

City University of New York (CUNY)

CUNY Academic Works

Publications and Research

Hunter College

1971

A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts

William (Bill) H. Williams
CUNY Hunter College

M. L. Goodman

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/hc_pubs/47

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts

W. H. WILLIAMS and M. L. GOODMAN

Reprinted from the Journal of the American Statistical Association
December 1971, Volume 66, Number 336
Pages 752-754

A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts

W. H. WILLIAMS and M. L. GOODMAN*

A simple method for the construction of empirical confidence intervals for time series forecasts is described. The procedure is to go through the series making a forecast from each point in time. The comparison of these forecasts with the known actual observations will yield an empirical distribution of forecasting errors. This distribution can then be used to set confidence intervals for subsequent forecasts. The technique appears to be particularly useful when the mechanism generating the series cannot be fully identified from the available data or when limits based on more standard considerations are difficult to obtain.

1. INTRODUCTION

The object of a Bell System study was to develop eighteen-month forecasts of the growth of main telephone stations, eighteen months being the approximate time required to construct additional switching capacity. It was imperative that the forecasts be accompanied by confidence intervals because such intervals were to be translated into time intervals giving the likely earliest and latest dates at which current plant capacity would be exceeded. These time intervals would, of course, be accompanied by the appropriate probability coefficient. In this study, the confidence statements were at least as important as the point forecast itself in influencing management's decision to install additional equipment.

The forecasting methodology in this application is in a rapid stage of evolution in which the techniques, models, and available data can be expected to improve substantially in the future. Nevertheless, daily operations demand that the present methods and their associated forecasts be used *now*. Hence, the question arises: if we know that the model is not as good as it should be, what can be said about the confidence intervals developed from that model? In such a situation, it is helpful to ask how well the procedure would have worked, had it been used in recent history. In this article, this question is considered relative to the standard method of constructing confidence statements, and a second method is proposed which can be used for comparative purposes.

Situations like the one just described seem to be much more frequent in socioeconomic studies than in the physical sciences, where modeling can be much more precise. In the social sciences, it is not at all uncommon to have a very good fit to a set of data and yet have the forecasts be poor. This fact is reflected in the title of the article.

* W. H. Williams is member of the Technical Staff and M. L. Goodman is associate member of the Technical Staff, both with the Statistics and Data Analysis Research Department, Bell Telephone Laboratories, Mountain Avenue, Murray Hill, N. J. 07974.

2. DATA

The data consisted of six time series of 169 points each. They were the number of residence and the number of business main telephones in service on the last day of the month for three Michigan cities—Flint, Grand Rapids, and Battle Creek.

Early analysis showed that it would be more appropriate to work with first differences rather than the original series. Consequently, the analysis presented is in terms of the first differences, $y_t = x_t - x_{t-1}$ where the $\{x\}_t$ series represents the raw data. The first difference, y_t , measures the gain or loss in telephones for month t .

3. THE FORECASTING MODEL

3.1 The Regression Model

Several forecasting models were investigated for predicting these series. Only the one which performed the best over all six cities is presented in this paper. This model has a simple regressive form,

$$y_t = \gamma_1 t + \gamma_2 t^2 + \sum_{i=1}^{12} \beta_i m_{it} + e_t \quad (3.1)$$

where the γ 's and the β 's are the regression coefficients, and the m 's are 0-1 indicator variables such that $m_{it} = 1$ if $t \equiv i \pmod{12}$, 0 otherwise, $i = 1, 2, \dots, 12$. The $\{y_t\}$ series all began with February, so that $m_{1t} = 1$ corresponds to February, $m_{2t} = 1$ to March, etc.

The errors, e_t , were assumed to have zero expectation and variance-covariance matrix $V\sigma^2$, with $V = \text{diag}(\rho^{-i-1})$ and ρ known. Unequal variances were used because it was felt that the more recent observations should be weighted more heavily in the forecasting. The parameters were then estimated by least squares in the standard way (see [2]) with $\rho = 0.96$. This value of $\rho = 0.96$ gives the confidence intervals highest accuracy, and was selected by numerical evaluation of the ranges 0.50(0.05)0.95 and 0.95(0.01)0.99. The estimates of the β 's are approximately exponentially weighted monthly averages.

3.2 Prediction from the Model

For each series, a first forecast was obtained by fitting the model to the first 24 points of the series and then extrapolating to forecast the gain in telephones for eighteen months into the future. Confidence intervals

based on the normal distribution were then constructed in the standard way, see for example, [1] or [2].

Subsequently, forecasts were made by including the next point in the time series, adjusting the regression coefficients, and repeating the forecasting procedure. In this manner, a series of eighteen-month forecasts were obtained for each of the six time series along with a confidence statement for each forecast.

Notice that this procedure has yielded six series of forecast errors, $E_t = F_t - A_t$, which describe the errors that would have been made if the method had been used over the period of the historical time series. In the subsequent parts of this article the forecasts and the actual errors remain unchanged. However, *two* methods for assessment of the forecasts are discussed.

4. CONFIDENCE INTERVALS

4.1 Standard Confidence Limits

Since we have constructed forecasts and confidence limits in the same way in which they would have been constructed if the procedure had been used over the period covered by the data, it is worthwhile to ask the number of times that the actual value fell outside of standard confidence intervals of size 0.80, 0.90 and 0.95. This was done for all the regressions for which a comparison of the actual versus the forecast was possible (169 - 24 - 18 = 127 per series). It was found that of the 762 comparisons, the 80 percent confidence bands covered the actual realized value 533 times, yielding an empirical confidence level of only .699 or 70 percent. Similarly, the empirical levels for the 90 percent and 95 percent intervals were considerably below the desired rates.

It might be argued that the confidence limits for the early forecasts are unreliable because they are based on so few points. Consequently, the calculations were repeated using only the last 85 predictions for each series and it was found that realized proportions were 70.8 percent, 80.0 percent and 85.7 percent. This slight improvement led us to try even shorter periods using the last 73, 61 and 49 predictions, but no further improvement in accuracy was noted. All of these results are shown in the top half of the table.

Similar results have been found with other data.

4.2 Empirical Confidence Intervals

The results of Section 4.1 show that if these standard confidence intervals were applied to the problem discussed in Section 1, the result would have been that too often plant capacity would have been exceeded at points of time outside the estimated intervals. Consequently, for comparison, an alternate method of constructing confidence limits was developed. The method bases the confidence intervals more directly on the forecast errors E_t . By accumulating these errors through time, an empirical distribution of the forecast errors can be obtained. If it is assumed that future forecast errors come from the same distribution as past forecast errors, then confidence

ASSESSMENT OF CONFIDENCE INTERVALS

| No. of errors used | Comparisons per series | Total comparisons | Times exceeded | | | Actual rate | | |
|--|------------------------|-------------------|----------------|-----|-----|-------------|------|------|
| | | | 80% | 90% | 95% | 80% | 90% | 95% |
| <u>Standard normal confidence limits</u> | | | | | | | | |
| | 127 | 762 | 229 | 163 | 123 | .699 | .786 | .839 |
| | 85 | 510 | 149 | 102 | 73 | .708 | .800 | .857 |
| | 73 | 438 | 142 | 99 | 71 | .676 | .774 | .838 |
| | 61 | 366 | 124 | 95 | 72 | .661 | .740 | .803 |
| | 49 | 294 | 99 | 74 | 56 | .663 | .748 | .809 |
| <u>Empirical confidence limits</u> | | | | | | | | |
| 24 | 85 | 510 | 107 | 78 | 61 | .790 | .847 | .880 |
| 36 | 73 | 438 | 90 | 60 | 45 | .795 | .863 | .897 |
| 48 | 61 | 366 | 74 | 52 | 35 | .798 | .858 | .904 |
| 60 | 49 | 294 | 56 | 29 | 14 | .810 | .901 | .952 |

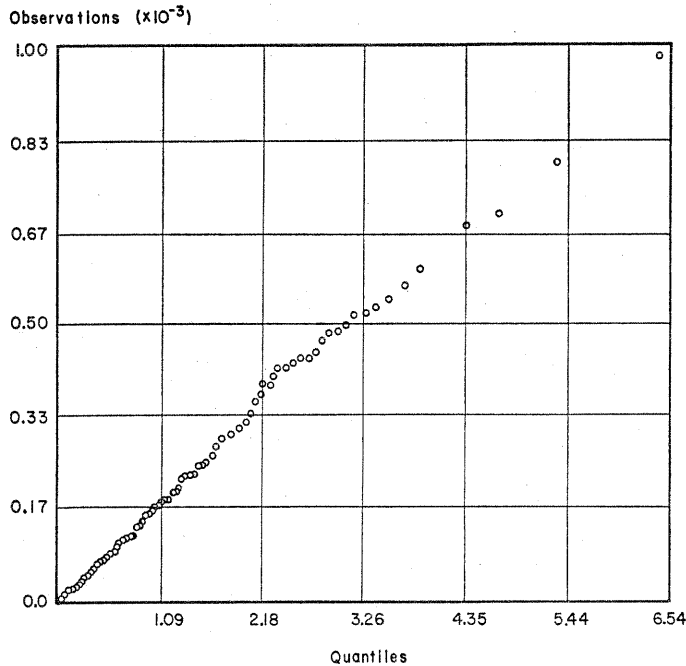
limits for future forecasts can be determined by using the percentage points of this empirical distribution.

It is important to note that the standard method assumes that the forecasting model describes the series adequately in the future, while the second method requires that the distribution of the forecasting errors remains the same in the future. It is hoped that this assumption will be somewhat weaker and will provide a more robust method.

Operationally, it was convenient to work with the absolute values of the forecast errors. The absolute errors seemed, in this case, to approximately follow a gamma distribution. As an example, for Grand Rapids, the figure shows the straight line probability plot of the ordered absolute forecast errors against the quantiles of the fitted gamma distribution. Subsets of consecutive forecast errors also seemed to follow a gamma distribution, suggesting that the assumption of a gamma distribution of errors was indeed stable over time. This information was used in the later formation of confidence intervals based on subsets of forecast errors. The general method, however, is not dependent upon the use of the absolute value of the forecast errors or on the gamma distribution. Some other function of these errors, such as their square might have been used, the empirical distribution fitted and used to construct confidence limits. The choice of function and the characteristics of the errors could presumably lead to virtually any shape distribution.

An available option with this procedure is to vary the number of errors used in estimating the parameters of the gamma distribution. The estimates which consider more values should be more reliable; however, the farther back one goes in the series the less tenable will be the assumption of a stable distribution of forecasting errors. As a result, different numbers were tried. First, 24 absolute forecast errors were used, and the parameters of the underlying gamma distribution estimated by the method of Wilk, Gnanadesikan and Huyett [3]. Then the 80, 90 and 95 percentage points of that distribution were found. These percentage points gave confidence intervals for the forecast made at the time of the 24th absolute error. Then

GAMMA PLOT OF ACTUAL FORECAST ERRORS, $N=127$
GRAND RAPIDS BUSINESS TELEPHONES



the 25th error was included in the sample, the first deleted, and the gamma percentage points re-estimated from this new set of 24 absolute errors. This yielded confidence intervals for the next forecast. By moving through time in this manner a series of confidence intervals was generated, just as had been done with the standard procedure. These empirical intervals were then assessed as before.

This time there were 85 points for comparison for each series, 510 points in all. The results are shown in the bottom half of the table. The empirical level is quite satisfactory for the 80 percent intervals, but we felt that the 90 percent and 95 percent levels might be improved by including more points in the gamma estimation. This might help to allow for longer term demographic changes in the cities. Consequently the calculations were repeated using 36, 48 and 60 forecast errors. The assumption of a

gamma distribution of absolute forecast errors seems particularly good when 60 forecast errors are used. The results are shown in the table.

5. COMPARISON AND SUMMARY

The new intervals are constructed so that they are directly determined by the historical forecast errors. One important characteristic of the suggested procedure is that it allows *any* error distribution to be introduced in a simple way. In particular it allows the use of longer-tailed and/or asymmetric distributions. In the example of this article a gamma distribution was used.

In addition the new method has the decided advantage in that it may be used in many situations where the standard method cannot, or is awkward to apply. For example, if forecasts are truncated for some reason (say at zero) truncated distributions have to be considered and confidence intervals may be difficult to obtain. The empirical method, however, may be applied in a straightforward manner.

A further point is that the method forces analysis of the actual forecasting errors which are themselves often very informative.

It should be pointed out that the problem could also be approached by modifying the standard confidence intervals so that they are based on a non-normal (and presumably longer-tailed) distribution. This of course would require considerably more mathematical analysis with *each* application; which in some cases would be quite difficult. It certainly seems appropriate to construct the empirical confidence limits along with the standard ones whenever the forecasting model is regarded with some suspicion.

REFERENCES

- [1] Goldberger, Arthur S., *Econometric Theory*, New York: John Wiley & Sons, Inc., 1964, 167-70.
- [2] Graybill, Franklin A., *An Introduction to Linear Statistical Models, Volume I*, New York: McGraw-Hill Book Company, Inc., 1961.
- [3] Wilk, M. B., Gnanadesikan, R. and Huyett, M. J., "Estimation of Parameters of the Gamma Distribution Using Order Statistics," *Biometrika*, 49 (December 1962), 525-45.