations done after t will be n-p. Now let v_T (t=1,2,...,p) be the first, second, etc. valuations before t^o and v_t (t=p+1,p+2, ... ,n) be the first, second, etc. valuation after t^o expressed as the cumulative percentage of the agreed-in-advance **contract value** of the project. For each v_t there will be an associated duration, d_t, expressed as a cumulative percentage of the agreed-in-advance contract duration of the project.

The use of **actual** contract values and durations is not considered here for two reasons (1) for the target contract, these are not known and would therefore have to be estimated - this introduces two further parameters into the problem, (2) it is considered undesirable by many consultants to commence a contract with forecast overruns on duration or expenditure - instead an estimated shortfall of expenditure is preferred at the contract completion. Of course, expenditure forecasts may be made beyond the contract completion date if required by simply extending the number of valuations to n+1, n+2, ...

The problem to be addressed, therefore, is how best to forecast the values of v_t (t=p+1,p+2,...,n) given the values of v_t (t=1,2,...,p) and the type and size of contract involved.

3. Analysis

Two basic types of analysis are used, which we term (1) analytic and (2) synthetic. The analytic analyses involve the analysis of data for a set of previous contracts. The synthetic analyses involve extrapolating from previous valuations for the current contract.

3.1. Analytic method

The analytic method relies on the analysis of historical data concerning completed contracts. It does not depend upon any valuation data concerning the target contract and is therefore appropriate in **pre-contract** forecasting. It is assumed that the shape of the expenditure curve used to model the data is some function of the characteristics (e.g. type and size) of the contracts providing the data. As the shape of the curve is completely described by at least two parameters, attention is focused on the function connecting the contract characteristics and the parameters. Here we use Skitmore's (1992) method to estimate the parameters (method 4 in Appendix B).

The advantages of this approach are that forecasts can be made from time point zero right to **beyond** contract completion if necessary.

A cross-validation technique (after Skitmore, 1992) is then used to establish which variables should be included in the final equation. The α and β coefficients are computed by analysis of the completed contract data **excluding one contract in turn** and applying the results obtained from the included contracts to the excluded contract. This provides a means of assessing improvements in predictive ability by **variable reduction**. This involves the following stepwise procedure

1. Start with regression terms α_0 and β_0
2. Let 'new' contract k=1
3. Assume trial values for the associated regression coefficients
4. Calculate a and b parameter values for a j contract (j not equal to k)
5. Calculate the sums of squares of forecast error, $ssqf_j$, for the j contract
6. Repeat 4 and 5 for remaining j contracts (j not equal to k) and sum $ssqf_j$ to give

tssq

7. Repeat 3 to 6 until tssq is minimum by a Newton-Raphson technique

8. Calculate a and b parameter values for the k contract using the estimated regression coefficients

9. Calculate cross validated $xssqf_k$ for the k contract

10. Repeat 3 to 9 for contracts k=2 to q, sum $xssqf_k$ and divide by the number of valuations used, N, to obtain the cross validated mean square error of forecast, $xmsq_{\alpha 0,\beta 0}$

11. Repeat 3 to 10 adding one new regression term each of $\alpha_1 V$, ... $\beta_5 T_4$ to obtain $xmsq_{\alpha 0,\beta 0,\alpha 1}$ ... $xmsq_{\alpha 0,\beta 0,\beta 5}$

12. Add regression term corresponding to the smallest xmsq into the equation **providing** this xmsq is less than xmsq on the previous iteration and repeat steps 2-11

13. When the smallest current xmsq is not less than xmsq on the previous iteration, switch into backward regression mode and repeat 2 to 12, excluding instead of entering regression terms. When the smallest current xmsq is not less than xmsq on the previous iteration, switch into forward regression mode again and repeat 2 to 12 adding a regression term. Continue switching until both forward and backward regression modes are terminated.

3.2. Synthetic methods

The synthetic methods do not use any historical data concerning completed contracts. Instead they typically rely on the extrapolation of valuation data concerning the target contract. Using least squares, the a_k and b_k parameters are estimated directly by minimising the error of **prediction**. At least 2 valuations need to have been completed before any parameter estimates can be made.

The synthetic methods use the valuation data in their raw and smoothed form, with varying numbers of data points, and weighted in various ways to allow for degradation over time.

3.2.1. Raw form

In the raw form, all the valuation data are used (i.e. s=p) for each of the six models. The a and b parameter values are estimated by method 1 (Appendix B) using the following procedure:

1. Start with contract k=1

2. Assume trial values for parameters a_k and b_k

3. Calculate $ssqp_k$

4. Repeat 2 and 3 until $ssqp_k$ is minimum

5. Calculate error of **forecast**, $ssqf_k$, using estimated a_k and b_k values

6. Repeat 2 to 5 for contract k=2,...,q, sum $ssqf_k$ and divide by the number of valuations used, N, to obtain the mean square error of forecast, msqf.

This process is repeated for each of the models under investigation. The model with the lowest msqf is then examined for improvement under different weighting regimes. Firstly the most recent valuation is weighted. Secondly, the two most recent valuations are weighted, etc. using the following procedure:

1. Initialise all weights w_i=1 (i=1, ... ,p)

2. Start with number of weights l=1

3. Start with contract k=1

4. Assume trial values for each weight w_i (i=1, ... ,l)

5. Assume trial values for parameters a and b parameter values for a j project (j not equal to k)

6. Calculate $wmsqp_j$

7. Repeat 5 and 6 until $wmsqp_j$ is minimum

8. Calculate $ssqf_j$ using the resulting a and b parameter values

9. Repeat 5-8 for remaining j projects (j not equal to k) and sum $ssqf_j$ to tssq

10. Repeat 4-10 until tssq is minimum

11. Calculate a and b parameter values for the k project by minimising the weighted error of prediction $wmsqp_k$

12. Calculate the error of forecast $ssqf_k$

13. Repeat 3-12 for projects k=2(1)q, sum $ssqf_k$ and divide by the number of valuation forecasts, N, to obtain msqf

14. Repeat 2-13 for number of weights l=2(1)10.

3.2.2. Smoothed form

In the smoothed form, Hudson's method is used, extended to cover 3 to 7 valuations.

The method proposed by Hudson is to use, "after a few months", the two most recent valuations smoothed, in conjunction with the three or five valuations, by means of a polynomial regression. These two smoothed valuations are then used to derive the two Hudson parameters by solving the resulting simultaneous equations - there being two equations and two 'unknowns'. In order to **replicate** Hudson two questions immediately arise: (1) how many months must elapse before the method is used, (2) how many terms should be used in the polynomial regression. In order to **extend** the Hudson method several other matters need to be considered: (3) will the use of more than two recent valuations improve forecasts, (4) will the use of a different number to three or five conjunctive valuations improve forecasts, (5) will the use of different models improve forecasts?

It turns out that (1) depends on whether we use a **percentage-wise** analysis or a **month-wise** analysis. A percentage-wise analysis involves considering the completion of projects in percentage terms, e.g. decile by decile. A month-wise analysis involves considering the completion of projects in monthly terms, or number of valuations made, eg, 3 valuations. Thus for a percentage-wise analysis we would count the number of valuations that have been made up to a certain percentage completion, e.g. 3 valuations have been made up to 10% completion of the contract duration. This distinction is important as shorter projects have less numbers of valuations made than longer projects for the same percentage duration, reducing the number of suitable cases in the database when large numbers of valuations are analysed. In the method described here, the analysis is arbitrarily stop when less than 50% of suitable cases are available.

Points (2) to (5) above are incorporated into a generalised optimisation model as follows:

Let m be the number of polynomial regression terms in the model, o be the number of smoothed valuations used to derive the model parameters and s-o will be the number of additional valuations used to perform the polynomial regression. Using the least squares criterion, the problem now is to find suitable values for m, o and s to minimise the mean square error of the model. Clearly m cannot exceed the total number of valuations, i.e. $m \leq s$ (where m=s the smoothed o valuations will be the same as the raw valuations). It should also be noted that in Hudson's original model o=2 and m=3 or 5. In our method we impose the arbitrary constraints:

1. $2 \leq m \leq 8$
2. $2 \leq o \leq 8$
3. $0 \leq s \leq 8$

For o=2, the Hudson parameters are obtained by solving simultaneously two equations of the form

$$a(x^2 - x) - \frac{6x^3 - 9x^2 + 3x}{b} = v - x \qquad \textbf{1}$$

For the other models, and the Hudson model where 2≥o, the model parameters are estimated by least squares (via a Newton-Raphson technique).

This involves the following procedure:

1. Set o=2

2. Start with s=2

3. Start with m=1

4. Start with contract k=1

5. Calculate regression coefficients, δ_{rk} by minimising sum of squares of smoothing error, $ssqv_j$ for r=1,2, ... ,m over the most recent s valuations for project k

6. Calculate v'_{ik} and hence a_k and b_k by minimising the smoothed error of prediction $sssqp_k$

7. Calculate the error of forecast $ssqf_k$

8. Repeat 5-7 for projects k=2,3, ... ,q, sum $ssqf_k$ and divide by the number of forecasts, N, to obtain msqf

9. Repeat 4-8 for m=2(1)7 providing m≤s

10. Repeat 3-9 for s=3(1)9 providing s≥o

11. Repeat 2-10 for o=2(1)8 providing sufficient valuations are available for at least half of the cases in the database.

This procedure is then repeated decile-wise for each of the 6 models.

3.3. Hybrid methods

Both the analytic and synthetic methods are inefficient in the use of the available data - analytic methods exclude the valuation data for contract k, and synthetic methods exclude the historical valuation data for the j=1,2, ... contracts in the database. Hybrid methods involve merging the two methods into one and offer the possibility of improved forecast accuracy through greater utilisation of the available data as well as through the combination of multiple individual forecasts (Clemen, 1989).

There are, of course, many ways in which such hybrid models may be devised although previous research in the use of combined forecasts generally indicates that simple models tend to outperform more sophisticated approaches (Clemen, 1989). In our case, an obvious approach is to start with an analytic method, as this is the **only** way of proceeding with no target contract valuation data, and smoothly extend this by the introduction of a synthetic formulation. The analytic method is extended by adding extra terms in the regression model for valuations already made. a_k and b_k values are calculated as before by eqns (12) and (13) and the 'error' associated with the previous valuations is found. These are then multiplied by the regression coefficients and the a_k and b_k values recalculated, i.e.

$$a_j = \alpha_0 + \alpha_1 V_j + \alpha_2 D_j + \alpha_3 T_{1_j} + \alpha_4 T_{2_j} + \alpha_5 T_{3_j} + \sum_{i=0}^{s-1} \alpha_{5+i+1} \left(\hat{v}_{p-i,j} - v_{p-i,j} \right) + \varepsilon_a \quad \textbf{(2)}$$

$$b_j = \beta_0 + \beta_1 V_j + \beta_2 D_j + \beta_3 T_{1_j} + \beta_4 T_{2_j} + \beta_5 T_{3_j} + \sum_{i=0}^{s-1} \beta_{5+i+1} \left(\hat{v}_{p-i,j} - v_{p-i,j} \right) + \varepsilon_a \quad \textbf{(3)}$$

The error of forecast is then minimised as before by cross-validation regression by the following stepwise procedure:

1. Start with regression terms α_0 and β_0

2. Start with contract k=1

3. Assume trial values for the regression coefficients in equation

4. Calculate a and b parameter values for a j contract (j not equal to k) using eqns (12) and (13)

5. Calculate difference between model and actual valuation for last s valuations, multiply by regression coefficients and add to a and b parameter values as eqns (2) and (3)

6. Using the revised a and b values, calculate $ssqf_j$ for the j contract

7. Repeat 4 to 6 for remaining j contracts (j not equal to k) and sum $ssqf_j$ to tssq

8. Repeat 3 to 7 until tssq is minimum

9. Calculate a and b parameter values for the k contract using the resulting regression coefficient estimates

10. Calculate cross validated $xssqf_k$ for the k contract

11. Repeat 3 to 10 for contracts k=2 to q, sum $xssqf_k$ and divide by the number of

valuations used, N, to obtain $xmsq_{\alpha 0,\beta 0}$

12. Repeat 3 to 11 adding one new regression term each of $\alpha_1 V$, ...

13. Add new regression term corresponding to the smallest xmsq into the equation **providing** this xmsq is less than xmsq on the previous iteration

14. When the smallest current xmsq is not less than xmsq on the previous iteration, switch into backward regression mode and exclude a regression term and repeat 2 to 13, excluding instead of entering terms. When the smallest current xmsq is not less than xmsq on the previous iteration, switch into forward regression mode again and include a regression term and repeat 2 to 13. Continue switching until both forward and backward regression modes are terminated.

4. Case study

The case study data comprise a set of 27 completed contracts in the United Kingdom, together with their associated monthly valuations (see Appendix A). The first column gives the project sequence number. The next two columns give the project contract value and duration. The contracts are categorised into four types of construction work: (1) steel-framed low rise buildings, (2) new build housing developments, (3) housing refurbishment projects, (4) multi-house 'pre-paint' maintenance contracts and these are shown in column four. The subsequent columns give the cumulative expenditure coordinate series in terms of percentage value completed (v) at the percentage duration completed (d).

The data were obtained from one single private practice quantity surveying firm and two local authority surveying departments. All contract values were rebased to 1974 prices by means of the R.I.C.S. Building Cost Information Service Tender Price Index. No adjustments were made for any inter-project variations such as winter working, industry holidays or delivery to site of steel or mechanical plant, since these adjustments were considered to be relatively small and have little effect on the results.

The results of applying the method to these data are as follows:

4.1. Analytic models

A summary of the six final models for each decile is given in Exhibit 1. Column 1 gives the percentage of the contract duration completed, column 2 gives the number of valuations to be forecasted over all the cases, and the suceeding columns give the best mean square cross-validation results for each of the six models using both forwards and backwards regression methods. The figures in bold print highlight the best results obtained for each percentile completed.

--- insert Exhibit 1 here ---

4.2. Synthetic models

4.2.1 Unweighted raw model

The results are shown in Exhibit 2 together with msqp and msqf values. The msqf statistic is directly comparable with xmsq in the analytic model as neither method involves the use of the **actual** forecast values in the estimation procedure.

--- insert Exhibit 2 here ---

4.2.2 Weighted raw model

The result of the raw weighted synthetic procedure for the Hudson model is given in Exhibit 3. Exhibit 4 summarises the best of these results for each decile.

--- insert Exhibit 3 and 4 here ---

4.3. Smoothed data

The msqf error is recorded over all 27 projects for s=0(1)7, o=2(1)8 and m=2(1)7 ($m \leq s+o$) for each decile completion. Where insufficient valuations are available for more than one half of the 27 projects, the msqf error is not recorded. Exhibit 5 shows the best results for each o value per decile for the Hudson model. Exhibit 6 summarises the best of these results for each decile.

--- insert Exhibits 5 and 6 here ---

4.4. Hybrid models

The result of the hybrid Hudson model is given in Exhibit 7. Exhibit 8 summarises the best of these results for each decile.

--- insert Exhibits 7 and 8 here ---

4.5. Discussion

The best results of all the tests are summarised in Exhibit 9 which clearly indicates that the best hybrid models easily outperform all others examined for these data in this case study except at 0 and 90 percent completion, and **all** the hybrid models outperform all others in the critical 10 to 50 percent completion range. The hybrid model also has desirable asymptotic characteristics, the error term reducing fairly smoothly over the duration of the contracts.

--- insert Exhibit 9 here ---

Restrictions on printing space precludes the graphical representation of all the forcasts for all the projects. Fig 1a-d, however, provides an example of the actual and forecasted values for project 25 using the best hybrid method for 0, 30, 50 and 90 percent completions.

Other features of interest in this study are that, from Exhibit 7, the best hybrid results are generally to be found when s is maximum, i.e. by incorporating potentially as many previous valuations as possible into the model. Exhibit 10 gives the variables finally entered into the logistic model for each decile, the o's marking the variables incorporated into the best model, and the resulting xmsq. This clearly shows that no α variables beyond α_{11} nor β variables beyond β_8 actually entered into the best models. This redundancy of variables seems to be a result of the arbitrariness of the forward step-wise procedure combined with the reduction in case data caused by the data demands on previous valuations, and suggests that, for these data it is not necessary to examine the addition of more than six and three previous valuations in the a and b models respectively.

--- insert Exhibit 10 here ---

Further redundancy also occurs in the α_1, α_3, and β_3 variables as these also do not enter into the best decile models.

Another point highlighted by the method is the apparent haphazard way in which variables appear in the various best decile models shown in Exhibit 10. Ideally, a simple algorithm would determine which variables should be in the model for each decile. This seems to be largely unpredictable with these data without going through the complicated and time consuming business of the cross-validation regression for each forecast (a project forecast based on the logistic model at % contract duration completion takes 1-2 hours computing time on a 486/50MH computer with these data). It is possible however that, with a larger data set, some general results may be obtained empirically.

5. Other error measures

In the absence of an agreed loss function, and to provide a check on the results obtained by mean square error analysis, the procedure reanalyses the data using mean absolute errors (mabs) and mean percentage absolute errors (mpabs) instead of mean square errors. In this case the results again indicated the clear superiority of the hybrid methods against the alternatives. The breakdown for each of the hybrid methods again indicated no clear preference between the six models and the summary of variables in the models also showed little consistency in the variables included in the model at each decile. The best results were again shown to occur generally when s is maximum but this time with little variable redundancy.

6. Comparison with current practice

Little is known of the methods used in current practice. Hudson (1978) urged the use of his method involving the solution of two simultaneous equations (Exhibit 5), but it is doubtful if this has ever been used in practice. More likely, the simple analytical method, again proposed by Hudson (1978), in which the values of the two parameters are provided in a table of contract value ranges, is used without modification in the light of incoming valuations. Perhaps the most sophisticated method in current use is that described by Townsend (1994) in relation to his own practice, Turner and Townsend, one of the largest in the United Kingdom. This method uses Hudson's simple analytical method first to generate forecasted values for the whole of the contract period. These forecasted values are then adjusted in proportion to the error in the most recent forecasted value. For example, if the forecasted value for the first valuation is $100000, and this turns out to have an actual value of $120000, then all the forcasted values for the future valuations for the project are increased similarly by 20%. Exhibit 11 summarises the results that would have occurred with these data using both the unadjusted Hudson method and the Townsend method. Comparison with Exhibit 9 indicates the extent of the difference in performance of the methods.

--- insert Exhibit 11 here ---

7. Summary and conclusions

In this paper we have proposed a method to deal with a situation where a construction project has commenced, some instalments have been paid, and estimates of future instalments are needed. Using cross-validation regression, the method compares three groups of alternative approaches - (1) analytic, (2) synthetic, and (3) hybrid - in conjunction with six models - (1) Hudson, (2) Kenley-Wilson, (3) Berny-Howes, (4) cumulative logistic, (5) cumulative normal, and (6) cumulative lognormal. This is demonstrated by means of a case study in which it is shown that, with the database used, the best hybrid models produce the most accurate *ex-ante* forecasts for contracts 10 to 90 percent complete, and **all** the hybrid models produce the most accurate *ex-ante* forecasts for contracts 10 to at least 50 percent complete for the three error measures used.

There are many possibilities for further work in this area:

o The developement of diagnostics, such as residual analysis, would be an important aid to future refinements in the modelling procedure.

o It would be useful to test the hybrid approach with each of the four types of projects in the 27 project sample. There could be unique issues among these 4 project types, particularly the retrofit (type 3) and maintenance (type 4) projects as distinct from the new projects (types 1 and 2). There were very serious schedule overruns on two of the type 4 projects (number 5 was 78% over schedule and number 6 was 84% over schedule).

- o It would be useful to introduce further variables representing the characteristics of the contract and further models.

- o Models of extra-contract duration and expenditure will extend the curves to the actual final payment and an efficient algorithm will determine the variables to be included in the model for a given duration completed.

- o An empirical study involving larger scale examination of world-wide data could reveal an underlying generalised model.

In all these cases the prospects for application into practice are bright. Even without a variable determining algorithm, a project expenditure flow can be forecast by the method described in this paper (with a good estimate of its *ex-post* accuracy) in 1-2 hours of computing time. With a VTA, this would be reduced to a few seconds of computing time. And with a generalised model, a forecast should be possible with a hand calculator or look-up tables. Meanwhile, however, practitioners may wish to use the methods described here in analysing their own data, to make their own comparison of models and develop their own models. For this reason sufficient details have been given in the paper to enable a complete replication to be made to test the system prior to the user's own empirical analysis.

Finally, there is reason to suppose that the methods and models used here may be extended to forecasts of work other than construction. University Department expenditure flows, for example, are notoriously difficult to forecast but may well be suitable for modelling in the manner described in this paper!

8. Acknowledgements

Many thanks go to Tony Sampson and David Skinner of Salford University for their comments on the first draft of this paper. I am also grateful for the helpful and constructive criticisms of the journal's Associate Editor and the two anonymous reviewers.

9. References

Berny, J. and R.Howes, R., (1982), Project management control using real-time budgeting and forecasting models, *Construction Papers*, 2(1), 19-40.

Clemen, R.T., (1989), Combining forecasts: a review and annotated bibliography, *International Journal of Forecasting*, 5, 559-83.

Hudson, K.W., (1978), D.H.S.S. expenditure forecasting method, *Chartered Surveyor: Building and Quantity Surveying Quarterly*, 5(3), 42-45.

Kenley, R., and O.D.Wilson, (1989), A construction project net cash flow model, *Construction Management and Economics*, 7, 3-18.

Khosrowshahi, F., (1991), Simulation of expenditure patterns of construction projects, *Construction Management and Economics*, 9, 113-132.

Miskawi, Z., (1989), An S-curve equation for project control, *Construction Management and Economics*, 7, 115-124.

Skitmore, R.M., (1992), Parameter prediction for cash flow forecasting models, *Construction Management and Economics*, 10(5), 397-413.

Townsend, P., (1994), Personal communication.

Tucker, S.N., (1986), Formulating construction cash flow curves using a reliability theory analogy, *Construction Management and Economics*, 4, 179-188.

Tucker, S.N., (1988), A single alternative formula for the Department of Health and Social Security S-curves, *Construction Management and Economics*, 6, 13-23.

Appendix A

Case data (1974 rebased)

Project number	Contract value (pounds)	Contract period (days)	Project type	v	d	v	d	v	d	v	d	v	d	v	d	v	d	v	d	v	d
1 379996	266	3	9.36	14.02	25.49	30.38	42.89	44.39	47.65	57.48	58.70	72.90	76.32	90.19	91.46	103.27	105.00	121.49	107.85	124.30	
2 682681	863	2	6.01	7.54	6.88	11.17	8.43	16.66	9.83	21.22	12.25	26.73	14.81	30.66	18.30	34.44	22.76	40.72	28.87	45.44	
			35.57	50.79	41.02	55.18	44.39	60.21	49.49	64.15	58.56	69.64	62.33	74.52	68.08	79.40	72.10	83.96	76.24	88.04	
			78.87	92.76	82.82	98.11	86.16	102.34	89.22	107.54	91.70	113.36	92.44	116.82	94.89	122.00	96.31	126.25	97.52	135.68	
3 200649	379	2	4.35	5.61	7.20	10.97	15.56	18.78	23.83	26.35	32.12	33.17	44.51	43.42	51.26	50.24	61.33	58.53	76.16	65.85	
			83.80	72.44	92.24	81.22	93.34	89.51	93.76	92.44											
4 526024	552	2	4.13	7.60	7.78	11.30	13.27	17.67	21.29	22.62	28.07	27.74	33.03	32.68	37.74	40.10	41.73	44.35	48.00	49.47	
			55.06	53.53	61.45	60.96	69.47	66.96	76.78	71.03	83.75	75.80	88.70	81.98	92.72	87.10	94.72	92.05	96.88	97.53	
5 82003	255	4	10.62	17.61	34.14	44.22	44.65	57.51	50.94	74.35	58.34	90.39	64.78	101.35	72.32	124.82	74.23	139.69	76.30	151.43	
			82.52	160.43	83.96	176.47															
6 80108	255	4	7.58	17.60	26.76	31.31	36.39	44.22	51.36	56.34	61.61	71.60	67.64	83.34	80.94	94.31	85.31	129.13	90.42	143.61	
			96.39	156.53	97.80	171.39	107.35	184.31													
7 95305	375	4	8.67	8.33	17.53	15.85	27.32	23.65	36.46	31.17	42.88	42.18	53.82	48.09	63.35	55.61	70.83	61.52	74.70	77.37	
			76.02	81.94	93.78	89.19	100.56	107.73													
8 91355	204	4	6.98	10.40	18.27	27.17	39.00	38.73	54.59	47.98	80.66	67.63	96.77	84.97	100.38	100.00					
9 247347	442	2	4.61	5.58	10.41	12.86	19.60	18.93	24.73	28.15	31.92	35.68	36.98	43.20	45.15	51.94	50.78	59.47	62.34	66.26	
			76.02	75.00	83.82	81.55	88.25	90.29	93.26	95.88	97.00	102.91	98.80	107.28							
10	323717	650	1	9.09	10.39	16.76	17.48	26.32	24.56	33.97	29.87	46.93	40.76	53.28	47.85	58.13	54.94	61.56	62.29	66.52	71.40
			69.40	76.97	72.86	84.83	77.07	92.67	81.55	101.80	86.05	109.64	89.70	116.72	95.96	125.09	100.32	134.71	103.94	140.28	
			110.20	146.86	112.93	155.97	113.45	163.32													
11	272052	409	1	7.20	9.62	16.89	17.58	27.23	25.00	39.11	38.46	49.02	44.22	58.28	50.27	69.53	62.09	76.84	69.22	84.05	76.92
			93.81	86.81	99.32	96.15	101.23	103.84	103.40	112.36											
12	57493	262	1	15.15	12.37	28.54	20.27	50.07	31.96	63.75	43.64	86.36	55.32	95.14	63.91	96.07	75.60	111.78	90.03		
13	143440	214	1	21.11	14.98	38.38	28.50	60.23	43.96	75.21	58.94	84.85	77.77	92.06	91.31	92.41	103.38				
14	280346	496	1	1.65	4.42	7.88	12.21	13.78	21.04	25.81	32.72	35.33	40.26	41.06	47.27	45.95	56.88	50.79	64.16	61.33	73.24
			71.66	81.03	78.75	88.06	90.68	96.88	96.68	104.67	98.09	112.73	98.78	121.04	100.01	128.83					
15	342743	465	1	12.77	6.08	19.40	11.74	28.17	19.50	39.45	26.42	49.82	32.29	58.54	38.36	66.70	46.96	72.11	52.20	81.33	58.70
			86.03	63.52	88.84	71.28	94.37	78.40	95.05	84.91	96.53	93.91	97.37	97.48							
16	217335	212	3	16.62	19.22	30.04	29.55	47.96	50.23	59.70	64.04	73.42	79.30	85.29	95.07	90.45	104.43				
17	832869	405	3	1.44	2.20	6.14	11.82	11.84	20.88	19.67	30.50	24.50	38.18	33.84	45.87	41.34	53.57	49.57	63.18	58.79	70.87
			65.96	80.50	72.25	88.18	76.88	95.87	82.34	105.49	85.46	111.26									
18	316025	267	3	9.39	8.58	19.41	17.96	30.00	31.43	44.67	46.53	55.20	58.37	62.16	71.43	71.61	82.86	86.75	97.55	99.94	108.98
19	327272	297	3	5.55	8.58	19.19	29.48	22.78	38.06	45.56	48.51	56.56	61.57	83.06	82.46	92.78	97.01	98.20	110.82		
20	117184	121	3	22.68	39.56	49.03	89.01	65.83	119.78	81.05	132.97										
21	260746	394	1	9.54	9.61	14.16	17.14	22.66	25.45	33.90	32.98	42.02	40.52	54.55	49.09	62.85	56.37	69.82	65.46	78.69	71.95
			88.15	78.97	94.60	87.28	98.00	95.33	103.20	102.34											
22	197198	221	3	5.18	11.66	23.38	20.40	32.37	45.67	45.67	61.71	57.52	74.82	71.65	83.57	87.38	99.61	101.13	107.38		
23	195997	242	3	2.62	11.65	22.37	20.39	31.91	45.63	42.80	61.65	59.43	74.76	70.70	83.49	85.52	99.52	104.82	117.48		
24	43960	420	3	29.61	6.42	61.91	9.04	93.65	16.33	111.66	19.49	115.91	27.33	119.06	33.98	132.74	41.82				
25	656715	766	2	2.28	3.58	7.63	8.88	9.82	15.53	13.76	20.31	19.96	24.58	24.14	27.82	31.48	35.32	34.98	37.54	41.35	44.54

			50.11	50.51	55.62	55.46	62.56	62.12	67.69	65.19	73.61	70.65	76.45	73.56	80.52	81.23	85.11	86.01	87.54	90.11	
			91.27	97.95	92.45	102.22	93.47	108.03	93.79	112.97	95.27	118.77	95.74	125.94	96.27	130.72					
26	841063	923	2	1.02	4.26	3.33	7.56	4.41	12.36	6.30	16.34	8.50	21.02	11.05	24.72	12.93	29.40	14.06	34.35	18.25	38.05
			20.99	41.07	26.40	45.75	29.26	50.01	31.67	53.57	37.42	58.51	42.43	62.51	46.49	67.17	53.66	71.15	58.80	75.00	
			61.22	80.22	66.85	84.07	73.57	88.20	79.47	92.04	82.41	96.02	87.22	99.86	90.32	103.44	93.34	108.38	95.07	112.36	
			96.90	116.08	98.63	120.74	99.64	126.79													
27	279650	436	2	4.79	9.70	17.86	17.09	26.06	26.78	29.65	32.91	37.03	41.33	47.68	44.90	54.74	56.38	60.68	65.05	70.20	72.45
			82.65	77.80	93.70	87.50	94.82	94.89	96.55	104.34	98.73	111.22									

Appendix B

The models

Two parameter models only are used. These are: (1) the Hudson formula (Hudson, 1978), (2) the Kenley-Wilson formula (Kenley and Wilson (1989), (3) the Berny-Howes formula (Berny and Howes, 1982), (4) the cumulative logistic, (5) the cumulative normal, and (6) the cumulative lognormal probability distributions.

Definition

The symbols a and b denote each of the two parameters in the models below and x is the duration, expressed as the ratio d/100.

The Hudson formula

$$v = 100[x+ax^2-ax-(6x^3-9x^2+3x)/b] \quad \textbf{(4)}$$

The Kenley-Wilson formula

$$v = 100F/(1+F) \quad \textbf{(5)}$$

where

$$F = e^a[d/(100-d)]^b \quad \textbf{(6)}$$

The Berny-Howes formula

$$v = 100x\{1+a(1-x)(x-b)\} \quad \textbf{(7)}$$

The cumulative logistic distribution

$$v = 100\frac{e^{a+b^x}}{1+e^{a+b^x}} \quad \textbf{(8)}$$

The cumulative normal distribution

$$v = \mathrm{erf}\{(x-a)/b\}100 \quad \textbf{(9)}$$

where erf denotes the cumulative normal probability function.

The cumulative lognormal distribution

$$v = \mathrm{erf}\{(\mathbf{ln}x-a)/b\}100 \quad \textbf{(10)}$$

For the cumulative lognormal distribution, it is found computationally convenient to limit the search within the limits 10>a>-10 and 10>b>0.

Parameter estimation

The parameters, a and b, for a 'new' project, k, are estimated by one or more of the following technique-dependent methods:

1. Minimise error of prediction

i.e. minimise sums of squares of prediction, $(\min)ssqp_k$

2. Minimise weighted error of prediction

i.e. minimise weighted mean square of prediction error, $(\min)wmsqp_k$

where the w_i's are estimated by minimising the cross-validated forecast error

$$(\min)xwssqf_k = \sum_{j=1,jnek}^{q} ssqf_j \qquad \textbf{(11)}$$

where q is the number of projects in the database.

3. Minimise smoothed error of prediction

i.e. minimise sums of squares of prediction error derived from smoothed valuations, $(\min)sssqp_k$

4. Regression on project characteristics (contract type, value and duration)

i.e.

$$a_k = \alpha_0 + \alpha_1 V_k + \alpha_2 D_k + \alpha_3 T_{1_k} + \alpha_4 T_{2_k} + \alpha_5 T_{3_k} + \varepsilon_a \qquad \textbf{(12)}$$

$$b_k = \beta_0 + \beta_1 V_k + \beta_2 D_k + \beta_3 T_{1_k} + \beta_4 T_{2_k} + \beta_5 T_{3_k} + \varepsilon_b \qquad \textbf{(13)}$$

where V is the rebased contract sum (in 100,000 units), D is the contract duration (in 100 day units), T_1, T_2 and T_3 are dummy variables for the project types, and the α's and β's are estimated by minimising the cross validated forecast error

$$(\min)xssqf_k = \sum_{j=1,jnek}^{q} ssqf_j \qquad \textbf{(14)}$$

or total error

$$(\min) xssq_k = \sum_{j=1, j \neq k}^{q} (ssqf_j + ssqp_j) \qquad \textbf{(15)}$$

Estimates of the Hudson formula constants are also obtained from Hudson's (1978) table. This method is termed here the standard Hudson formula.

The error terms

Error measures

For the jth project in our database of q previously completed projects (j=1,2, ... ,q) we are interested in 4 error measures:

Error of prediction

This is denoted by the sum of squares of prediction error

$$ssqp_j = \sum_{t=1}^{s} (\hat{v}_{t,j} - v_{t,j})^2 \qquad \textbf{(16)}$$

where

s is the number of previous valuations considered ($s \leq p$) and

$$v_{ij} = f(a_j, b_j, d_{ij}) \qquad \textbf{(17)}$$

e.g. for the Hudson curve:

$$v_{ij} = 100\left(x_{ij} + a_j x_{ij}^2 - a_j x_{ij} - \frac{6 x_{ij}^3 - 9 x_{ij}^2 + 3 x_{ij}}{b_j} \right) \qquad \textbf{(18)}$$

where

$$x_{ij} = \frac{d_{ij}}{100} \qquad \textbf{(19)}$$

Error of forecast

This is denoted by the sum of squares of forecast error

$$ssqf_j = \sum_{t=p+1}^{n} (\hat{v}_{t,j} - v_{t,j})^2 \qquad \textbf{(20)}$$

Weighted error of prediction

This is denoted by the weighted mean square of the prediction error

$$wmsqp_j = \frac{\sum_{i=0}^{l-1} (\hat{v}_{p-i,j} - v_{p-i,j})^2 w_i + \sum_{i=l}^{s-1} (\hat{v}_{p-i,j} - v_{p-i,j})^2}{\sum_{i=0}^{l} w_i + s - l - 1} \qquad \textbf{(21)}$$

where l is the number of immediately previous valuations to be weighted (l≤s) and w is the weight given to each previous valuation

Smoothed error of prediction

This is denoted by the sum of squares of prediction error derived from o smoothed immediately previous valuations

$$sssqp_j = \sum_{i=0}^{o-1} (\hat{v}_{p-i,j} - v_{p'-i,j})^2 \qquad \textbf{(22)}$$

where o≥s and

$$v_{p'-i,j} = \sum_{r=0}^{m} \delta_{rj} v_{p-i,j}^{r} \qquad \textbf{(23)}$$

m being the number of regression coefficients used in the smoothing (m<s). The regression coefficients, δ_{rj}, are estimated by minimising

$$ssqv_j = \sum_{i=0}^{s-1} (v_{p'-i,j} - v_{p-i,j})^2 \qquad \textbf{(24)}$$

Exhibit 1: Mean square forecast errors of analytic models

%	N	Hudson	Kenley &	Berny & Wilson	Logistic Howes	Normal	Lognormal						
		fwd	bkwd	fwd	bkwd	fwd	bkwd	fwd	bkwd	fwd	bkwd	fwd	bkwd
0	287	111.0	111.0	185.0	185.0	223.0	157.9	228.0	200.4	195.3	195.3	213.0	213.0
10	268	101.9	126.7	219.2	206.5	226.1	160.1	229.9	206.8	199.1	199.1	213.0	213.0
20	235	117.0	100.1	184.9	178.8	195.4	130.9	193.9	186.1	178.6	169.5	181.6	181.6
30	206	124.9	124.2	176.1	166.6	185.1	132.2	182.1	177.2	165.6	157.3	170.5	165.8
40	180	138.7	163.6	140.9	146.7	163.4	141.7	156.4	155.0	139.9	132.8	142.1	137.2
50	145	124.1	87.4	98.1	98.1	101.2	103.0	95.8	95.2	125.9	96.4	95.4	99.4
60	118	125.1	95.4	95.9	101.8	97.5	97.5	93.9	96.6	92.3	92.9	89.1	89.1
70	92	104.5	76.5	85.7	85.7	124.6	107.5	82.6	92.8	79.9	79.9	80.4	80.4
80	61	114.4	89.3	85.3	85.3	104.9	94.0	116.2	87.4	74.0	72.0	73.9	70.9
90	31	54.4	53.3	52.0	52.9	59.2	42.4	56.9	60.7	38.5	38.6	38.5	38.2
100	0	-	-	-	-	-	-	-	-	-	-	-	-

bold lowest msqf

Exhibit 2: Mean square errors of synthetic models

%	N	Hudson		Kenley &	Berny & Wilson		Logistic Howes	Normal	Lognormal					
	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf
0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	6	44	0.0	****.*	0.0	837.8	0.2	****.*	0.0	2444.7	0.0	2109.3	0.0	305.0
20	42	177	4.2	429.7	4.6	276.6	4.7	303.2	4.2	702.9	4.3	532.1	4.3	847.6
30	78	189	7.2	77.5	6.4	211.6	9.1	321.2	6.4	392.6	6.3	330.1	6.3	644.0
40	105	175	6.8	88.6	8.9	185.4	14.4	301.7	9.4	272.6	9.1	250.2	9.6	473.5
50	141	144	7.2	66.3	16.7	60.9	15.4	48.2	18.8	74.5	18.1	65.3	18.0	277.8
60	168	117	7.7	54.1	15.1	52.8	13.6	47.9	17.6	39.4	16.7	37.1	17.8	213.0
70	194	91	6.4	54.9	14.0	57.1	12.8	53.1	16.7	38.1	15.6	38.3	18.7	174.6
80	225	60	7.9	56.7	13.9	51.1	13.4	53.9	16.4	23.7	15.5	22.6	21.7	135.6
90	256	31	10.1	76.6	14.2	61.2	13.8	72.3	16.1	20.6	15.2	18.9	26.3	101.7
100	287	0	16.5	-	18.0	-	19.6	-	16.2	-	15.1	-	31.4	-

bold lowest msqf

Exhibit 3: Mean square errors of weighted synthetic Hudson model

%									l														
		0	1	2	3	4	5	6	7	8	9	10											
	N	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp	msqf	msqp
0		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	177	***.*	**.*	***.*	**.*	***.*	**.*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	189	614.3	7.8	614.3	7.8	614.4	6.8	94.7	9.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	175	644.6	12.6	644.6	12.6	644.5	12.6	61.0	14.3	60.6	16.0	-	-	-	-	-	-	-	-	-	-	-	-
50	144	60.5	14.6	60.5	14.6	60.5	14.6	46.0	15.0	36.8	15.4	36.9	16.8	39.1	17.5	-	-	-	-	-	-	-	-
60	117	48.8	14.7	48.8	14.7	49.3	14.7	**.*	**.*	41.1	14.7	40.7	15.7	36.8	15.6	39.0	16.2	-	-	-	-	-	-
70	91	57.2	14.7	57.2	14.7	56.7	14.7	56.7	14.7	46.8	15.0	41.5	14.6	42.4	15.3	42.1	15.7	42.1	5.3	-	-	-	-
80	60	66.7	16.0	66.7	16.0	66.7	16.0	67.2	16.5	67.2	16.5	55.8	15.5	51.9	15.0	49.8	15.1	47.1	5.0	47.9	5.8	-	-
90	31	94.4	20.9	94.4	20.9	93.6	20.4	94.4	17.3	93.6	16.9	73.1	16.5	73.9	17.3	67.8	16.0	62.1	7.7	62.1	7.7	62.1	7.7

bold lowest msqf

Exhibit 4: Summary of best weighted synthetic models

%	Hudson		K&W		BEHO		Logistic		Normal		Lognormal	
	l	msqf	l	msqf	l	msqf	l	msqf	l	msqf	l	msqf
10	-	-	-	-	-	-	-	-	-	-	-	-
20	*	**.*	0	**276.6**	0	303.2	2	648.9	1	519.6	0	847.6
30	3	**94.7**	2	134.9	3	103.0	3	272.5	3	142.2	3	516.0
40	4	60.6	4	**56.4**	4	64.4	3	96.9	3	96.2	4	294.3
50	4	36.8	5	40.7	4	39.7	5	44.1	5	**31.8**	5	190.2
60	7	39.0	6	40.2	5	39.3	7	**27.1**	6	27.8	2	151.8
70	7	42.1	8	42.4	5	39.6	7	28.7	8	**24.7**	2	86.5
80	8	47.1	8	42.3	8	42.1	1	**16.6**	4	17.4	2	34.5
90	8	62.1	10	45.1	8	53.1	10	11.1	8	**9.4**	2	14.6

bold lowest msqf

Exhibit 5: Percentage-wise Hudson analysis - Hudson model

%	o=2			o=3			o=4			o=5			o=6			o=7			o=8		
	s	m	msqf	s	m	msqf	s	m	msqf	s	m	msqf	s	m	msqf	s	m	msqf	s	m	msqf
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	0	2	791.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	1	2	97.2	0	3	125.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	2	2	45.1	1	2	48.8	0	2	59.7	-	-	-	-	-	-	-	-	-	-	-	-
50	2	2	38.4	2	5	35.8	1	3	36.7	1	2	38.9	0	2	40.4	-	-	-	-	-	-
60	4	3	36.0	2	2	38.8	2	3	38.0	1	5	37.1	0	4	39.0	0	3	40.9	-	-	-
70	5	3	34.7	4	3	35.9	2	2	40.2	1	2	40.0	1	7	39.9	1	4	40.3	0	6	41.0
80	5	2	32.9	4	2	42.4	4	4	39.6	2	2	44.5	3	3	45.1	1	2	45.3	0	2	45.4
90	6	7	31.4	5	5	53.8	4	2	56.4	3	2	58.5	3	3	59.5	2	3	58.3	1	2	60.7

bold lowest msqf

Exhibit 6: Percentage-wise Hudson analysis

%	Hudson				K&W				BEHO				Logistic				Normal				Lognormal			
	o	s	m	msqf	o	s	m	msqf	o	s	m	msqf	o	s	m	msqf	o	s	m	msqf	o	s	m	msqf
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	2	0	2	791.3	2	0	2	**360.9**	2	0	2	542.3	2	0	2	653.5	2	0	2	537.5	2	0	2	817.3
30	2	1	2	97.2	2	1	2	93.6	2	1	2	**84.9**	2	1	3	161.2	2	1	2	148.0	2	1	2	431.5
40	2	2	2	45.1	3	1	2	47.7	2	2	2	**43.2**	2	2	2	66.8	3	1	3	72.3	2	2	4	203.0
50	3	2	5	35.8	2	4	2	36.7	3	2	5	35.7	3	2	5	31.4	3	2	5	**29.4**	2	2	3	120.1
60	2	4	3	36.0	2	3	2	37.9	3	3	3	36.3	6	1	2	35.1	3	3	3	**33.8**	2	2	3	65.4
70	2	5	3	34.7	3	4	3	32.9	3	4	3	35.9	6	2	6	30.6	6	2	6	**28.7**	2	5	3	49.0
80	2	5	2	32.9	2	6	2	26.8	4	5	3	38.2	2	7	7	**12.0**	2	7	7	12.8	2	6	6	12.9
90	2	6	7	31.4	2	6	4	16.7	2	8	7	47.1	2	8	5	**2.1**	2	6	7	2.5	3	6	2	10.4

bold lowest msqf

Exhibit 7: Mean square forecast errors of hybrid Hudson model, forward cross-validation regression

%											s											
	0	1	2	3	4	5	6	7	8	9	10											
	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq	N	xmsq
0	287	111.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	268	101.9	195	55.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	235	117.0	233	71.4	177	67.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	206	124.9	204	55.9	189	38.2	149	27.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	180	138.7	180	45.4	175	36.7	148	35.0	130	27.6	-	-	-	-	-	-	-	-	-	-	-	-
50	145	124.1	145	43.1	144	37.4	137	33.5	119	35.7	105	31.4	99	32.6	-	-	-	-	-	-	-	-
60	118	125.1	118	30.3	117	31.8	117	35.1	104	28.8	91	27.6	85	25.3	80	29.8	-	-	-	-	-	-
70	92	104.5	92	40.2	91	33.8	91	33.8	89	35.7	74	34.3	67	29.9	65	35.0	65	35.0	-	-	-	-
80	61	114.1	61	35.1	60	36.4	60	34.5	60	38.5	59	52.0	46	24.0	44	25.3	43	21.5	43	21.5	-	-
90	31	54.4	31	57.3	31	57.3	31	57.3	31	57.3	30	38.7	27	37.8	22	26.9	20	20.0	20	20.0	20	20.0

Exhibit 8: Summary of best hybrid models (xmsq)

%	Hudson nw	Hudson xmsq	K&W nw	K&W xmsq	BEHO nw	BEHO xmsq	Logistic nw	Logistic xmsq	Normal nw	Normal xmsq	Lognormal nw	Lognormal xmsq
0	0	111.0	0	185.0	0	223.0	0	228.0	0	195.3	0	213.0
10	1	55.5	1	52.4	1	54.0	1	55.6	1	40.4	1	69.4
20	2	67.8	2	54.0	2	47.7	2	46.8	2	50.2	2	65.3
30	3	27.7	3	40.5	3	30.2	3	37.8	3	35.1	3	40.7
40	4	27.6	4	28.1	4	29.6	4	25.5	4	27.1	4	37.1
50	5	31.4	6	26.0	4	23.4	5	22.4	6	17.7	5	24.7
60	6	25.3	6	13.5	6	20.1	7	12.8	4	18.7	7	13.6
70	6	29.9	8	26.1	6	16.1	7	23.6	8	10.0	7	17.2
80	8	21.5	8	7.0	8	19.5	8	5.2	8	12.3	6	12.0
90	8	20.0	8	2.6	8	13.6	8	2.5	8	3.4	8	4.3

bold lowest xmsq

Exhibit 9: Summary of best results (msq)

%	Analytic (forecast) (Exhibit 1)	Synth (Raw) (Exhibit 2)	Method Synth (Wgtd) (Exhibit 4)	Synth (Hudson) (Exhibit 5) (o=2)	Synth (Smthd) (Exhibit 6) (2≤o)	Hybrid (forcst) (Exhibit 8)
0	111.0	-	-	-		111.0
10	101.9	305.0	-	-		40.0
20	100.1	276.6	276.6	360.9	360.9	46.8
30	124.2	77.5	94.7	84.9	84.9	27.7
40	132.8	88.6	56.4	43.2	43.2	25.5
50	87.4	48.2	31.8	33.3	29.4	17.7
60	89.1	37.1	27.1	34.8	33.8	12.8
70	76.5	38.1	24.7	31.1	28.7	10.0
80	70.9	22.6	16.6	12.2	12.0	5.2
90	38.2	18.9	9.4	2.1	2.1	2.5

Exhibit 10: Variables in the logistic hybrid model (msqf)

Variable					%						
	0	10	20	30	40	50	60	70	80	90	Total
α_0	o	o	o	o	o	o	-	o	o	o	9
α_1	-	-	-	-	-	-	-	-	-	-	0
α_2	-	-	-	-	-	-	o	-	-	-	1
α_3	-	-	-	-	-	-	-	-	-	-	0
α_4	-	-	o	-	o	-	o	-	-	-	3
α_5	o	-	-	-	o	-	-	-	o	-	3
α_6		o	o	o	o	o	o	o	-	o	8
α_7			-	o	-	-	o	-	o	-	3
α_8				-	-	-	o	-	o	-	2
α_9					-	-	-	-	o	-	1
α_{10}						-	-	-	-	o	1
α_{11}						-	-	-	o	-	1
α_{12}							-	-	-	-	0
α_{13}									-	-	0
α_{14}											
α_{15}											
β_0	o	o	o	o	o	o	o	o	o	o	10
β_1	-	o	o	-	-	-	o	o	-	-	4
β_2	-	-	-	-	o	-	-	-	-	-	1
β_3	-	-	-	-	-	-	-	-	-	-	0
β_4	-	-	-	-	o	o	-	-	-	-	2
β_5	-	o	-	-	-	o	o	o	o	o	6
β_6		-	-	-	-	o	-	-	o	-	2
β_7			o	-	-	-	-	-	-	-	1
β_8				-	o	-	-	-	-	o	2
β_9					-	-	-	-	-	-	0
β_{10}						-	-	-	-	-	0
β_{11}						-	-	-	-	-	0
β_{12}							-	-	-	-	0
β_{13}									-	-	0
β_{14}											
β_{15}											
msqf	228.0	55.6	46.8	37.8	25.5	22.4	12.8	23.6	5.2	2.5	

Exhibit 11: Results of using methods in current practice

%	Hudson's simple analytical method			Townsend's method		
	MSQF	MABS	MPABS	MSQF	MABS	MPABS
0	-	-	-	-	-	-
10	307.7	11.5	26.0	12143.1	56.9	1010.9
20	284.6	11.9	24.1	1440.7	25.5	41.3
30	281.2	12.3	23.3	953.0	21.1	32.0
40	269.2	12.5	21.9	542.6	16.0	23.6
50	259.7	13.1	21.6	161.6	8.7	11.7
60	272.2	13.5	20.1	106.1	7.4	9.5
70	274.7	13.4	19.3	101.5	7.8	9.6
80	268.0	13.0	17.6	47.6	4.9	5.8
90	198.4	11.3	14.0	26.2	3.8	4.4