A FRESH APPROACH TO COMBINING FORECASTS

Frank G. Landram, West Texas A&M University
Francis Mendez, Texas State University
Vivek Shah, Texas State University
Suzanne V. Landram, Walter Reed Army Institute of Research / NMRC

ABSTRACT
This article illustrates improved methods of combining forecasts. Its objective is to describe how exponential smoothing methods are optimized within least squares equations. Traditionally, forecasts have been optimized first before being combined. Advantages of using a single seasonal index variable rather than multiple (11 for monthly) indicator variables are also illustrated. By using these improvements plus judgmental modification in unrestricted least squares equations, combined forecasts models are derived that rival all other econometric models in out-of-sample predictions. Hence, this paper describes improvements that produce practical models with increased out-of-sample accuracy, broader applicability, and judgment modification capabilities.
JEL Classification: C53

INTRODUCTION

In today's dynamic economy, each business must select the forecasting methods that help their particular situation. This forecasting dilemma is further complicated by the fact that most economic conditions are constantly changing. Therefore, the practice of combining forecasts that are conducive to a variety of economic conditions has gained popularity (Batchelor and Dua 1995). Unrestricted least squares is an extremely popular and highly regarded method of combining forecasts (Granger and Ramanathan 1984). The objective of this paper is to describe improvements to this effective forecasting method by optimizing exponential smoothing forecasts within an unrestricted least squares equation. This method is original with the authors and has yet to be explored. Also, the use of a single seasonal index variable rather than traditional indicator (dummy) variables has only recently been committed to writing (Landram. et al., 2004, 2008a) and is effectively employed in this article. These improved methods expand the capabilities of combined forecast models enabling them to become more practical and effective.

Consider the quarterly forecasting equation

$$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t \qquad (1)$$

where $\hat{Y}_t$ are the computed response values, X_t are for trend values (X_t = 1, 2, ... n), S_j are the quarterly seasonal indices repeated each year, and C_t are cyclical factors. In general, the accuracy of least squares predictions by (1) are superior to the accuracy obtained from the traditional time series decomposition method:

$$T_tS_jC_t = T_t{*}S_j{*}C_t, \quad (2)$$

where S_j and C_t are defined in (1) above. T_t are trend estimates; $T_t = b_0 + b_1X_t$. When the accuracy of (2) approximates that of (1), include the multiplicative term (2) in (1);

$$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t + b_4T_tS_jC_t. \quad (3)$$

Both additive and multiplicative relationships among time series components are described by (3). Although an additional degree of freedom is consumed, (3) will command more (never less) accuracy than its traditional decomposition counterpart described by (2). Stated differently, (3) provides a superior alternative to the traditional time series decomposition method of forecasting.

Exponential Smoothing

The inclusion of exponential smoothing measures as a time series cyclical component often brings additional out-of-sample accuracy to the model;

$$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t + b_4F_t, \quad (4)$$

where F_t represents simple exponential smoothing forecasts:

$$F_t = \alpha Y_{t-1}+(1-\alpha)F_{t-1} \quad (5)$$

However, exponential smoothing measures should be optimized within the least squares equation. A complete description of this optimization process is illustrated later by an example along with how it enhances the accuracy of out-of-sample predictions. Again, the practice of combining forecasts that are conducive to a variety of economic conditions will increase the accuracy of most predictions. When including forecasts from exponential smoothing methods, it is best to optimize these forecasts within the least squares framework. Although the simple exponential smoothing method is the easiest to illustrate and comprehend, all exponential smoothing methods can be combined and optimized within least squares equations. Since F_t and the cyclical factor C_t both measure cyclical movement, they are often collinear with one needing deletion.

Indicator Variables

Using indicator (dummy) variables, the approximate accuracy of (1) is obtained from (6):

$$\hat{Y}_t = b_0 + b_1X_t + b_2D_2 + b_3D_3 + b_4D_4 + b_5C_t, \quad (6)$$

where $D_j = 1$ if quarter j, j = 2, 3, 4, otherwise 0, and X_t and C_t are defined in (1). The indicator variable method of including seasonal variation is described in most statistical, forecasting, and econometric textbooks. However, the need for k-1 dummy variables to represent a k level qualitative variable brings severe limitations. Monthly seasonal variation must be represented by 11 dummy variables; 22 are needed if interaction is involved. Therefore, indicator variables in forecasting soon become impractical.

Benefits

The objective of this paper is to explain how greater accuracy is obtained by optimizing exponential smoothing forecasts within least squares models. Traditionally, these forecasts are optimized first and then included in the model. In achieving this objective, other innovative methods just beginning to frequent the literature are described. This paper brings further attention to employing the single seasonal index variable S_j in (1) rather than multiple indicator variables. Observe that S_j in (1) and (3) provide superior alternatives to the traditional decomposition method of forecasting. This model also possesses the judgmental capability for including structural breaks and event modeling variables. Using unrestricted least squares, these improvements produce practical forecasting models with increased accuracy for future predictions. Since these models are easy to derive and understand, they will quickly gain popularity in the workplace.

COMBINING FORECASTS

Granger and Ramanathan (1984**)** argue that combined forecasts from several methods outperform forecasts from a single method. They point out that values from discarded forecasting models still contain useful information about the underlying behavior of Y_t. When biased forecasts are included in a least squares equation, the intercept adjusts for the bias. Hence, it is important to use least squares equations with an intercept – unrestricted least squares. The authors totally agree with Granger and Ramanathan that the common practice of obtaining a weighted average of alternative forecasts should be abandoned in favor of least squares equations with an intercept. However, in this article the research of combining forecasts is carried one step further. Forecasts from naive time series methods are optimized within the least squares equation.

Concepts and Notations

The inclusion of X_t, S_j and C_t in (1) is easily justified. When combining forecasts, time series components (trend, seasonal, and cyclical) along with other forecasts generally make a significant contribution in explaining the behavior of the response variable Y_t. Of course, this assumes the model is not overspecified. Overfitting inflates forecast error variances (Landram et al, 2008b) and is the downside to combining forecasts. Therefore, out-of-sample criteria are employed (Landram, et al., 2005).

Least Squares Seasonal and Interaction

When seasonal variations changes with time, include an interaction variable between the cyclical C_t and the seasonal indices S_j:

$$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t + b_4C_tS_{t,} \qquad (7a)$$

where $C_tS_t .= C_t{*}S_t$. If X_t and C_t are held constant, a change in Y_t given a change in S_j is dependent upon the value of C_t:

$$\hat{Y}_t = [b_0 + b_1X_t + b_3C_t] + [b_2 + b_4C_t]S_j. \qquad (7b)$$

Similar statements may be made concerning the interaction between X_t and S_t. Observe, that a three way interaction term ($X_t{*}S_j{*}C_t$) approximates the decomposition of a multiplicative time series forecast described by (2). Barsky and Miron (1989) convincingly argue against deseasonalizing time series data. They maintains that considerable knowledge can be obtained from the interaction between seasonal and cyclical components.

Event Modeling

Forecasting macroeconomic time series is notoriously difficult. Structural breaks in the deterministic components are often the most damaging. An example in predicting future housing starts is discussed later. In this example, the following forecast model is employed with dummy variables to adjust for our recent decline in the housing market:

$$\hat{Y}_t = b_0 + b_1S_j + b_2F_t + b_3D_j + b_4S_jD_j + b_5H_j, \quad (8)$$

where H = 1 if $\geq$ quarter 4, 2005, otherwise 0 . The dummy variable H_j adjusts for the abrupt decline in the housing market. Also in an effort to make quarters 3 and 4 more pronounced, the dummy variable D is employed along with the interaction variable S_jD_j; where D_j = 1 if quarter 3 or 4, otherwise 0. The time series cyclical component C_t is collinear with the exponential smoothing F_t values and therefore deleted. The trend variable X_t was found insignificant and also deleted. Forecasting model (8) is explained later in greater detail along with the optimization of F_t within the least squares framework.

Statistical Modeling

Forecasting methods known to have different accuracies when used alone may perform quite differently when combined with other variables in a least squares equation. Here lies a major weakness of most combination of forecasts methods. Most forecasters assign their own prescribed weights thereby not recognizing the capability of unrestricted least squares. Additional insight into combined forecasts and into regression analysis in general is obtained by defining redundant and synergistic variables in the following manner.

(a) Redundant or multicollinear variables partially duplicate the information of other variables. Therefore, these variables possess insignificant partial t values and significant simple t values. They are significantly related with Y_t when used alone but can not be used effectively in a combination with other variables.

(b) Synergistic variables are the reverse. These variables possess significant partial t values but insignificant simple t values. They are insignificant when used alone but significantly related with Y_t when used in a combination with other variables. These variables make a unique contribution even though they are inaccurate when used alone. Furthermore, synergism in regression is more common than one might think (Shieh 2002).

(c) Other variables are significantly related with Y_t at both the partial and simple levels. The reverse is also true; some variables may be insignificant when used alone and when used in a combination with other variables.

The above definitions provide additional insight into regression analysis and in the rationale used when combining forecasts. A forecast may be highly accurate alone but inaccurate (redundant) when combined with other forecasts. Other forecasts may be inaccurate alone but highly accurate (synergistic) when used in a combination with other forecasts. Hence, when combining forecasts, let the unrestricted least squares model assign appropriate weights to the forecasts. This discussion explains why inaccurate forecasts are sometimes needed and why accurate forecast are sometimes not needed.

Comments

Since combined forecast models are subject to the same statistical modeling properties as other least squares regression models, consider the following comments.

1. Single forecasts should be optimized not with the objective of being used alone but with the objective of being used with other explanatory variables. Criteria used in this

optimization should not only be goodness of fit but also goodness of prediction criteria.

2. In regression the sum of squares error (SSE) cannot increase and usually decreases when additional variables enter the model. Therefore, R^2 is biased. As additional variables enter the model, R^2 becomes upward biased with regard to in-sample fitted values. This may lead to the model becoming over specified.

3. Combined forecast models are subject to the same bias of omission (specification error) as other regression models. Since the predicted values of various forecast methods are used as input variables when combining forecasts, emphasis should be placed on the proper selection and optimization of these methods.

4. In combining forecasts, variables are obtained from single forecasts, time series components, leading indicators, and econometric data. Also, be cognizant that these models possess event modeling, judgment modification, and even piecewise regression capabilities.

5. Kutner et al. (2005) wisely suggest that any violation of the classical regression assumptions be treated first as specification error and then as needing structural modification.

6. Since goodness of fit does not guarantee goodness of prediction, PRESS (Myers, 1990) and P^2 (Landram, et al. 2005) are used in measuring out-of-sample predictions.

Structured Judgmental Modification

A major problem with all forecasting methods is their inability to determine changing conditions in advance. Forecasting methods and time series values appropriate for one period are not necessary appropriate in another period. Therefore, structured judgmental modification is needed to help predict when these changes will occur and the effect these changes will have on the behavior of their forecasts.

Moving Seasonals

When using a constant seasonal index, it is assumed the seasonal variation is not moving -- is not becoming stronger or weaker. However, if a seasonal index for say quarter 1 possesses a trend, a moving seasonal index needs to be constructed. These effective, but often forgotten, indices are described in older textbooks (Croxton and Cowden, 1955) and employed when average seasonal indices do not adequately describe current seasonal variations. When forecasting future values of Y_t, moving seasonal indices may be obtained judgmentally. This allows the model to possess structured judgmental modification capabilities.

Judgmental Modification

While there is widespread acceptance that structured judgmental modification of statistical models improves forecasts, there are issues concerning how the process should be structured (Lawrence, Edmundson, and O'Connor 1986). Bunn and Wright (1991) remind readers that model specification, variable selection, how far back to go in a time series, and special event modeling are judgmental. The use of moving seasonal indices and optimizing exponential smoothing forecast within a least squares equation is in agreement with the structured visual aids promoted by Edmundson (1990). The idea is to obtain judgmental modification at the level of time series components. Thus, treatment of forecast values and time series components as explanatory variables in regression enables forecasters to employ structured judgment modifications at effective levels.

HOUSING STARTS: COMPARING FORECASTS

In an effort to convey ideas concerning the manner in which forecasts are optimized within a least squares equation, the following example is given. Housing starts is a strong indicator of our construction industry, public sentiment, and the health of our nation. Figure 1 reveals housing starts have little trend but a substantial cyclical movement. Although the declining US economy may have a slow recovery, this downward trend does not adequate predict future housing starts. Therefore, the combined forecasts model described in (8) is employed in predicting future housing starts. In Figure 1, predictions from (8) are shown by the line projected from 2008 through 2015.

Time Series Components

Table 1 reveals the relevance of the various time series components including the simple exponential smoothing component F_t. Note, R^2 increases from 0.671 to 0.919 and P^2 increases from 0.624 to 0.881when the exponential smoothing variable F_t is entered and optimized within the least squares equation. This optimization procedure is discussed below. As shown by (9) in Table 1, when the exponential smoothing forecasts are optimized first ($\alpha = 0.39$) and then entered into the model, R^2 and P^2 increase to 0.842 and 0.821, respectively.

TABLE 1

COMPARING FORECAST MODELS

Forecasting Model		R^2	Adj $\bar{R}^2$	PRESS	P^2
(5)	$F_t = \alpha Y_{t-1}+(1-\alpha)F_{t-1}$	0.454	0.454		
(1)	$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t$	0.671	0.660	206,280	0.624
(9)	$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t + b_4F_t^*$	0.842	0.833	98,573	0.821
(4)	$\hat{Y}_t = b_0 + b_1X_t + b_2S_j + b_3C_t + b_4F_t$	0.919	0.915	63,505	0.881
(8)	$\hat{Y}_t = b_0 + b_1S_j + b_2F_t + b_4D_j + b_4S_jD_j + b_2H_j$	0.948	0.945	33,233	0.940
	F_t^* where $\alpha = 0.39$	F_t where $\alpha = 0.95$			

Modeling

Both a variable elimination procedure and event modeling are employed to obtain (8):

$$\hat{Y}_t = b_0 + b_1S_j + b_2F_t + b_3D_j + b_4S_jD_j + b_5H_j, \tag{8}$$

The constant seasonal index S_j needed help in accentuating seasonal variations. This assistance is obtained from the dummy variable (D = 1 if quarter 3 or 4) that underscored seasonal variability. As stated above, H_j, is used to adjust for our recent decline.

Optimization

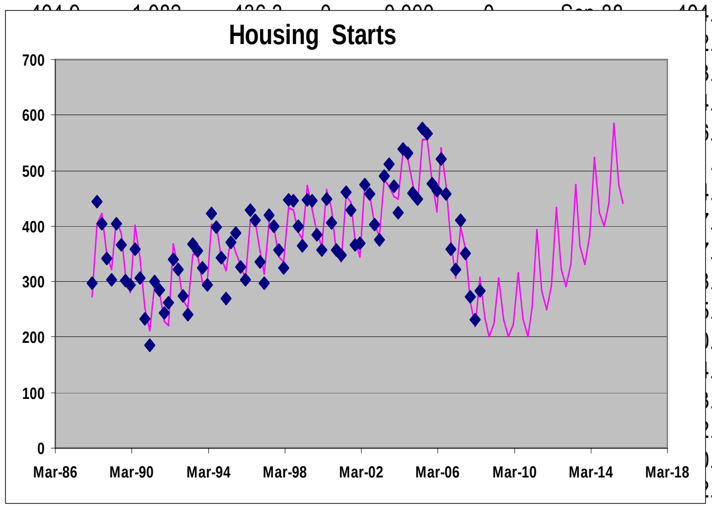
The optimization of exponential smoothing methods within a least squares structure is conducive to spreadsheet operations. Figure 1 depicts these operations with the simple exponential smoothing equation (5) being placed in column D. Let the value of α be in cell D2. Let R^2 and adjusted R^2 be in say cells B2 and B3. These values are computed in another spreadsheet space or worksheet and merely referenced in cells B2 and B3. The PRESS and P^2 statistics are also referenced in a similar manner. Although this procedure is illustrated with spreadsheet operations, it can easily be conducted on statistical software. Each time one changes the value of α, spreadsheet operations DATA>DATA_ANALYSIS>REGRESSION are performed and R^2 is examined. Using this procedure R^2 is maximized when α = 0.95. Table 1 shows that α = 0.39 when R^2 for (5) is optimized and the F_t values are entered into (9). The value of α becomes 0.95 when R^2 is optimized within (4) and (8).

Figure 1. Housing Starts (1988 to Q2, 2008)

R^2 = 0.948 α = 0.95 **PRESS=** 33,233

Adj R^2 = 0.945 P^2 = 0.944

Dates	Y	S_j	F_t	D	SD	**H**	Dates	Y	Y^
Mar-88	297.2	0.857	297.2	1	0.857	0	Mar-88	297.2	272.4
Jun-88	443.6	1.137	297.2	1	1.137	0	Jun-88	443.6	404.0
Sep-88	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	423.2
Dec-88	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	349.9
Mar-89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	321.0
Jun-89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	412.6
Sep-89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	386.1
Dec-89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	311.4
Mar-90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	280.3
Jun-90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	401.9
Sep-90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	341.3
Dec-90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	252.5
Mar-91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	211.8
Jun-91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	294.2
Sep-91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	281.5
Dec-91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	228.3
Mar-92	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	220.1
Jun-92	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	367.8
Sep-92	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	323.1
Dec-92	274.9	0.924	322.8	0	0.000	0	Dec-92	274.9	265.9
Mar-93	240.6	0.857	277.3	1	0.857	0	Mar-93	240.6	252.4
Jun-93	367.2	1.137	242.4	1	1.137	0	Jun-93	367.2	349.0
Sep-93	355.6	1.082	361.0	0	0.000	0	Sep-93	355.6	347.6
Dec-93	324.3	0.924	355.9	0	0.000	0	Dec-93	324.3	299.1



Autocorrelation

Note that all exponential smoothing equations possess a residual lag component; (5) may be written as $F_t = \alpha e_{t-1} + F_{t-1}$, where $e_{t-1} = (Y_{t-1} - F_{t-1})$. Residual lag components help eradicate autocorrelation as do moving seasonal indices and cyclical components. Although there is no guarantee, there is a great likelihood that combined forecast models that include an exponential smoothing variable will be autocorrelation free.

DISCUSSION AND CONCLUDING REMARKS

The improved methods of forecasting discussed above produce combined forecast models with increased accuracy, greater applicability, and judgmental modification capabilities. Be cognizant that the seasonal index variable S_j provides a superior alternative to the traditional time series decomposition forecasting method described by (2). The S_j variable is also superior to (6)—describing seasonal variation with dummy variables.

Parsimonious Models

As compared to other combined forecast methods that employ weights, unrestricted least squares produce the most accurate in-sample fitted values (Granger and Ramanathan 1984). Hence, the preferred means of improving these combined forecast models are to maintain the least squares in-sample accuracy and employ judgmental modification variables to obtain greater out-of-sample accuracy. However, the belief that accuracy is increased by combining a large number of forecasts is not true for out-of-sample predictions when the model is overspecified. Overfitting inflates forecast error variances and diminishes the accuracy of point estimates (Landram, et al 2008b). This adds credence to Zellner's (1991) KISS principle of "keep it sophisticatedly simple." In general, parsimonious models yield more accurate out-of-sample predictions. Hansen (2007) maintains that quality of prediction is inversely correlated with quality of fit. Therefore, model selection criteria are employed to prevent overfitting.

Implications

Landram et al. (2008a) revealed that the use of time series components within least squares equations far exceeds the accuracy obtained by using Solver--a spreadsheet linear programming algorithm. These models are also superior to (a) traditional time series decomposition methods and (b) models that describe seasonal variation with dummy variables. Furthermore, they are easier to derive and interpret than Box-Jenkins ARIMA models. Indeed, these models rival all econometric and time series models in predictive accuracy. Given the above attractive attributes, these models will quickly be adopted in the classroom and more importantly in the workplace.

CONCLUSION

The intent of this study is to describe how an exponential smoothing variable F_t is optimized within a least squares equation. This improved method of forecasting produces combined forecast models that rival all econometric and time series models in both in-sample and out-of-sample accuracy. The likelihood of autocorrelation is also diminished when an exponential smoothing variable is included in the model. Forecasting is a vital ingredient in all facets of business. Indeed, all budgeting and planning operations begin with assumed accurate forecasts. Therefore, the improved methods described above will make the combination of forecasts increasingly popular in the business community.

REFERENCES

Barshy, R. and J. Miron. 1989. “The Seasonal Cycle and the Business Cycle.” *Journal of Political Economy* 97: 503-534.

Barton, H.C. 1941. "Adjustment for Seasonal Variation." *Federal Reserve Bulletin* June 1941: 23-38.

Bohara, A., R. McNown and J. Batts. 1987. "A Re-evaluation of the Combination and Adjustment of Forecasts." *Applied Economics* 19: 437-455.

Batchelor, R., P. Dua. 1995. “Forecast Diversity and the Benefits of combining Forecasts.” *Management Science* 41: 68-75.

Bopp, A.E. 1985. "On Combining Forecasts: Some Extensions and Results." *Management Science* 31: 1492-1498.

Bunn, D. and G. Wright. 1991. "Interaction of Judgemental and Statistical Forecasting Methods: Issues & Analysis." *Management Science* 37: 501-518.

Croxton, Frederick E. and D. J. Cowden. 1955. *Applied General Statistics*. Englewood N.J.: Prentice-Hall.

DeLurgio, S. 2008. *Forecasting Principles and Applications*. Boston: Irwin-McGraw.

Edmundson, R. 1990. "Decomposition: A Strategy for Judgemental Forecasts." *Journal of Forecasting* 9 (4): 305.

Granger, C. and R. Ramanathan. 1984. "Improved Methods of Combining Forecasts." *Journal of Forecasting* 3: 197-204.

Hansen, P. 2007. “Criteria-Based Shrinkage for Forecasting.” *Fifth European Central Bank Workshop on Forecasting Techniques: Forecast Uncertainty in Macroeconomics and Finance*.

Kutner, M., C. Nachtsheim, J. Nete and W.Li. 2005. *Applied Linear Statistical Models*, 5th edition. Boston, Irwin-McGraw.

Landram, F., A. Abdullat and V. Shah. 2004. “Using Seasonal and Cyclical Components in Least Squares Forecasting Models.” *Southwestern Economic Review* 31: 189-196.

Landram, F., A. Abdullat and V. Shah. 2005. “The Coefficient of Prediction for Model Specification.” *Southwestern Economic Review* 32: 149-156.

Landram, F., R. Pavur and B. Alidaee. 2008a. “Combining Time Series Components for Improved Forecasts.” *Decision Sciences: Journal of Innovative Education* 6: 197-204.

Landram, F., R. Pavur, and B. Alidaee. 2008b. “An Algorithm for Enhancing Spreadsheet Regression with Out-of-Sample Statistics.” *Communication in Statistics: Simulation and Computation* 37: 1578-1592.

Lawrence, M. J., Edmundson, R. H. and O'Connor M. J. 1986. "The Accuracy of Combining Judgmental and Statistical Forecasts." *Management Science* 32: 1521-1532.

Myers, R. 1990. *Classical and Modern Regression with Application*, 2nd edition, Boston: Duxbury Press.

Ramanathan, R. 2002. *Introductory Econometrics with Applications*, 5th edition, Mason, Ohio: South-Western.

Shieh, Gwowne. 2001. "The Inequality Between the Coefficient of Determination and the Sum of Squared Simple Correlation Coefficients." *The American Statistician* 42: 121-124.

Zellner, A. 1992. “Statistics, Science and Public Policy.” *Journal of the American Statistical Association* 87: 1-6.