Choosing Among Competing Econometric Forecasts:
Regression Based Forecast Combination Using Model Selection [*]

Norman R. Swanson and Tian Zeng

Pennsylvania State University

August 1997

ABSTRACT

Based on Monte Carlo simulations using both stationary and nonstationary data, a model selection approach which uses the SIC to select a "best" group of forecasts in the context of forecast combination regressions dominates a number of other techniques, including the standard t-statistic approach and the simple averaging rule. Our results are robust across various restricted and unrestricted versions of standard linear forecast combination specifications, and are not affected by the use of differenced or levels data. In addition to the Monte Carlo experiments, we conduct a real-time forecasting experiment using sequences of forecasts of nine macroeconomic variables. The forecasting models which we examine include linear VAR models, nonlinear artificial neural network models, and the Survey of Professional Forecasters (SPF). Based on *ex ante* performance, our empirical evidence suggests that the SIC dominates the other forecast combination strategies which are examined. Furthermore, when the SIC approach is used in the context of forecast encompassing type regressions, we conclude that "individual" forecasts are often preferred to combined weight forecasts, and that the SPF model and the econometric models each win approximately the same number of times. We also illustrate that collinearity among competing forecasts arises, even when only three competing forecasts are used in a linear forecast combination regression, and suggest that this is one of the reasons why information criteria are useful for constructing combined weight forecasts. This point is illustrated as t-statistics used to test the significance of alternative forecasts of simulated as well as actual data are often below a 10% critical value, suggesting that either none, or all competing forecasts should be used to construct combined weight forecasts, for example.

KEYWORDS: forecast combination, Akaike and Schwarz Information Criteria, ex ante forecasting, unrevised data, artificial neural networks, Survey of Professional Forecasters

JEL Classification: C12,C22,C45,C53.

Correspondence to: Norman R. Swanson, Department of Economics, College of the Liberal Arts, Pennsylvania State University, 521 Kern Graduate Building, University Park, PA 16802. Email: nswanson@psu.edu

1. Introduction

Simplicity and ease of application has always been considered desirable in the formulation of statistical models. This is readily apparent in the forecasting literature, for example, where many applications make use of forecast combination and related forecast encompassing techniques, which are both theoretically appealing, as well as easy to apply. Perhaps for both of these reasons, the number of studies which examine forecast combination is vast. Overviews of many of the most popular techniques and methods are contained in Granger and Ramanathan (1984), Clemen (1989), Granger (1989), and Diebold and Lopez (1995), and the references contained therein. In this paper we consider variations of what might well be referred to as one of the "workhorses" of linear forecast combination. Consider

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i f_{i,t} + \varepsilon_t \tag{1.1}$$

where y_t is some economic variable of interest (which may be integrated of order zero or one, say), and $\{f_{i,t}\}$ is a sequence of forecasts of y_t made prior to time t, and constructed using some model, i, $i = 1,...,p$. It is well known that when p is large, the extent of collinearity among competing forecasts may be so severe as to effectively cripple the ability of least squares to provide reasonable estimates of the weighting coefficients. This drawback of the regression approach is one of the potential problems which we address. In particular, we examine the ability of Schwarz and Akaike Information Criteria (SIC: Schwarz (1978) and Rissanen (1978), AIC: Akaike (1973, 1974), respectively) to effectively pre-sort the $\{f_{i,t}\}$ sequences into groups of "information rich" and "information poor" forecasts. Then, some type of forecast combination can be carried out using only those forecasts from the "information rich" group, hence alleviating to some extent any collinearity problems. Since the model selection approach taken here is based on exposing irrelevant forecasts, the techniques can also be interpreted as a type of forecast encompassing test (in mean).

Although one of our main goals is to address collinearity among forecasts, our efforts can also be interpreted as an effort to compare alternative forecast combination techniques. Thus, while we often focus on collinearity in our discussion, it should be stressed that the model selection approach taken here applies more generally, as well. In this sense, the results are meant to provide broad additional evidence of the usefulness of model selection in forecast combination. Given this interpretation of our approach, it is

worth noting that Schmittlein, Kim, and Morrison (1990) used Monte Carlo simulation to examine the use-fulness of the AIC relative to mean square error for forecast combination. Our paper differs from theirs in a number of ways. For example, in addition to the AIC, we also examine and compare SIC, t-statistic, simple averaging, and simple least squares combination techniques. Also, we examine the combination tech-niques using not only Monte Carlo experimentation, but also a real time forecasting experiment based on actual data.

In the context of optimal forecast combination, we also suggest other model selection based tech-niques which are based on *ex ante* forecast evaluation. For example, we suggest that a partition of the competing forecasts into the two groups discussed above may be achieved by examining the performance of different combinations of forecasts in a real-time forecasting scenario. Such an approach requires separating the $\{f_{i,t}\}$, sequences into two samples, one to be used for rolling in-sample forecast combination regressions and *h*-step ahead *ex ante* forecast construction, and the other to be used for real-time forecast combination evaluation. Of note is that this approach is substantially more computer intensive than the "in-sample" based SIC and AIC approaches, and as such may be more costly to implement.

As discussed in Swanson and White (1995, 1997), the model selection approach taken here does not rely on classical testing theory, in the sense that the traditional approach of fixing the test size (given that the limiting size is known) and rejecting the null hypothesis at that size regardless of sample size is not adopted. This feature is particularly appealing when model selection is used as an alternative to some sort of t-statistic based tests to select "best" or encompassing forecasts. An example of the use of t-statistics, in the current context, is given in Fair and Shiller (1990), where the information in *h*-step ahead forecasts from a number of econometric models is compared by examining heteroskedasticity and serial correlation consistent standard errors and t-statistics. As the model selection approach is not based on the use of t-statistics, the type of collinearity problems mentioned above are avoided. Thus, the SIC and AIC criteria, for example, have two separate and appealing uses in the current context. First, the criteria can be used to narrow down any large group of forecasts *before* standard forecast combination equations are estimated, and optimal forecasts constructed. Second, the criteria can be used to separately compare the information in competing forecasting models. Furthermore, one might combine standard hypothesis testing with model

selection by first pre-selecting a group of "information rich" forecasts using model selection, and then by using t-statistics to further narrow down the group to some "optimal" size. Another desirable feature of the model selection approach is that it does not require the specification of a correct model for its valid application as does the traditional hypothesis testing approach. A limitation of the model selection approach is that it can sometimes be difficult to assess the Type I error associated with testing the implicit assumption that two models being considered perform equally well based on observed differences in realized model selection criteria. This defect is of the same order as using a traditional test whose size is known only asymptotically.

Our results are rather unambiguous. Based on extensive Monte Carlo simulations using both stationary and nonstationary data, a model selection approach which uses the SIC to select a "best" group of forecasts in the context of forecast combination regressions dominates a number of other techniques. The other techniques considered include, among others, the t-statistic approach used by Fair and Shiller (1990), and a simple averaging rule. Interestingly, the dominance of the SIC approach is not affected when we use various alternative types of linear forecast combination regression models, and when we use nonstationary rather than stationary data.

In addition to the Monte Carlo experiments, we conduct a real-time forecasting experiment using sequences of forecasts of nine macroeconomic variables. The forecasting models which we examine include linear VAR models, nonlinear artificial neural network models, and the Survey of Professional Forecasters. Based on *ex ante* performance, our empirical evidence suggests that the SIC dominates the other linear forecast combination strategies which are considered. Furthermore, when the SIC approach is used in the context of forecast encompassing type regressions, we conclude that "individual" forecasts are often preferred to combined weight forecasts, and that the SPF model and the econometric models each win approximately the same number of times. We also note that collinearity among competing forecasts is great, even when only three competing forecasts are used in a linear forecast combination regression. This point is illustrated as t-statistics used to test the significance of alternative forecasts (e.g. see Fair and Shiller (1990)) are below a 10% critical value for 8 of the 9 macroeconomic variables which we examine. This in turn strengthens the case for using the SIC to "weed out" *extraneous* forecasts when constructing

combined weight forecasts. Further discussion of this type of procedure, whereby forecasts are de-selected because they add insufficient additional new information, is given in Clemen (1985), where the term *extraneous* is coined.

By adopting a model selection approach to forecast combination, we believe that we contribute not only to the discussion of the usefulness of different forecast combination strategies, but also the the methodology of forecast combination. One dimension of this contribution is that we address the issue of collinearity among competing forecasts, and offer a surprisingly simple model selection based approach for dealing with the problem. Contributions are also attempted in several other related interesting directions. For instance, we compare adaptive and nonadaptive linear and nonlinear econometric models with professional forecasts, in the context of encompassing type regressions. We also examine the usefulness of "parsimony" in the construction of both individual as well as combined weight forecasts, and suggest that while "parsimony" is generally a good rule for constructing individual forecasts, the opposite may be the case if one's goal is the construction of "combined weight" forecasts.

The rest of the paper is organized as follows. Section 2 expands upon the types of forecast combination and encompassing regressions which are considered below. The third section outlines a series of Monte Carlo experiments, and discusses the results from the experiments. The fourth section examines the proposed methods in an empirical setting, by comparing and contrasting a group of linear and nonlinear econometric forecasts with professional forecasts for nine different macroeconomic variables. Section 5 concludes.

2. Forecast Combination

For ease of exposition, assume that there are p unbiased one-step ahead forecasts of y_t, say $f_{i,t}$, $i = 1,...,p$. The forecasts are combined using regression weights which are estimated using regressions of the basic form

$$y_{t+1} = \alpha_0 + \sum_{i=1}^{p} \alpha_i f_{i,t} + \varepsilon_{1t} \qquad\qquad \text{(FC1-L)}$$

$$y_{t+1} = \sum_{i=1}^{p} \alpha_i f_{i,t} + \varepsilon_{2t} \qquad\qquad \text{(FC2-L)}$$

$$y_{t+1} = \sum_{i=1}^{p} \alpha_i f_{i,t} + \varepsilon_{3t} \quad , \quad \sum_{i=1}^{p} \alpha_i = 1 \qquad\qquad \text{(FC3-L)}$$

These are the three basic regression models considered by Granger and Ramanathan (GR: 1984), and versions of them date back to Bates and Granger (1969). Although (FC3-L) has been used extensively in practice, GR show that (FC1-L) is usually preferable, even if the different forecasts are all unbiased. This is because the forecast errors associated with combination weighted forecasts based on (FC3-L) may be correlated with the individual $f_{i,t}$ series. This in turn means that the forecast errors can themselves be forecast using $f_{i,t}$, so that the combination is not optimal. However, moving to (FC2-L) loses the unbiasedness property, and thus (FC1-L) is preferred. It should perhaps be noted that the optimality results derived in Bates and Granger (1969) and other earlier papers depend sometimes crucially on the unbiasedness assumptions. In such cases, estimation techniques also become important.

Diebold (1988) and Coulson and Robbins (1993), on the other hand, show that ε_{1t} may be an ARMA process, in particular when y_t is serially correlated. Diebold suggests that the time series properties of the residuals should be identified, and the forecast combination regression appropriately re-estimated. Coulson and Robbins further note that the inclusion of the lagged dependent variable as an additional regressor may be called for. Both of these approaches are relevant, and are admitted within the framework discussed below.

If the data used are integrated processes, such that their differences are stationary, GR and Fair and Shiller (1990) suggest that differences should perhaps be used in the combination regressions. Their argument is clearly relevant if t-statistics are to be examined, but is also relevant if the forecaster is interested in growth rates rather than levels forecasts, say. To allow for difference stationary data, we also consider three obvious variations of the above regression models (e.g. see Fair and Shiller (1990)). They are:

$$y_{t+1} - y_t \ = \ \alpha_0 + \sum_{i=1}^{p}\alpha_i\,(f_{i,t} - y_t) + \varepsilon_{1t} \tag{FC1-D}$$

$$y_{t+1} - y_t \ = \ \sum_{i=1}^{p}\alpha_i\,(f_{i,t} - y_t) + \varepsilon_{2t} \tag{FC2-D}$$

$$y_{t+1} - y_t \ = \ \sum_{i=1}^{p}\alpha_i\,(f_{i,t} - y_t) + \varepsilon_{3t} \ , \ \ \sum_{i=1}^{p}\alpha_i = 1 \tag{FC3-D}$$

Many generalizations of the simple regressions considered here abound in the literature. For instance, Engle, Granger, and Kraft (1984) and Diebold and Pauly (1987) examine time-varying combining weights, and Deutsch, Granger, and Teräsvirta (1994) consider nonlinear combining regressions. It

should be noted, though, that the model selection combination techniques discussed in the following paragraphs can easily be modified to encorporate these and other related generalizations.

2.1 Forecast Combination Using Akaike and Schwarz Information Criteria

The first model selection approach to forecast combination which we consider is based on complexity penalized likelihood criteria. In particular, we consider the AIC and the SIC.

$$AIC = T \log s^2 + 2k \ ,$$
$$SIC = T \log s^2 + k \log T \ ,$$

where k is the total number of parameters in the equation of interest, and s^2 is the regression mean-squared error. The procedure begins by estimating regressions of the form given by equations (FC1-L)-(FC3-L) and (FC1-D)-(FC3-D). In total,

$$\sum_{i=1}^{p} \frac{p!}{(p-i)!i!} = TR$$

regressions should be run, so that each possible linear combination of the p competing forecasts are considered. (Nonlinear combinations are clearly also possible, but are not considered here.) Then, model selection criteria based on each regression are calculated, and the "best" group of forecasts is selected. Combined weight forecasts can then be constructed based directly on a secondary regression which uses the selected group of forecasts. Alternatively, the second regression could be further used to pare down the number of regressors by using some sort of robust t-statistic to test whether the relevant coefficients are zero. In this case, a third regression could be run using the "final" set of forecasts in order to estimate a "final" set of combining weights. In addition to suggesting which forecast weights should be set to zero, this approach offers a direct test of informational content of the competing forecasts. If the "best" group of forecasts does not include a particular subset of forecasts, then that subset may loosely be said to be encompassed by the "best" group, at least in mean. However, as noted by Fair and Shiller (1990), this last interpretation requires that forecasts make no use of future information. This assumption may be crucial in many practical applications, and is discussed in more detail below.

As mentioned above, this approach does not require specification of a correct model for its valid application. Another desirable feature is that the probability of selecting the truly best model approaches one as the sample size increases, if the model-selection approach is properly designed. This is contrary to

the standard practice of fixing a test size, and rejecting the null hypothesis at that fixed size, regardless of sample size. Further, Granger, King, and White (1995) suggest that although standard hypothesis testing has a role to play in terms of testing economic theories, it is more difficult to justify using standard hypothesis tests for choosing between two competing models. One reason for their concern is that one model must be selected as the null, and this model is often the more parsimonious model. However, it is often difficult to distinguish between the two models (because of collinearity, near-identification, etc.), so that the null hypothesis may be unfairly favored. For example, it is far from clear that pre-test significance levels of 5% and 1%, say, are optimal, as pointed out Fomby and Guilkey (1978) in the context of Durbin-Watson tests. The model selection approach neatly avoids these types of sticky issues, and thus offers a potentially interesting alternative to standard hypothesis testing.

In another recent paper, Sin and White (1995) consider the use of penalized likelihood criteria for selecting models of dependent processes. In the context of strictly nested, overlapping or non-nested, linear or nonlinear, and correctly specified or misspecified models they provide sufficient conditions on the penalty to ensure that the model selected attains the lower average Kullback-Leibler Information Criterion, with probability (approaching) one. Further, as special cases, their results describe the Akaike and Schwarz Information Criteria.

In the context of forecast combination, if it turns out that these model selection criteria perform well, then we have direct evidence that model selection criteria offer a valid alternative to standard hypothesis testing approaches to testing encompassing restrictions, and also offer a reasonable approach for dealing with collinearity, when a large number of competing forecasts are being examined. Further, as model selection statistics are very easy to calculate, regardless of the properties of the data, evidence of the usefulness of AIC and SIC criteria should be of some interest to forecasters.

2.2 Forecast Combination Using an Ex Ante Forecasting Approach

An alternative model selection approach to forecast combination is based on *ex ante* forecasting. This approach is analogous to the use of the AIC and SIC, except that the "best" group of forecasts is selected using rolling windows of data in an *ex ante* forecast combination scenario. As above, assume that a sample of competing forecast sequences, each of size T observations is available. Start with some frac-

tion of *T*, say *qT* forecasts, 0<*q*<1. The procedure can then be summarized in five steps.

Step (1): Run TR regressions of the form given by one of equations (FC1-L)-(FC3-L) and (FC1-D)-(FC3-D), where TR is the complete set of all possible linear combining regressions, as discussed above.

Step (2): Form TR *h*-step ahead combined weight forecasts and forecast errors, where *h* is some pre-selected forecast horizon.

Step (3): Roll the sample forward one period (either keeping a fixed sample size, or by augmenting the previous sample with one new observation). Then, construct another TR real-time forecasts and forecast errors by repeating steps (1) and (2).

Step (4): Continue the above procedure until the entire sample is exhausted, resulting in TR sequences of *ex ante* forecasts and forecast errors, say $\{f_{j,k}^*\}_{k=1}^{k=T-qT}$ and $\{e_{j,k}^*\}_{k=1}^{k=T-qT}$, j = 1,...,TR.

Step (5): Choose a selection criterion with which to compare the sequences of forecasts and/or forecast errors and "select" the model from among the TR alternatives which optimizes the criterion. Examples of possible criteria are mean squared out-of-sample forecast error, mean absolute forecast error deviation and turning point predictive accuracy. In general, the criterion which is used should be tailored to the specific end use of the forecasts.

This approach results in the selection of some "best" group of forecasts, and also offers an alternative to the encompassing type tests based on the AIC and the SIC which are discussed above. For instance, we can compare the predictive accuracy of two competing forecasts, say *f* 1 and *f* 2. The three alternative regressions run in Step 1 would use (i) *f* 1, (ii) *f* 2, and (iii) both *f* 1 and *f* 2. The outcome would be the selection of either one of the two forecasts, or both. This approach is recognizable as a direct model selection analog to the classical predictive accuracy tests discussed by Granger and Newbold (1986), Mizrach (1991) and Diebold and Mariano (1995), except for the potentially useful feature that more than two forecasts, say *f* 1, ..., *fn*, can easily be compared.

Engle and Brown (1986) consider the related problem of comparing penalized mean squared forecast errors from competing forecasting models of electricity demand using among others, the AIC and SIC. Our approaches differ from theirs in a number of ways. First, we emphasize forecast combination and optimal weight selection. Second, we examine groups of forecasts, and admit that any single forecasting

model may not be selected as "best" if combinations of competing forecast sequences are examined in addition to single forecast sequences. Third, we attempt to deal with the issue of collinearity among forecasts. Finally, we consider "in-sample" likelihood criteria as well as "out-of-sample" or *ex ante* criteria.

In our simulations we are able to implement a truly *ex ante* forecasting procedure. However, as is well known, the use of real-time forecasting methods becomes suspect when actual economic data which are used have been subjected to two-sided moving average filtering (as is the case with most seasonal adjustment filters), periodic rebasing, and periodic revision, for example. Thus, we again stress that these and related problems may make valid implementation of the types of tests discussed above quite difficult in practice. It should be noted, though, that many research groups and individuals are currently working extensively on constructing real-time macroeconomic data bases (e.g. see the discussion in Swanson (1996)). Furthermore, in our empirical examination we ensure that real-time data are used by constructing a real-time dataset.

3. Monte Carlo Experiments

3.1 Experimental Design

Our Monte Carlo experiments are all based on 5000 simulations, and begin with variations of the following data generating process (DGP):

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \quad , \quad u_t = \eta_t + \theta\eta_{t-1}$$

where $\eta_t \sim IN(0,1)$ and $\beta_0 = 0.1$. As a starting point, data are generated using the following parameter combinations:

$$\left[\beta_1 = 0.5 \ , \ \theta = 0.3\right] \tag{DGP1}$$

$$\left[\beta_1 = 0.5 \ , \ \theta = 0.0\right] \tag{DGP2}$$

$$\left[\beta_1 = 1.0 \ , \ \theta = 0.3\right] \tag{DGP3}$$

$$\left[\beta_1 = 1.0 \ , \ \theta = 0.0\right] \tag{DGP4}$$

Both I(1) and I(0) data are generated, and a moving average error component is present in (DGP1) and (DGP3). Since u_t is an invertible MA process, y_t in (DGP1) and Δy_t in (DGP3) can be expressed as infinite order AR processes. In our experiments, we construct forecasts using a variety of linear and nonlinear AR

models, which either (i) nest the true specification (when the data are generated according to (DGP2) or

(DGP4)) or (ii) do not nest the true specification (when the data are generated according to (DGP1) or

(DGP3)). In particular, one (h=1) step-ahead forecasts are constructed using the following models:

$$y_{t+h} = \gamma_0 + \gamma_1 y_t + \nu_{t+h} \tag{F1}$$
$$y_{t+h} = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \nu_{t+h} \tag{F2}$$
$$y_{t+h} = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 y_{t-2} + \nu_{t+h} \tag{F3}$$
$$y_{t+h} = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \gamma_3 y_{t-2} + \cdots + \gamma_p y_{t-p+1} + \nu_{t+h} \tag{F4}$$
$$y_{t+h} = \gamma_0 + y_t + \nu_{t+h} \tag{F5}$$
$$y_{t+h} = \gamma_0 + \gamma_1 \mid y_t \mid / \mid y_t + 2 \mid + \nu_{t+h} \tag{F6}$$
$$y_{t+h} = \gamma_0 + \gamma_1 \ln(y_t^2) + \nu_{t+h} \tag{F7}$$

The last two models are used to produce "strawman" forecasts. If either of the last two models is preferred

in any experiment, then we have direct evidence that the particular forecast combination method being

used is not robust to the inclusion of possibly "nonsense" forecasts, for example (see discussion below). It

should be noted, though, that combination experiments which include data generated by (F1)-(F5) (i.e.

excluding the "nonsense" models) are performed for each DGP which we consider. Thus, our simulation

results are not driven by our use of "nonsense" models in certain cases. In model (F4), p is selected using

an in-sample SIC. An alternative would be to use the AIC for lag selection, say, rather than the SIC. How-

ever, as it is well known that more parsimonious models often perform better in a forecasting scenario, we

use the SIC. In all experiments, models (F1)-(F7) are estimated using a training set of 100 realizations of

the true DGP and an h-step ahead forecast is constructed. The sample is then moved forward one period,

the forecasting model is re-estimated, and another h-step ahead forecast is constructed. This process is

repeated 100 times, yielding sequences of 100 1-step ahead forecasts. These forecasts are then used in

forecast combination regressions, as well as for *ex ante* analysis based on the realizations of the true DGP

which correspond to the forecast sequences.

To be more specific, the generated sequences of forecasts are used in the forecast combination

regressions given above by (FC1-L)-(FC3-L) and (FC1-D)-(FC3-D). All of the combination regressions of

this form use three different sequences of forecasts, selected from among (F1)-(F7). For each DGP given

by (DGP1)-(DGP4), four different variations of these three different sequences are tried. In particular, for

DGP1, the combination regressions (see (C1)-(C7) below) are based on forecasts using the following

models: [1] (F2), (F4), and (F5); [2] (F1), (F2), and (F6); [3] (F2), (F4), and (F7); and [4] (F4), (F6), and (F7). For DGP2, the variations tried are: [1] (F1), (F2), and (F5); [2] (F1), (F2), and (F6); [3] (F1), (F5), and (F7); and [4] (F1), (F6), and (F7). For DGP3, the variations tried are: [1] (F1), (F2), and (F5); [2] (F1), (F2), and (F6); [3] (F2), (F4), and (F5); and [4] (F4), (F6), and (F7). Finally, for DGP4, the variations tried are: [1] (F1), (F2), and (F5); [2] (F1), (F5), and (F6); [3] (F2), (F5), and (F7); and [4] (F5), (F6), and (F7). In total, 192 experiments are carried out. They can be broken down as 4 ((DGP1)-(DGP4)) x 4 (combinations of three forecasts from (F1)-(F7) used in combination regressions) x 2 (forecast horizons - h=1,4) x 6 (forecast combination regressions (FC1-L)-(FC3-L) and (FC1-D)-(FC3-D)) = 192.

It should be noted that based on comparisons of this sort it may not be reasonable to assume that the forecasting model which should be picked is an AR(1), say, even if the true DGP is (DGP1) or (DGP3). Although this statement is perhaps surprising, our evidence suggests that based on mean squared forecast error (MSE), mean squared absolute forecast error deviation (MAD) and mean absolute forecast percentage error deviation (MAPE) measures the AR(1) model is not always preferred, when such calculations are based on our out-of-sample period of 100 observations (see discussion of results). This in turn suggests that if we simply run our 196 experiments and then choose as "best" the forecast combination technique which "chooses" the AR(1) model most often, we may be in error, particularly if our objective function involves producing minimum MSE real-time forecasts, for example. To address this type of problem, we tabulate the out-of-sample MSE, MAD, MAPE for all of the forecast combinations which are chosen as "best".

In order to provide a rough benchmark as to how well our model selection approaches are performing, we select the "best" group of variables by using 5 different combination techniques. They are:

(M-AVG): This is the simple averaging rule which constitutes constructing the arithmetic average of the competing forecasts. This rule may be taken as the benchmark in this study, as it is clearly the easiest to apply, and as there is some evidence that simple averages often outperform more complicated combination techniques (e.g. see Russell and Everett (1987) and Hallman and Kamstra (1989)).

(M-SIC): Use the SIC to compare $TR = \sum_{i=1}^{p} \frac{p!}{(p-i)!i!} = 7$ (p =3), regressions for each combination regression type ((CR1-L)-(CR3-L) and (CR1-D)-(CR3-D)). Choose the "optimal" regression based on the SIC.

The regressors from this "optimal" regression form our "best" group of forecast sequences. For example, the TR=7 regressions for combination regression type (FC1-L) are

$$y_t = \alpha_0 + \alpha_1 f_{F1,t} + \alpha_1 f_{F2,t} + \alpha_1 f_{F3,t} + \varepsilon_{1t} \tag{C1}$$

$$y_t = \alpha_0 + \alpha_1 f_{F1,t} + \alpha_1 f_{F2,t} + \varepsilon_{1t} \tag{C2}$$

$$y_t = \alpha_0 + \alpha_1 f_{F1,t} + \alpha_1 f_{F3,t} + \varepsilon_{1t} \tag{C3}$$

$$y_t = \alpha_0 + \alpha_1 f_{F2,t} + \alpha_1 f_{F3,t} + \varepsilon_{1t} \tag{C4}$$

$$y_t = \alpha_0 + \alpha_1 f_{F1,t} + \varepsilon_{1t} \tag{C5}$$

$$y_t = \alpha_0 + \alpha_1 f_{F2,t} + \varepsilon_{1t} \tag{C6}$$

$$y_t = \alpha_0 + \alpha_1 f_{F3,t} + \varepsilon_{1t} \ , \tag{C7}$$

where $f_{F1,t}, f_{F2,t}, f_{F3,t}$ correspond to forecasts constructed using models (F1)-(F3).

(M-AIC): Select the "best" group of forecasts to use in any later forecast combination regressions as above, but use the AIC rather than the SIC.

(M-ALL): Use all three forecasts as the "best" group. Note that here we simply construct least squares forecast combination weights using all competing forecasts, without any concern for collinearity, significance, etc.

(M-TST): Select the optimal group of forecasts to use in any later forecast combination using the technique employed by Fair and Shiller (1990). In particular, a forecast combination regression is first run which contains all of the competing forecasts. The "best" group of forecasts from among these (which, throughout our experiments, is later used in a secondary forecast combination regression to construct real-time combined forecasts from which out-of-sample MSE, MAD, and MAPE criteria are calculated) is selected by using heteroskedasticity and serial correlation robust standard errors (see Fair and Shiller (1990)) and t-statistics. The size of the t-test of the null hypothesis that the coefficient in question is zero is set at 10%. If all of the competing forecasts fail the t-test, then we use all three forecasts as the "best" group. Note that in these cases, we have direct evidence that the t-test approach suffers from severe collinearity problems. Interestingly, we will see that the collinearity issue is very important in both our Monte Carlo experiments as well as in our empirical analysis.

(M-MSE): Select the optimal group of forecasts by running (C1)-(C7) and constructing MSE/(100-k) based on the in-sample period of 100 observations, and for k equal to the number of estimated parameters. This statistic is an in-sample version of the mean squared error statistic used to measure the out-of-sample

performance of the models selected as "best". Thus, the two versions of MSE criteria used are quite different, and should not be confused with one another.

To keep the simulation run-time manageable, we do not construct experiments which use the *ex ante* model selection approach discussed in Section 2.2. However, the approach has been used in Swanson, Ozyildirim and Pisu (1996), who find that *ex ante* forecasts offer a convenient short-cut to standard tests of the null hypothesis of Granger noncausality. Swanson et al. also find that the AIC and the SIC perform very well when used to test the null of Granger noncausality.

All estimations are carried out using least squares as the optimization criterion. Although many generalizations of the forecast combination techniques discussed in the previous paragraphs are available, we simplify our analysis by considering only the six techniques mentioned. However, as discussed above, we feel that these techniques are sufficiently diverse to give us a good indication as to the relative performance of (M-SIC), (M-AIC), and (M-MSE), for example, for "pre-selecting" a "best" group of forecasts to include in subsequent forecast combination regressions, and for conducting model selection based on encompassing type regressions.

3.2 Simulation Results

Tables 1-4 contain summaries of the 192 Monte Carlo experiments. (Tables which report complete simulation results are available upon request from the authors.) In practice, we do not know *a priori* whether the data which we are examining in any particular forecasting experiment are integrated or not. To account for this feature, we examine combination regressions based on levels data (FC1-L)-(FC3-L) as well as based on differenced data (FC1-D)-(FC3-D). Also, we use the same three basic combination regression types considered by Granger and Ramanathan (1984), even though the combined forecasts constructed using (FC3-D) may be sub-optimal relative to those based on (FC1-D), for example. In order to keep the length of our tables manageable, the results in Tables 1-4 are based on the combination regression which yielded the best possible ex ante MSE performance (say, M_OPT), were the forecaster to have perfect foresight in each simulation. Of course, this does not bias our presented findings, as any different combination regression and any given combination technique may be preferred for any given simulation in each of the 192 experiments, but rather serves as a convenient way to summarize a large set of

experiments. (A complete appendix with all findings, as well as M-OPT values is available upon request from the authors.)

Our experimental findings can be broadly summarized as follows. First, the Schwarz Information Criterion (M-SIC) approach to selecting the "best" group of forecasts from amongst competing models clearly outperforms all others considered here based on MSE and MAD, while the simple averaging rule is "best" based on MAPE. In particular, 4 different forecast models based on (F1)-(F7) were used for each of our 4 DGPs, and simulation results for each DGP are reported on in each of tables 1 to 4. Combining the results of all 4 tables, the following findings emerge: (i) for MSE, M-SIC wins 10 times, M-AVG 5 times, and M-ALL 1 time, (ii) for MAD, M-SIC wins 11 times and M-AVG 5 times, and (iii) for MAPE, M-SIC wins 4 times and M-AVG 12 times. This finding is interesting for two reasons. First, M-AVG does not dominate other combination techniques, as one might believe it would, given recent evidence reported in the literature (see above), and second, the choice of loss function is clearly important when choosing combination technique. Indeed, other loss functions such as profit maximization might be expected to select yet different combination techniques. However, the clear finding that M-SIC dominates based on the oft used MSE measure should perhaps be stressed.

Second, the other two model selection approaches, including the less parsimonious M-AIC, and the M-MSE technique, do not appear to offer a valid alternative to standard hypothesis testing approaches for selecting amongst competing forecasting models. This finding is evidenced by noting that based on all 16 model-DGP combinations, M-TST wins 11 times, when only M-TST, M-AIC, and M-MSE are compared. (Although this finding is not altered when M-ALL is also included in the comparison.) Put another way, the M-TST approach appears to clearly outperform all other methods considered, except of course, M-SIC and M-AVG.

Third, only about one half of the point estimates reported in Tables 1-4 seem to be very close together. This is perhaps best illustrated graphically. In Figure 1, the first panel plots MSE values for all 16 model-DGP combinations (in the order in which they are presented in Tables 1-4), for M-SIC, M-AIC, M-TST, and M-MSE. M-ALL is not included because it does consistently more poorly than the other techniques, while The second panel plots M_AVG against M-SIC. Note that M-AVG does very poorly for

point 12, which corresponds to the use of F5, F6, and F6 and DGP3. Interestingly, from the first panel, it is clear that all techniques do quite poorly in this case, with M-SIC being clearly dominant, however. All techniques except for M-SIC also do badly for all of DGP3 and DGP4 (represented by points 9 to 16). Interestingly, these are the nonstationary DGPs. Based on the stationary DGPs, MSEs for M-AVG and M-SIC are generally close together, although M-SIC is somewhat lower for points 3 to 6 in the plot. Again, it is worth noting that the other combination techniques also perform poorly relative to M-SIC for points 3 to 6 (which overlap models used in DGP1 and DGP2). Thus, one of the features of M-SIC that is apparent upon inspection of Figure 1 is that M-SIC is a stable technique that is not prone to the outliers or "blow-ups" which appear to plague all of the other combination techniques.

Fourth, although the detailed results are not reported here, we note that: (1) In many cases the "best" model, in the sense of out-of-sample performance, is not the true DGP. In particular, for (DGP1) and (DGP3), where the true DGP is nested within the alternative forecasting models for many of the experiments, the specification chosen as "best" using the model selection techniques is not the true specification (e.g. an AR(2) is often preferred to an AR(1), even if the true DGP is an AR(1)). As mentioned above, this rather surprising result suggests that it is not enough to simply determine which technique selects the true specification most frequently, particularly if the preferred forecasts are ones which minimize standard out-of-sample forecast performance measures (here we use MSE, MAD, and MAPE). This, in fact, is the main reason why our experiments include the *ex ante* performance measures, rather than simple counts of the number of times that each technique "picks" the true model. One possible reason for this type of finding is that we are considering only forecasts of the first moment, and are considering only h=1. Perhaps an AR(1) specification would be preferred if models of the entire distribution were estimated, for example. (2) The nonlinear specifications (F6) and (F7) occasionally appear in the forecast combination chosen as "best" by all of our forecast combination techniques. One possible conclusion based on these two findings is that while parsimony appears to be a good rule when selecting *combinations* of forecasts (i.e. (M-SIC) picks a better "group" of forecasts than (M-AIC)), individual forecast construction should be carried out using liberal specifications (e.g. use the AIC to select the optimal number of lags, rather than the SIC), if the goal is to construct "optimal" forecasts based on forecast combination regressions. This strategy differs

from the standard approach taken when a *single* forecasting model is constructed, and weighting regressions are *not* used. In particular, in the latter case it is well known that more parsimonious models often perform "best", while we suggest that when constructing "weighted" forecasts, less parsimonious models should also be estimated.

Even though forecast combination regression types (FC2) and (FC3) are not optimal in the sense described by Granger and Ramanathan (1984), it is worth noting that our findings are not affected by which type of forecast combination regression is run (although we only report results for M-OPT cases in Tables 1-4). This is one reason why we feel that our results serve as a good indicator of the usefulness of (M-SIC) in the context of more complex regression based forecast combination techniques. Overall, the model selection approach to forecast combination appears to offer a tractable an appealing alternative to various other standard approaches, including (M-AVG) and (M-TST), and should be considered seriously when constructing forecasts. However, further research is needed before a final choice between (M-AVG) and (M-SIC), say, is made, and indeed it may turn out that both techniques turn out to be preferred, with a choice between the two occurring on an individual case by case basis. In the next section we use econometric and professional forecasts to further compare a number of the techniques discussed above.

4. Empirical Evidence

In a recent paper, Swanson and White (SW: 1996) examine the relative performance of individual *ex ante* econometric and professional forecasts using a model selection approach. They compare a "best" linear model which has parameters re-estimated before each new real-time forecast is constructed, an adaptive linear model, which respecifies and re-estimates the linear model at each point in time, and an adaptive artificial neural network model, which includes nonlinear regressors. Although we are here interested in forecast combination rather than individual performance, and in the selection of "best" groups of forecasts using the AIC, SIC, and other techniques, the set-up of SW's analysis is a reasonable starting point for us. We use the same dataset as SW, and also use the same econometric models to produce our sequences of econometric forecasts. The next section outlines the setup of the empirical part of our paper, and the *Data* and *Model* sections are largely excerpted from SW, where more details can be found.

4.1 Real Time Forecasting Models and Data

4.1.1 Data

For the period 1960:1 to 1993:3, unrevised quarterly U.S. data on unemployment, interest rates, industrial production, nominal gross national product, corporate profits, real gross national product, personal consumption expenditures, the change in business inventories, and net exports of goods and services are considered. Table 5 expands on the series definitions. It should be noted that most of the variables are seasonally adjusted. Although the implications of this are not obvious, we nevertheless have ensured that future information due to the application of two-sided seasonal adjustment filters, for example, has not seeped into our data set. In order to collect the data, each monthly issue of the Survey of Current Business from 1960 to 1993 was examined. Each time a "new", or first available, observation for any of the series was reported, we added one more observation to our data set. In addition to forecasts constructed using econometric models, we examine *median* professional forecasts. The professional forecasts were provided by Dean Croushore, and are collected in the Survey of Professional Forecasters (SPF) data set (Croushore (1993)). It should be pointed out that our use of only median forecasts rather than the entire panel is a simplification that may lead to testing bias in certain cases. For instance, Keane and Runkle (1990) avoid aggregation bias by using the full panel of the SPF when testing the rationality of price forecasts, and find results different than when only mean forecasts are used.

Any comparison of forecast accuracy depends on the timing of the forecasts. Thus, comparisons of our econometric and "professional" forecasts depend crucially on the timing and availability of the data used to construct the competing forecasts. The real-time feature of our forecasts makes it relatively easy to pinpoint the timing used in our analysis. Generally, SPF surveys are mailed around the beginning of the current quarter, say period t. Responses are requested by shortly after the *middle* of the current quarter and consist of forecasts for periods t, $t+1$, ..., $t+5$. However, since respondents are asked to forecast current quarter values *during* the current quarter, their information sets may contain a large amount of the *same* information that is later used by the government to construct the actual data for period t. Our approach to this issue is to compare SPF forecasts for $t+1$ that are made during period t (which we call "next period" forecasts), as well as forecasts for t that are made during period t (which we call "current period" fore-

casts). All econometric forecasts are constructed using data *through* period t to forecast $t+1$. This approach differs from SW, where only "next period" professional forecasts are compared, and allows us to essentially "straddle" both sides of any potential timing incompatibility problem, when comparing the alternative forecasts. Of final note is that we use an "unrevised" dataset. This is not truly "real-time", as a real-time dataset should consist of an entire new dataset at each point in time, corresponding to *all* of the information known at that point in time, both revised, partially revised, and unrevised (see Swanson (1996) for a discussion of these and related issues).

4.1.2 Models

The linear models specified in this paper are all special cases of the following model:

$$y_{t+h-1} = \alpha_0 + \sum_{i=1}^{K1} \beta_i \, y_{t-i} + \sum_{i=1}^{K2} \gamma_i \, x_{t-i} + \sum_{i=1}^{K3} \delta_i \, z_{t-i} + u_{t+h-1}, \tag{4.1}$$

where y_t is one of the nine macroeconomic variables, and h is the horizon of our forecast, in quarters. The independent variables, x_t and z_t, are two other variables chosen from our set of nine macroeconomic variables (see the discussion in SW).

In all, 21 versions of (1) are estimated. The first model corresponds to a random walk, where $\alpha_0 = 0$, $\beta_1 = 1$, β_i, $i = 2,..., K1 = 0$, and $\gamma_i = \delta_i = 0$, $i = 1,...,5$. The next five models are AR(K1) processes, where K1=1,...,5 and $\gamma_i = \delta_i = 0$, for all i. Two-variable VAR(5) models are also considered, where alternately: (i) K1=K2=1,...,5 and $\delta_i = 0$, $i = 1,..., K3$, and (ii) K1=K3=1,...,5 and $\gamma_i = 0$, $i = 1,..., K2$. The final five models considered are three-variable VAR(5) specifications, where K1=K2=K3=1,...,5.

In this study, we consider these models as special cases of a fairly broad array of forecasting models, while realizing that various other linear models that we don't examine here are also available. The random walk model is called a "no change" model, while all other models are referred to by the ordered triplet (K1,K2,K3), and are called nonadaptive linear VAR models.

We also examine a class of flexible models called "artificial neural network" (ANN) models. Because of their flexibility and simplicity, and because of demonstrated successes in a variety of empirical applications (see White (1989) and Kuan and White (1994) for some specifics), ANNs have become the focus of considerable attention as a possible vehicle for forecasting economic variables, and in particular, financial

variables. Here, we treat these models as a potentially interesting black box, delivering a specific class of nonlinear regression models. In particular, the ANN regression models considered here have the form:

$$f(\tilde{w}, \theta) = \tilde{w}'\kappa + \sum_{j=1}^{q} G(\tilde{w}'\pi_j)\lambda_j \qquad (4.2)$$

where $\tilde{w} = (1, w')'$ is a (r+1)x1 vector of explanatory variables,

$$w = (y_{t-1}, \dots, y_{t-K1}, x_{t-1}, \dots, x_{t-K2}, z_{t-1}, \dots, z_{t-K3})',$$

$\theta = (\kappa', \lambda', \pi')'$, $\lambda = (\lambda_1, \dots, \lambda_q)'$, $\pi = (\pi_1', \dots, \pi_q')'$, q is a given integer, and G is a given nonlinear function, in our case, the logistic cumulative distribution function (c.d.f.) $G(z) = 1/(1 + \exp(-z))$.

A network interpretation of (4.2) which is also given in SW is as follows. "Input units" send signals $\tilde{w} = (w_0(=1), w_1, \dots, w_r)$ over "connections" that amplify or attenuate the signals by a factor ("weight") π_{ji}, $i = 0, \dots, r$, $j = 1, \dots, q$. The signals arriving at "intermediate" or "hidden" units are first summed (resulting in $\tilde{w}'\pi_j$) and then converted to a "hidden unit activation" $G(\tilde{w}'\pi_j)$ by the operation of the "hidden unit activation function", G. The next layer operates similarly, with hidden activations sent over connections to the "output unit." As before, signals are attenuated or amplified by weights λ_j and summed. In addition, signals are sent directly from input to output over connections with weights κ. A nonlinear activation transformation at the output is also possible, but we avoid it here for simplicity. In network terminology, $f(w, \theta)$ is the "network output activation" of a "hidden layer feedforward network" with "inputs" w and "network weights" θ. The parameters π_j are called "input to hidden unit weights," while the parameters λ_j are called "hidden to output unit weights." The parameters κ are called "input to output unit weights."

Our approach is to apply model (4.2) to the problem of forecasting y_{t+h-1} using explanatory variables w corresponding to the variables considered in the linear forecasting models described above, and with the highest number of "hidden units", $q = 5$ (note that q is chosen by the network, can take any value for 0 to 5, and can vary as we roll through the forecast horizon). When $\lambda_i = 0$, $i = 1,...,5$, (so that $q = 0$ for all t) we have a three variable adaptive linear VAR(5) model as a special case (this model is designated as our adaptive linear VAR model).

The parameters of our linear models are estimated using the method of least squares. The nonlinear

models are estimated using a combination of least squares and the method of steepest descent (for select-

ing "hidden unit" parameters).

4.1.3 Forecast Combination and Model Selection

In SW, sequences of 45 1- and 4-step ahead econometric forecasts corresponding to the period

1982:3-1993:3 are constructed using rolling windows of 40, 68, and 76 quarters of data. Here we use the

"best" from among those forecasts, based on SW's findings. In particular, we examine three sequences of

forecasts for each of the nine macroeconomic variables. The forecasts are based on: (i) the "best" linear

nonadaptive model (chosen from (4.1) using a training set of data); (ii) the ANN model (which is adaptive,

and is often linear, as there are many cases where no hidden units are found), and (iii) the Survey of Pro-

fessional Forecasters. In all cases, forecasts are compared to fully revised CITIBASE data which were

downloaded in the fourth quarter of 1994.

The forecast sequences are used in two different ways. First, a "training set" of the forecasts (ranging

from $O = 15$ to 30 observations) is used to form an optimal forecast combination regression based on each

of the techniques (M-SIC), (M-AIC), (M-ALL), (M-TST), and (M-MSE). [1] Separate "best" combination

regression equations are then specified for h=1, and for each type of regression given by (FC1-L)-(FC3-L)

and (FC1-D)-(FC3-D). The "best" combination weights are then used to produce a single 1-step ahead

forecast. The "weights" from the "best" regression are fixed throughout the out-of-sample period. (It

should be of interest, though, to allow the weights to be re-estimated at each point in time, as the sample

"rolls" forward.) This process is continued by rolling forward one-step at a time, until the end of the S=45

period forecast sample, at which stage we have sequences of S-O real-time forecasts based on the various

forecast combination strategies. These new forecasts are then compared with the actual data values for the

ex ante period, using the out-of-sample MSE, MAD, and MAPE criteria discussed above. This experiment

enables us to conduct a real-time analysis of the different forecast combination techniques using actual

data, and should provide further evidence as to the relative usefulness of alternative forecast combination

techniques.

_____

[1] It should be noted that in this section we focus our attention on the relative merits of all of the combination
techniques discussed above with the exception of (M-AVG). Thus, we focus our attention on the various combination
techniques which involve regression. This should not be taken as implying that (M-AVG) is not useful, however, as we
have seen above that only (M-SIC) generally outperforms (M-AVG) when selecting forecast combinations.

Second, the "training set" is fixed at 45 observations, so that the entire sequence of 1-step ahead forecasts are used. Then, the various forecast combination techniques (e.g. (M-AIC) and (M-TST)) are used to select "best" groups of forecasts, for all regression types ((FC1-L)-(FC3-L) and (FC1-D)-(FC3-D)). This allows us to directly examine the relative performance of our three different forecasting models, but does not involve any *ex ante* evaluation of the different forecast combination techniques. As discussed above, if the (M-SIC) technique picks a "best" group of forecasts which includes the nonadaptive linear econometric and the SPF forecasts, say, then we have direct evidence that these two types of forecasts dominate the ANN forecasts. In our analysis, we classify the methods according to either (i) the *nonadaptive* linear model wins, (ii) the *adaptive* linear or nonlinear ANN model wins, (iii) the SPF model wins, (iv) both econometric models win, or (v) some other combination of forecasts which includes the SPF wins. The analysis is carried out for all forecast combination techniques, combination regression types, forecast horizons, and variables.

4.2 Empirical Results

4.2.1 Comparison of Five Alternative Forecast Combination Methods

As discussed above, sequences of 45-O real-time forecasts for the nine variables are constructed by using an "in-sample" period of O observations (O ranges from 15 to 30) to first select a "best" group of forecasts using each forecast combination technique, and then by using the "best" specification to construct 1-step ahead real-time rolling forecasts for the rest of the 45 observation sample. The real-time forecasts are then used to construct out-of-sample MSE, MAD, and MAPE criteria. In Tables 6 and 7 we report summary results for O=25. Complete results, as well as results for O=15 - O=30 are available from the authors, but are not included here. Overall, the results for O=25 are very similar to the results obtained when other "training sets" of forecasts are used. Furthermore, our empirical results do not change in any noteworthy manner when SPF forecasts for time $t+1$ (see Table 7) are used (see above discussion) instead of SPF forecasts for time t (see Table 6).

Since Table 6 tabulates findings for the case where SPF forecasts for t are constructed using data available midway through quarter t, the SPF forecasts use more information than is available to our econometric models. For combination regression types (FC1-L) and (FC1-D), the (M-SIC) picks the

forecast combinations which are MSE-, MAD-, and MAPE-dominant for the "most" variables in 6 of 6 cases. For example, using this measure, (M-SIC) gets 1 of its 6 "wins" in the {MSE, FC1-L} cell of Table 4, as (M-SIC) selects the optimal "group" of forecasts for 7 of 9 macroeconomic variables. Of note is that all of the numbers in this cell sum to 19, which is clearly much greater than the number of variables (nine). This is because when two methods both pick the best group of forecasts for a variable, they are both afforded a "win". Across all different combination regression types (e.g. FC1-L, FC1_D, etc.), (M-SIC) "wins" 10 of 17 times, with 5 of these coming from "ties". Not surprisingly, (M-TST) appears to be the second best technique, although (M-SIC) clearly dominates. Thus, as found above based on Monte Carlo experiments, SIC appears to offer a potentially interesting alternative method for selecting the "best" group of forecasts.

4.2.2 Comparison of Professional and Econometric Forecasts

In Section 4.2.1 we only examined the relative performance of the alternative forecast combination techniques. In Tables 8-9 the relative performance of the alternative forecasting models (SPF, nonadaptive linear, and adaptive linear and nonlinear) is compared. This is done by setting the "training set" of forecasts used equal to the entire sequence of 45 forecasts. Then, the only result of our analysis will be the "selection" of optimal "groups" of forecasts for each macroeconomic variable, based on each alternative forecast combination technique ((M-ALL), (M-AIC), (M-SIC), (M-TST), and (M-MSE)).

Table 8 reports results for "own period" SPF forecasts, with "next period" SFP forecast results gathered in Table 9. The results from the two tables are similar, and thus we focus our discussion on Table 8. The entries in the table can be interpreted as follows. Consider forecast combination regression (FC1-D). The number of "wins" associated specifically with the (M-SIC) forecast combination technique sum to nine (corresponding to the nine macroeconomic variables). Indeed, we see that the "LINEAR" econometric model is individually preferred for 2 variables, the "ADAPTIVE" econometric model is individually preferred for 4 variables, "BOTH" econometric models are combined optimally for 0 variables, the "SPF" model is individually preferred for 3 variables, and any "OTHER" combinations of forecasts are preferred for 0 variables. These figures provide direct evidence that based on h=1, (M-SIC), and (FC1-D), the optimal combined weight forecasts for our variables are in all cases constructed using individual forecasts,

with the SPF model and the econometric models each winning approximately the same number of times. Interestingly, if we again consider (FC1-D), but use (M-TST) as our model selection technique, *combinations* of SPF and econometric forecasts are selected for 8 of 9 variables. One reason for this finding is that (M-TST) is usually faced with a high degree of collinearity, resulting in low t-statistics for *all* weights, and hence in "best" groups of forecasts which contain all three forecasting models. This is precisely the type of problem that may frequently arise in practice, and highlights the potential usefulness of (M-SIC) for choosing optimal "groups" of forecasts. It is perhaps surprising, though, that collinearity is so prevalent, given that only three alternative forecasts are used. Although the detailed results are not tabulated, it is worth noting that the forecasts selected as best by (M-SIC) and based on (FC1-D) in Table 8 (*h*=1) are: (1) - nonadaptive "LINEAR" models are individually preferred for interest rates (R) and nominal GNP (NGNP) (2) - "ADAPTIVE" econometric models are individually preferred for unemployment (U), industrial production (IP), corporate profits (Π), and personal consumption expenditures (PCE) (3) - "SPF" forecasts are individually preferred for real GNP (RGNP), change in business inventories (ΔBI), and net exports of goods and services (Net X).

5. Conclusions

In this paper we have taken a model selection approach to selecting "optimal" groups of forecasts for inclusion in forecast combination regressions. The approach avoids issues associated with collinearity among competing forecasts, and is of particular interest when groups of more than two variables are being compared and/or combined. In closing, we offer the following conclusions. First, an approach to forecast combination based on the Schwarz Information Criterion appear to dominate a number of linear alternatives, including the t-statistic approach adopted by Fair and Shiller (1990), a simple approach which uses all forecast combination weights without concern for their significance, an arithmetic average approach, and various other model selection techniques, when mean square forecast error (MSE) and mean absolute forecast error deviation (MAD) loss functions are specified. The arithmetic average approach is second best, and is actually preferred to the model selection approach when the loss function specified is mean absolute percentage forecast error (MAPE). All other techniques perform rather poorly relative to the two best techniques, regardless of loss function. Second, collinearity among competing forecasts is an

important factor, even when groups of only 3 forecasts are considered, partially accounting for the inability of the t-statistic approach to adequately select models which are "best" based on *ex ante* MSE, MAD, and MAPE loss functions. Third, in an empirical experiment which uses nine macroeconomic variables, we find that individual forecasts are often selected as "best" (i.e. forecast combination techniques favor including only one forecast with nonzero weight), even when multiple professional and econometric forecasts are used to form optimal forecast combinations. For example, in the context of the linear forecast combination regressions considered here which are based on differenced data, adaptive econometric models (which include artificial neural networks as well as linear models which are re-specified and re-estimated at each point in time) are individually selected as "best" approximately the same percentage of the time as are professional forecasts from The Survey of Professional Forecasters. Naïve linear as well as combined weight models are selected as "best" much less frequently.

The work reported here is meant as a starting point. Given the apparent success of model selection in the linear forecast combination context, a variety of further questions present themselves for subsequent research. For instance, are the results reported here robust to the use of more complex combination regression specifications? Does the use of time varying parameter models, or nonlinear combination equations have an effect on our results? Also, why does the SIC technique so clearly dominate model selection strategies based on the AIC and on the in-sample MSE?

References

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in B. N. Petrov and F. Csaki, eds., 2nd International Symposium on Information Theory Budapest: Akademiai Kiado, 267-281.

Akaike, H., 1974, A new look at the statistical model identification, IEEE Transactions on Automatic Control AC-19, 716-23.

Bates, J.M. and Granger, C.W.J., 1969, The combination of forecasts, Operations Research Quarterly 20, 451-468.

Clemen, R.T., 1985, Extraneous expert information, Journal of Forecasting 4, 329-348.

Clemen, R.T., 1989, Combining forecasts: a review and annotated bibliography, International Journal of Forecasting 5, 559-583.

Coulson, N.E. and Robins, R.P., 1993, Forecast combination in a dynamic setting, Journal of Forecasting 12, 63-67.

Croushore, D., 1993, Introducing: the survey of professional forecasters, Business Review, The Federal Reserve Bank of Philadelphia, November-December, 3-15.

Deutsch, M., Granger, C.W.J., and Teräsvirta, T., 1994, The combination of forecasts using changing weights, International Journal of Forecasting 10, 47-57.

Diebold, F.X. and Lopez, J.A., 1995, Forecast evaluation and combination, Handbook of Statistics, Amsterdam: North-Holland, forthcoming.

Diebold, F.X. and Mariano, R.S., 1995, Comparing predictive accuracy, Journal of Business and Economic Statistics 13, 253-263.

Diebold, F.X. and Pauly, P., 1987, Structural change and the combination of forecasts, Journal of Forecasting 6, 21-40.

Engle, R.F., Granger C.W.J., and Kraft, D.F., 1984, Combining competing forecasts of inflation using a bivariate ARCH model, Journal of Economic Dynamics and Control 8, 151-165.

Engle, Robert F. and Brown, S.J. 1986, Model selection for forecasting, Applied Mathematics and Computation 20, 313-327.

Engle, R.F. and Granger, C.W.J., 1987, Co-integration and error correction: representation, estimation, and testing, Econometrica 55, 251-76.

Fair, R.C. and Shiller, R.J., 1990, Comparing information in forecasts from econometric models, American Economic Review 80, 375-389.

Fomby, T.B. and Guilkey, D.K., 1978, On choosing the optimal level of significance for the Durbin-Watson test and the Bayesian alternative, Journal of Econometrics 8, 203-213.

Granger, C.W.J., 1969, Investigating causal relation by econometric models and cross spectral methods, Econometrica 37, 428-438.

Granger, C.W.J., 1989, Combining forecasts: twenty years later, Journal of Forecasting 8, 167-173.

Granger, C.W.J., King, M.L. and White, H., 1995, Comments on testing economic theories and the use of model selection criteria, Journal of Econometrics 67, 173-187.

Granger, C.W.J. and Newbold, P., 1986, Forecasting economic time series, San Diego: Academic Press.

Granger C.W.J. and Ramanathan, R., 1984, Improved methods of combining forecasts, Journal of Forecasting 3, 197-204.

Hallman, J. and Kamstra, M., 1989, Combining algorithms based on robust estimation techniques and co-integrating restrictions, Journal of Forecasting 8, 189-198.

Keane, Michael P. and Runkle, D.E., 1990, Testing the rationality of price forecasts, American Economic Review 80, 714-735.

Kuan, C.-M. and White, H., 1994, Artificial neural networks: an econometric perspective, Econometric Reviews 13, 1-91.

Mizrach, B., 1991, Forecast comparison in L_2, mimeo, Department of Finance, Wharton School, University of Pennsylvania.

Rissanen, J., 1978, Modeling by shortest data description, Automatica 14, 465-471.

Russell, T.D. and Adam, E.A., 1987, An empirical evaluation of alternative forecasting combinations, Management Science 33, 1267-1276.

Schmittlein, D.C., Kim, J., and Morrison, D.G., 1990, Combining forecasts: operational adjustments to theoretically optimal rules, Management Science 39, 1044-1056.

Schwarz, G. (1978), Estimating the dimension of a model, The Annals of Statistics 6, 461-4.

Sin, C.-Y. and White, H., 1995, Information criteria for selecting possibly misspecified parametric models, Journal of Econometrics 71, 207-225.

Swanson, Norman R., 1996, Forecasting using first available versus fully revised economic time series data, Studies in Nonlinear Dynamics and Econometrics 1, 47-64.

Swanson, N.R., Ozyildirim, A., and Pisu, M., 1996, A comparison of alternative causality and predictive accuracy tests in the presence of integrated and co-integrated economic variables, mimeo, Pennsylvania State University.

Swanson, N.R. and White, H., 1995, A model selection approach to assessing the information in the term structure using linear models and artificial neural networks, Journal of Business and Economic Statistics 13, 265-275.

Swanson, N.R. and White, H., 1996, A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, Review of Economics and Statistics, forthcoming.

Swanson, N.R. and White, H., 1997, Forecasting economic time series using adaptive versus nonadaptive and linear versus nonlinear econometric models, International Journal of Forecasting, forthcoming.

White, H., 1989, Learning in artificial neural networks: a statistical perspective, Neural Computation 1, 425-464.

Table 1: Forecast Combination Method Performance Based on DGP1

Combination Method	Forecast Horizon $h=1$		
	MSE	MAD	MAPE
Forecast Models Used: F1, F2, F5			
M-ALL	1.0776	0.8263	10.0492
M-AVG	1.0460	0.8164	4.5512
M-AIC	1.0611	0.8201	9.7587
M-SIC	1.0428	0.8131	9.7512
M-TST	1.0599	0.8197	9.4044
M-MSE	1.0705	0.8236	9.8089
Forecast Models Used: F1, F2, F6			
M-ALL	1.0753	0.8257	6.1093
M-AVG	1.0567	0.8201	2.9076
M-AIC	1.0531	0.8175	5.9041
M-SIC	1.0390	0.8122	5.8746
M-TST	1.0459	0.8147	5.4685
M-MSE	1.0646	0.8216	5.9430
Forecast Models Used: F1, F5, F7			
M-ALL	1.0889	0.8283	0.1198
M-AVG	1.0129	0.8032	3.6376
M-AIC	1.0742	0.8230	0.1152
M-SIC	1.0587	0.8181	0.1147
M-TST	1.0800	0.8263	0.1183
M-MSE	1.0824	0.8258	0.1184
Forecast Models Used: F1, F6, F7			
M-ALL	1.7690	0.8833	3.2899
M-AVG	1.1642	0.8602	2.4550
M-AIC	1.2766	0.8472	2.1319
M-SIC	1.0960	0.8243	1.8327
M-TST	1.6677	0.8731	3.1665
M-MSE	1.5502	0.8666	2.5396

[1] **Results based on the 6 different forecast combination techniques (M-ALL, M-AVG, ...) are tabulated. The different combination techniques are discussed above. All reported results are values for three loss functions, mean square forecast error (MSE), mean square absolute forecast deviation (MAD), and mean absolute forecast percentage error (MAPE), and are based on 5000 simulations. All entries are constructed by forming sequences of 100 1-step ahead forecasts of data generated according to DGP1. The sequences of forecasts are based on forecast combinations which use forecast models given by models F1-F7. This ex ante comparison fixes the "best" group of regressors chosen by each technique, but re-estimates the weights as rolling ex ante 1-step ahead forecasts are constructed.**

Table 2: Forecast Combination Method Performance Based on DGP2

Combination Method	Forecast Horizon $h = 1$		
	MSE	MAD	MAPE
Forecast Models Used: F1, F2, F6			
M-ALL	1.0751	0.8256	6.7047
M-AVG	1.0247	0.8078	3.5055
M-AIC	1.0553	0.8183	6.6045
M-SIC	1.0424	0.8137	6.6175
M-TST	1.0489	0.8160	6.9411
M-MSE	1.0648	0.8217	6.6746
Forecast Models Used: F2, F4, F5			
M-ALL	1.0751	0.8256	3.3695
M-AVG	1.0823	0.8305	5.2901
M-AIC	1.0627	0.8210	3.2525
M-SIC	1.0486	0.8158	3.0795
M-TST	1.0579	0.8195	3.1562
M-MSE	1.0703	0.8238	3.2535
Forecast Models Used: F2, F4, F7			
M-ALL	1.0568	0.8200	7.1865
M-AVG	1.0252	0.8080	3.4641
M-AIC	1.0524	0.8183	7.2157
M-SIC	1.0513	0.8179	7.2104
M-TST	1.0515	0.8180	7.0539
M-MSE	1.0538	0.8189	7.2575
Forecast Models Used: F4, F6, F7			
M-ALL	1.0794	0.8273	6.1360
M-AVG	1.0408	0.8140	3.3751
M-AIC	1.0545	0.8185	5.7305
M-SIC	1.0418	0.8142	5.7874
M-TST	1.0550	0.8189	6.3809
M-MSE	1.0652	0.8222	5.8669

[1] **See notes to Table 1.**

Table 3: Forecast Combination Method Performance Based on DGP3

Combination Method	Forecast Horizon $h = 1$		
	MSE	MAD	MAPE
Forecast Models Used: F1, F2, F5			
M-ALL	1.0463	0.8150	2.4451
M-AVG	1.0189	0.8055	0.0979
M-AIC	1.0428	0.8136	2.4098
M-SIC	1.0409	0.8129	2.3966
M-TST	1.0486	0.8158	2.3289
M-MSE	1.0447	0.8143	2.4318
Forecast Models Used: F1, F5, F6			
M-ALL	1.0868	0.8271	0.1219
M-AVG	1.4126	0.9137	0.1238
M-AIC	1.0701	0.8214	0.1165
M-SIC	1.0578	0.8171	0.1163
M-TST	1.0788	0.8258	0.1241
M-MSE	1.0795	0.8243	0.1170
Forecast Models Used: F2, F5, F7			
M-ALL	1.1299	0.8437	2.7961
M-AVG	1.6663	0.9926	0.1228
M-AIC	1.0938	0.8312	2.3924
M-SIC	1.0694	0.8225	2.1664
M-TST	1.1033	0.8351	2.6714
M-MSE	1.1126	0.8375	2.5744
Forecast Models Used: F5, F6, F7			
M-ALL	1.3981	0.8757	3.4308
M-AVG	3.0900	1.2322	0.1786
M-AIC	1.2647	0.8502	2.7117
M-SIC	1.1884	0.8288	2.2337
M-TST	1.3341	0.8643	3.2854
M-MSE	1.3090	0.8620	2.9739

[1] **See notes to Table 1.**

Table 4: Forecast Combination Method Performance Based on DGP4

Combination Method	Forecast Horizon $h = 1$		
	MSE	MAD	MAPE
Forecast Models Used: F1, F2, F5			
M-ALL	1.0637	0.8219	0.0594
M-AVG	1.0649	0.8238	0.0487
M-AIC	1.0667	0.8230	0.0593
M-SIC	1.0656	0.8226	0.0591
M-TST	1.0709	0.8246	0.0597
M-MSE	1.0655	0.8226	0.0592
Forecast Models Used: F1, F2, F6			
M-ALL	1.0924	0.8321	5.4117
M-AVG	1.1953	0.8685	0.0516
M-AIC	1.0857	0.8296	5.2353
M-SIC	1.0843	0.8291	5.2297
M-TST	1.1534	0.8551	2.9785
M-MSE	1.0878	0.8304	5.2188
Forecast Models Used: F2, F5, F7			
M-ALL	1.2805	0.8969	6.0666
M-AVG	1.2621	0.8918	0.0509
M-AIC	1.2543	0.8871	5.1414
M-SIC	1.2123	0.8726	4.8071
M-TST	1.2516	0.8866	5.6268
M-MSE	1.2673	0.8921	5.3276
Forecast Models Used: F4, F6, F7			
M-ALL	1.5107	0.9150	4.9194
M-AVG	1.7265	1.0194	0.0650
M-AIC	1.4033	0.8942	4.8264
M-SIC	1.2268	0.8724	4.5207
M-TST	1.4279	0.8963	4.6306
M-MSE	1.4827	0.9068	5.0583

[1] **See notes to Table 1.**

Table 5: Variable Definitions and Mnemonics [1]

Variable	Description
NGNP	Gross National Product: SA, $billions, nominal, Quarterly.
RGNP	Gross National Product: SA, $billions, real, various base years, Quarterly.
PCE	Personal Consumption Expenditures: SA, $billions, real, various base years, Quarterly.
Π	Corporate Profits After Taxes: SA, $billions, nominal, Quarterly.
IP	Industrial Production Index: SA, index, various base years, Averaged from monthly.
R	Aaa Corporate Bond Yield: Moody's, %, Averaged from monthly.
ΔBI	Change in Business Inventories: SA, $billions 1987, Quarterly.
Net X	Net Exports of Goods and Services: SA, $billions 1987, Quarterly.
U	Civilian Unemployment Rate: SA, %, Averaged from monthly.

[1] **All data are collected from various issues of the Survey of Current Business. SA stands for seasonally adjusted. The full sample is 1960:1-1993:3. All ex-post model selection uses the sample 1982:3-1993:3. Linear and adaptive network models are compared to median forecasts from the Survey of Professional Forecasters (SPF). In 1992:1 participants in the SPF were asked to switch from forecasting GNP to GDP. In order to continue the ex-post sample through 1993:3, GDP median forecasts from the SPF for 1992:1-1993:3 were modified by adding the actual** *rest of world* **figures as they became available in the Survey of Current Business, so that the GDP forecasts were** *roughly* **transformed to GNP forecasts (This approximation does not have any notable effect on our estimation results, partly because $\Delta logs$ of the GNP variables are taken.) In the forecasting experiments, $\Delta logs$ are used for the first 5 variables in the table. Of the rest, ΔBI and Net X are re-based to 1987 dollars using a simple calculation based on a comparison of overlapping quarters of data in the two different base years. For all of the real national income and product accounts variables, base year changes occurred in 1965:3 (from 1954 to 1958 dollars), in 1975:4 (from 1958 to 1972 dollars), in 1985:4 (from 1972 to 1982 dollars), and in 1991:3 (from 1982 to 1987 dollars). For a more complete discussion of the data the reader is referred to Swanson and White (1995) and Swanson (1996).**

Table 6: Winning Model Selection Techniques Based on Ex Ante Performance[1]

SPF Forecasts are for Period *t* and are Based on Information Available Midway Through Quarter *t*

h	Ex Ante Criterion	Comb Tech	FC1-L	FC1-D	FC2-L	FC2-D	FC3-L	FC3-D
h=1	MSE	M-ALL	1	1	2	3	4	5
		M-AIC	5	8	5	4	4	3
		M-SIC	7	8	6	3	4	2
		M-TST	3	2	4	3	5	5
		M-MSE	3	5	5	3	3	2
	MAD	M-ALL	1	2	2	2	4	4
		M-AIC	5	7	5	4	4	4
		M-SIC	7	7	6	5	4	3
		M-TST	3	3	4	4	5	5
		M-MSE	3	4	5	3	3	3
	MAPE	M-ALL	3	1	4	1	3	4
		M-AIC	3	8	3	4	4	4
		M-SIC	5	8	3	5	4	3
		M-TST	5	2	3	5	6	5
		M-MSE	4	5	3	3	3	3

[1] **Each block of entries (e.g. FC1-L and M-ALL to M-MSE) tabulates the number of times that each model selection technique picks the "best" combination of forecast models, based on ex ante mean square forecast error (MSE), mean absolute forecast error deviation (MAD), and mean absolute percentage forecast error (MAPE). The first block of entries sums to 19, rather than 9, which is the number of variables forecast in the experiment (see Table 5). This is because more than one model selection technique sometimes pick the "best" group of forecasts. The "best" group of forecasts are selected from among linear and nonlinear econometric forecasts and SPF forecasts. The M-MSE combination technique (Comb Tech) is in-sample based, and the statistic used is normalized by T-k, where T is the in-sample size (here set to 100 observations), and k is the number of parameters estimated. The other combination techniques are discussed above. All calculations are based on the forecast period 1982:3-1993:3 (45 observations), and forecast horizons *h*=1 and *h*=4. "Best" groups are selected using a "training set" of 25 observations, and then fixed. The remaining 20 observations are then used to construct *ex ante* forecasts and model selection criteria (i.e. the *ex ante criteria* used are MSE, MAD, and MAPE). Thus, the *ex ante* comparison fixes the "best" group of regressors chosen by each technique, and does not re-estimate the weights as rolling *ex ante* 1- and 4- step ahead forecasts are constructed. The procedure is repeated for various forecast combination regression models, denoted by (FC1-L)-(FC3-L) for levels, and (FC1-D)-(FC3-D) for difference regressions. FC3 corresponds to our most general specification, where a constant is added to the regression, and the weights are not restricted to sum to one (see above).**

Table 7: Winning Model Selection Techniques Based on Ex Ante Performance[1]

SPF Forecasts are for Period $t+1$ and are Based on Information Available Midway Through Quarter t

h	Ex Ante Criterion	Comb Tech	FC1-L	FC1-D	FC2-L	FC2-D	FC3-L	FC3-D
h=1	MSE	M-ALL	2	3	2	6	6	5
		M-AIC	4	4	4	2	1	3
		M-SIC	5	6	5	2	3	4
		M-TST	2	4	4	1	2	2
		M-MSE	3	3	4	3	1	4
	MAD	M-ALL	1	2	3	5	5	5
		M-AIC	4	5	3	2	3	4
		M-SIC	5	7	4	3	5	5
		M-TST	3	3	5	3	2	2
		M-MSE	5	4	3	3	2	4
	MAPE	M-ALL	2	0	3	1	5	4
		M-AIC	2	7	2	3	2	5
		M-SIC	3	9	2	6	4	6
		M-TST	4	1	6	5	3	2
		M-MSE	5	6	2	2	3	3

[1] **See notes to Table 6.**

Table 8: Winning Forecast Models by Model Selection Technique and Forecast Combination Regression[1]

SPF Forecasts are for Period t and are Based on Information Available Midway Through Quarter t

Model	Comb Tech	FC1-L	FC1-D	FC2-L	FC2-D	FC3-L	FC3-D
LINEAR	M-ALL	0	0	0	0	0	0
	M-AIC	2	1	3	1	0	0
	M-SIC	4	2	5	2	0	0
	M-TST	1	1	5	3	0	0
	M-MSE	2	0	2	0	0	0
ADAPTIVE	M-ALL	0	0	0	0	0	0
	M-AIC	2	4	1	3	0	0
	M-SIC	3	4	1	3	0	0
	M-TST	0	0	0	1	0	0
	M-MSE	2	2	1	1	0	0
BOTH	M-ALL	0	0	0	0	0	0
	M-AIC	3	0	2	0	3	4
	M-SIC	0	0	0	0	3	4
	M-TST	2	0	2	1	3	3
	M-MSE	3	3	3	4	2	2
SPF	M-ALL	0	0	0	0	0	0
	M-AIC	2	2	2	4	0	0
	M-SIC	2	3	3	4	0	0
	M-TST	0	0	1	0	0	0
	M-MSE	1	2	1	2	0	0
OTHER	M-ALL	9	9	9	9	9	9
	M-AIC	0	2	1	1	6	5
	M-SIC	0	0	0	0	6	5
	M-TST	6	8	1	4	6	6
	M-MSE	1	2	2	2	7	7

[1] See notes to Table 7. "Best" groups of forecasts from among the nonadaptive linear (LINEAR), the adaptive linear and nonlinear (ADAPTIVE), and the Survey of Professional Forecasters (SPF) methods are chosen for all nine variables, using the entire forecast period - 1982:3-1993:3 - for the "training set". When the "best" group contains both econometric forecasts (i.e. LINEAR and ADAPTIVE), then a "win" is tallied for "BOTH". In addition, "wins" are recorded for each "individual" forecasting model, when only that particular forecasting model is chosen as "best". All other cases (e.g. when the "best" group of forecasts contains the LINEAR and the SPF forecasting models) are recorded in the category "OTHER". The entries in the table indicate the number of times that the different models "win" based on all nine macroeconomic variables, using different forecast combination techniques, and using different forecast combination regressions. For example, for FC1-D and h=1, note that the AIC based technique suggests that the nonadaptive linear model wins once, the adaptive linear and nonlinear model wins four times, the SPF model wins twice, and there are two instances where the "best" group contains both econometric and SPF forecasts (shown as an "OTHER" entry). The total sums to nine, corresponding to all nine variables.

Table 9: Winning Forecast Models by Model Selection Technique and Forecast Combination Regression[1]

SPF Forecasts are for Period $t+1$ and are Based on Information Available Midway Through Quarter t

Model	Comb Tech	FC1-L	FC1-D	FC2-L	FC2-D	FC3-L	FC3-D
LINEAR	M-ALL	0	0	0	0	0	0
	M-AIC	3	1	3	1	0	0
	M-SIC	5	1	4	1	0	0
	M-TST	1	1	5	3	0	0
	M-MSE	2	1	3	1	0	0
ADAPTIVE	M-ALL	0	0	0	0	0	0
	M-AIC	2	2	2	2	0	0
	M-SIC	2	3	2	3	0	0
	M-TST	0	0	0	1	0	0
	M-MSE	2	1	2	0	0	0
BOTH	M-ALL	0	0	0	0	0	0
	M-AIC	2	0	0	0	3	3
	M-SIC	0	0	0	0	3	3
	M-TST	2	0	1	0	3	3
	M-MSE	3	2	0	3	3	3
SPF	M-ALL	0	0	0	0	0	0
	M-AIC	2	3	2	3	0	0
	M-SIC	2	4	2	4	0	0
	M-TST	0	0	0	0	0	0
	M-MSE	0	1	1	1	0	0
OTHER	M-ALL	9	9	9	9	9	9
	M-AIC	0	3	2	3	6	6
	M-SIC	0	1	1	1	6	6
	M-TST	6	8	3	5	6	6
	M-MSE	2	4	3	4	6	6

[1] **See notes to Table 8.**