Macroeconomic Forecasting Using Many Predictors

July 2000
(Revised August 15, 2000)

Mark W. Watson*
Department of Economics and Woodrow Wilson School
Princeton University

* This paper is based on research carried out jointly with James H. Stock, who I thank for comments on this paper. I thank Jean-Philippe Laforte for research assistance. This research was supported by the National Science Foundation (SBR-9730489).

(Version WC_2b)

1. Introduction

The last twenty-five years has seen enormous intellectual effort and progress on the development of small-scale macroeconometric models. Indeed, standing in the year 2000, it is not too much of an overstatement to say that the analysis of small macroeconometric models in a stationary environment is largely a completed research topic. In particular, we have complete theories of estimation, inference and identification in stationary vector autoregressions (VARs). Threshold autoregressions and Markov switching models capture the limited amount of nonlinearity in macroeconomic relations, at least for countries like the United States. We have accumulated a vast amount of experience using these small-scale models for empirical analysis. Identified VARs have become the workhorse models for estimating the dynamic effects of policy changes and for answering questions about the sources of business cycle variability. Both univariate autoregressions and VARs are now standard benchmarks used to evaluate economic forecasts. Although work remains to be done, great progress has been made on the complications associated with nonstationarity, both in the form of the extreme persistence often found in macroeconomic time series and in detecting and modeling instability in economic relations.

Despite this enormous progress, it is also not too much of an overstatement to say that these small-scale macroeconometric models have had little effect on practical macroeconomic forecasting and policymaking.[1] There are several reasons for this, but the most obvious is the inherent defect of small models: they include only a small number of variables. Practical forecasters and policymakers find it useful to extract information from many more series than are typically included in a VAR.

This mismatch between standard macroeconometric models and real world practice has led to two unfortunate consequences. First, forecasters have had to rely on informal methods to distill information from the available data, and their published

[1] Some might take issue with this broad assertion, and point to, for example, the VAR forecasting framework used for many years at the Federal Reserve Bank of Minnesota. I would argue, however, that such examples are the exception rather than the rule.

forecasts reflect considerable judgement in place of formal statistical analysis. Forecasts are impossible to reproduce, and this makes economic forecasting a largely non-scientific activity. Since it is difficult to disentangle a forecaster's model and judgement, a predictive track record can tell us more about a person's insight than about the veracity of his model and the nature of the macroeconomy. The second unfortunate consequence is that formal small-scale macroeconometric models have little effect on day-to-day policy decisions, making these decisions more *ad hoc* and less predicable than if guided by the kind of empirical analysis that follows from careful statistical modeling.

In this paper I discuss a direction in macroeconometrics that explicitly incorporates information from a large number of macroeconomic variables into formal statistical models. The goal of this research is to use the wide range of economic variables that practical forecasters and macroeconomic policymakers have found useful, while at the same time maintaining the discipline and structure of formal econometric models. This research program is not new (several of the important contributions date from the 1970's and earlier), but it is still immature. There are few theoretical results and fewer empirical results, at least compared to small-scale models. Yet, the results that we do have suggest that there may be large payoffs from this "large model" approach. The purpose of this paper is to summarize some these results – particularly those relating to forecasting – to highlight some of the potential gains from these large models.

Section 2 begins the discussion by contrasting two approaches to macroeconometric forecasting. The first, the standard small-model method, constructs a forecasting equation using only a few variables; the second, a large-model method, uses information in a large number of variables. The main problem to be solved when constructing a small model is to choose the correct variables to include in the equation. This is the familiar problem of variable selection in regression analysis. Economic theory is of some help, but usually suggests large categories of variables (money, interest rates, wages, stock prices, etc.) and the choice of a specific subset of variables then becomes a statistical problem. The large-model approach is again guided by economic theory for choosing categories of variables, and the statistical problem then becomes how to combine the information in this large collection of variables.

Sections 3 and 4 present empirical evidence on the relative merits of the small and large model approaches. Section 3 uses monthly U.S. data on 160 times series from 1959-1998 and examines the properties of regressions that include all of these 160 variables as regressors. The empirical analysis focuses on two questions. First, are the regressions characterized by only a few non-zero coefficients? If so, then only a few variables are needed, and there is no gain from looking across many series (except perhaps to choose the best variables to include in the model). Alternatively, if the regressions appear to have a large number of small, but non-zero coefficients, then a large model approach that incorporates all of the variables may be more useful. Importantly, the empirical evidence summarized in Section 3 suggests that there are many non-zero regression coefficients, and thus provides support for the large model approach. The second question taken up in this section focuses on the size of the marginal gains from including additional regressors. That is, are the predictors sufficiently correlated so that the bulk of their predictive content is contained in, say only 30 variables, or are there large additional gains from including 50 or 100 or 150 variables? Here, the empirical evidence suggests that the marginal predictive gain of including an additional regressors is a sharply decreasing function of the number of predictors, a result that is shown to be consistent with a factor analytic structure linking the regressors.

Section 4 uses these same data in a simulated out-of-sample forecasting experiment. This experiment focuses on 12-month ahead forecasts constructed over the 1970-1997 sample period. Forecasts are constructed using three small models and one large model. The small models are a univariate autoregression (which serves as a benchmark), a model that includes the variables present in a typical VAR (output, prices, interest rates and commodity prices), and a model that includes a set of leading indicators suggested by previous forecasting comparisons. The large model is a version of the factor forecasting model developed in Stock and Watson (1998). This model summarizes the regressors using a few common factors that are used to construct the forecasts. For the majority of series studied, the factor model produces forecasts that are considerably more accurate than any of the small models.

The final section of the paper begins with additional discussion of the results. It then continues with some speculative remarks about the future of this research program and lists several outstanding problems.

2. Small and Large Models

2.1 General Framework

This paper considers forecasting in a standard linear regression framework

$$y_{t+1} = x_t' b + e_{t+1} \tag{2.1}$$

where y is the variable to be forecast using the vector of variables x, and the error term satisfies $E(e_{t+1} \mid \{x_t, y_t\}_{t=-\infty}^{t}) = 0$. The dating of the variables in (2.1) emphasizes that it is a forecasting equation and is specified as a 1-period forecast horizon for notational ease. (The empirical work presented below uses a 12-month forecast horizon.) The equation does not include lagged values of y, and this too is for notational convenience. More substantively, (2.1) is specified as a time invariant linear regression and so abstracts from both instability and from nonlinearity. The sample data are $\{y_t, x_t\}_{t=1}^{T}$, and the goal is to forecast y_{T+1}. The forecasts are constructed as $\hat{y}_{T+1} = x_T' \hat{b}_T$ where $\hat{b}_T$ is an estimate of b constructed from the sample information. Forecast loss is quadratic, and so the forecast risk function is

$$R(\hat{b}_T, b) = s_e^2 + E[(\hat{b}_T - b)' x_T' x_T (\hat{b}_T - b)] \tag{2.2}$$

Thus far, this all standard.

The only non-standard assumption concerns the number of regressors: x is an $n \times 1$ vector, where n is large. Thus, in contrast to standard large-T analysis of the regression model, here the analysis is carried out using a framework that assumes both large n and large T.

With this notation fixed, we now consider small-model and large-model approaches the problem of estimating b and to forecasting y.

2.2 Small-Model Approach

To interpret (2.1) as a small model, suppose that only k of the elements of the vector b are non-zero, where k is a small number. If the indices of the non-zero elements are known, then analysis is straightforward: x is partitioned as $x' = (x_1', x_2')$ where x_1 contains the k elements of x corresponding to the non-zero values of b and x_2 contains the remaining "irrelevant" variables. The regression coefficients are estimated by regressing y onto x_1, excluding x_2 from the regression. Imagining that k remains fixed as T grows large, this yields a consistent estimator of b, and sampling error disappears from the limiting value of the forecasting risk: $lim_{T\to\infty} R(\hat{b}_T,\ b) = s_e^2$.

The only remaining statistical problem becomes choosing the elements of x_1 from the vector x. This is the well known variable-selection problem, and there are many standard methods for consistent variable selection. To highlight the key results for the forecasting problem, consider a special version of (2.1) with *iidN(0,1)* disturbances and strictly exogenous orthogonal and standardized regressors (so that $T^{-1}\sum x_t' x_t = I_n$). Then, the OLS estimators are $\hat{b}_{T,i} = T^{-1}\sum_{t=1}^{T} x_{it} y_t$, and the normality of the errors implies that $\hat{b}_{T,i} \sim iidN(b_i, T^{-1})$. Consider a simple variable-selection rule that chooses the r variables with the largest estimated coefficients. Suppose that r is fixed (not a function of T) with $r>k$. The key features of this variable-selection procedure follow from well known results about the asymptotic behavior of order statistics. Let J denote the set of indices for the elements of b that are equal to 0. Let $b_{nT} = \max_{i\in J} |\hat{b}_{T,i}|$, then

$$\left(\frac{T}{2\log(n)}\right)^{1/2} b_{nT} \xrightarrow{as} 1$$

as n and T® ¥ (Galambos (1987)). Thus, if $n=o(e^T)$ then $b_{nT} \xrightarrow{as} 0$.

Three important asymptotic properties follow directly from this result. First, with high probability the correct regressors (those in x_1) will be chosen for inclusion in the regression. Second, the estimated coefficients on the remaining r-k irrelevant selected variables will have coefficients that are close to zero. Finally, the implied forecasts have a limiting risk that is unaffected by sampling error in the estimated coefficients.

Taken together, this is good news for the small model approach: if the number of variables that enter (2.1) is small and if the sample size is large, then sampling error associated with estimating the regression coefficients is likely to be small. Moreover, this conclusion continues the hold even if the variables chosen to enter (2.1) are determined by variable selection over a large number of possible models.

2.2 Large-Model Approaches

2.2.1 Factor Models

Macroeconomists naturally think of the comovement in economic time series as arising largely from a relatively few key economic factors like productivity, monetary policy, etc. One way of representing this notion is in terms of a statistical factor model

$$x_t = \Lambda F_t + u_t \tag{2.3}$$

which explains comovement among the variables x_t using a small number of latent factors F_t.[2] While the classic factor model (Lawley and Maxwell (1971)) assumes that the elements of u_t are mutually uncorrelated both cross sectionally and temporally, this can be relaxed to allow temporal dependence among the u*'s* (Sargent and Sims (1977), Geweke (1977), Engle and Watson (1981)), limited cross sectional dependence

[2] Versions of this framework have been explicitly used in empirical economics beginning with Sargent and Sims (1977) and Geweke (1977), but models like this are implicit in earlier discussion of the business cycle. For example, Stock and Watson (1989) discuss the index of NBER's index of coincident of leading indicators using a model much like (2.3)

(Chamberlain and Rothschild 1983), Connor and Korajczyk (1986)) or both (Forni, Hallin, Lippi and Reichlin (1998), Stock and Watson (1998).)

Pushing the factor model one step further, these latent factors might also explain the predictive relationship between x_t and y_{t+1}, so that

$$y_{t+1} = a F_t + e_{t+1} \tag{2.4}$$

where $E(e_{t+1} \mid \{y_t, x_t, F_t\}_{t=-\infty}^{t}) = 0$. As discussed in Stock and Watson (1998), (2.3) and (2.4) is a potentially useful framework for large model forecasting because it imposes important constraints on the large number of regression coefficients in (2.1). Notably, if we write $E(F_t \mid x_t) = g x_t$ (so that regression of F onto x is linear), then (2.4) implies that $E(y_{t+1} \mid x_t) = a E(F_t \mid x_t) = ag x_t$, so that the regression coefficients b in (2.1) satisfy $b = ag$.

This model yields values of b that are quite different than the small model specification. In that model, most of the elements of b are 0, and only a few non-zero elements account for the predictive power in the regression. In the factor model, all of the elements of b are non-zero in general, but they are each small. This follows because the coefficients g in the regression $E(F_t \mid x_t) = g x_t$ will, in general, be $O(n^{-1})$ (the estimated value of F will be an average of the n elements in x), and $b = ag$. Section 3 will examine sample versions of (2.1) for several macro variables and ask whether the distribution of the elements of b appears to be more consistent with the small model (many zero and a few large non-zero values) or the large model (many small non-zero values).

If the factor model can describe the data, then forecasting becomes a three step process. First, an estimate of $E(F_t \mid x_t)$, say $\hat{F}_t$, is constructed from the x data. Second, the coefficients a in (2.4) are estimated by regressing y_{t+1} onto $\hat{F}_t$. Finally, the forecast is formed as $\hat{y}_{T+1} = \hat{a} \hat{F}_T$.

Somewhat remarkably, this procedure can work quite well, even when $\hat{F}_t$ is a very naïve estimator of $E(F_t \mid x_t)$. For example, Stock and Watson (1998) show that under limited cross sectional and temporal dependence (essentially *I(0)* dependence both temporally and cross-sectionally), an identification condition on the factor loadings *(L'L* must be well-behaved), and the number of regressors is sufficiently large ($n = O(T^r)$ for any $r > 0$), then $\hat{a}\hat{F}_T - b'x_T \xrightarrow{p} 0$. That is, to first order, the feasible forecast constructed from the factor model is equal to the forecast that would be constructed knowing the true value of b. This is the same result obtained in the small model. Even though many more parameters must be estimated in the factor model, the same result is possible because the large amount of cross-sectional information in the *x's* is exploited to estimate the predictive relation linking y_{t+1} and x_t.

An important practical problem is determination of the number of factors to use in the analysis. There are two ways to do this. In the context of the forecasting problem, Stock and Watson (1998) propose a modification of the standard information criterion applied to the regression of y_{t+1} onto $\hat{F}_t$. They show that the resulting estimated estimate is consistent and the resulting forecasts are asymptotically efficient. Alternatively, Bai and Ng (2000) propose estimators that are based on the fit of (2.3) and show consistency under assumptions similar to those used in Stock and Watson (1998).

2.2.3 Non-Factor Models

Large-model methods have also been proposed that do not rely on the factor framework. The most obvious is based on the OLS estimator of b, which in general is well-defined when *n£T*. However, when the ratio of estimated parameters to observations *(n/T)* is large, sampling error in the estimated coefficients will be large, and this suggests that significant improvements may be possible. From the classic result in Stein (1955), OLS estimates and forecasts are inadmissible when $n \geq 3$, and shrinkage estimators (such as the classic James-Stein estimator (1960)) dominate OLS. While these results are largely irrelevant in the large-T and small-n model (since the risk of the OLS forecast converges to the risk of the infeasible known-b forecast), their relevance resurfaces in the large-T and large-n model. For example, when $n = rT$ with $0 < r < 1$, then the OLS estimator of b is not consistent and the sampling error in b continues to have a first order effect on the forecast risk as $T \to \infty$. Shrinkage estimators can produce forecasts with asymptotic relative risks that dominate OLS, even asymptotically.

This idea is developed in Knox, Stock and Watson (2000) where empirical Bayes estimators for b and forecasts from these estimators are constructed. They study parametric and non-parametric empirical Bayes estimators and provide conditions under which the large-*n* assumption can be used to construct efficient forecasts (efficient in the sense that in large samples they achieve the same Bayes-risk as the optimal forecasts constructed using knowledge of the distribution of the regression coefficients). To date, the empirical performance of these methods have not been systematically studied, and so it is premature to judge their usefulness for macroeconomic forecasting.

With this background we now proceed to a discussion of some empirical evidence on the relative merits of the small-model and large-model approaches to forecasting.

3. Full-Sample Empirical Evidence

Two empirical questions are addressed in this section. First, in regressions like (2.1), does it appear that only a few of the coefficients are non-zero? If so, then forecasts should be constructed using a small model. Alternatively, are there many non-zero (but perhaps small) regression coefficients? If so, then the large model framework is appropriate. Second, and more generally, how does the regression's predictive R^2 change as the number of regressors is increased?

3.1 Data[3]

These questions are investigated using data on 160 monthly macroeconomic U.S. time series from 1959:1-1998:12. The data represent a wide range of macroeconomic activity and are usefully grouped into 8 categories: Output and Real Income (21 series); Employment and Unemployment (27 series); Consumption, Sales and Housing (22 series); Inventories and Orders (27 series); Prices and Wages (22 series); Money and

[3] These data were used in Stock and Watson (1999b).

Credit (9 series); Interest Rates (19 series); and Exchange Rates and Stock Prices/Volume (13 series). These categories are perhaps overly broad, but it will be useful to have many series in each category, and so this coarse aggregation is necessary.

The data were transformed in three ways. First, many of the series are seasonally adjusted by the reporting agency. Second, the data were transformed to eliminate trends and obvious nonstationarities. For real variables, this typically involved transformation to growth rates (the first difference of logarithms), and for prices this involved transformation to changes in growth rates (the second difference of logarithms). Interest rates were transformed to first differences or to "spreads." Finally, some of the series contained a few large outliers associated with events like labor disputes, other extreme events, or with data problems of various sorts. These outliers were identified as observations that differed from the sample median by more than 6 times the sample interquartile range, and these observations were dropped from the analysis.

A detailed description of the data can be found in the appendix.

3.2 Forecasting Regressions

Equation (2.1) was modified in two ways for the empirical analysis. First, the dependent variable was changed to focus on 12-month ahead forecasts, and second, autoregressive lags were added to the regression. The modified regression takes the form:

$$y_{t+12}^{12} = x_t'b + f(L)y_t + e_t \qquad (3.1)$$

where y_t is the transformed variable of interest (as described above), and y_{t+12}^{12} denotes a further transformation of the variable appropriate for multi-step forecasting. For example, if z_t denotes the raw (untransformed) value of the variable of interest and y_t is the monthly rate of growth of z_t, then $y_{t+12}^{12} = ln(z_{t+12}/z_t)$ is the rate of growth over the forecast period. When y_t is the level of z_t then $y_{t+12}^{12} = z_{t+12}$, and when y_t denotes the change in the monthly growth rate, then $y_{t+12}^{12} = ln(z_{t+12}/z_t) - 12ln(z_t/z_{t-1})$.

3.3 Empirical Distribution of t-statistics

Equation (3.1) was estimated by OLS for all 160 series in the data set. Each regression included a constant, current and 5 lagged values of y_t, and an x vector containing the remaining 159 series (transformed as described above). Heteroskedastic and autoregressive consistent t-statistics were computed for the estimated b's. This yielded 25440 t-statistics (160 regressions each including 159 x-regressors). Let t_{ij} denote the t-statistic for the i'th b in the j'th equation. Then, approximately, t_{ij}~$N(m_{ij},1)$ where m_{ij} is value of the non-centrality parameter for the t-statistic. If $m_{ij} \neq 0$ then variable i is a useful predictor for y_{jt}, otherwise variable i is not a useful predictor. We can determine the fraction of useful predictors by determining the fraction of non-zero values of m_{ij}. Of course m_{ij} is not directly observed, but it is related to the t-statistics by $t_{ij} = m_{ij} + e_{ij}$ where (approximately) e_{ij}~$N(0,1)$ and is independent of m_{ij}. This implies that the distribution of m can then be estimated from the empirical distribution of t using deconvolution techniques. (The densities of t and m are related by $f_t(x) = \int f(x-s) f_m(s) ds$, where f_τ is the density of t, f_m is the density of m, and f is the standard normal density.)

Figure 3.1 shows the estimated CDF of m estimated from the 25440 values of t_{ij} using the deconvolution method outlined in Diggle and Hall (1993). The estimated CDF suggests that there are a large number of non-zero, but small values of b. Over 55% the non-centrality parameters exceed 0.5, 33% of the values lie between 0.5 and 1.0, 20% between 1.0 and 2.0, and only 2% of the values are above 2.0. Thus, the large-model assumptions seem to be a better characterization of these data than do the small-model assumptions.

Table 3.1 summarizes the distribution of the absolute values of the t-statistics for selected variables. The first row of the table shows quartiles of the standard normal as a benchmark, and the next row shows the quartiles of the empirical distribution of all of the 25440 t-statistics from the estimated regressions. The median absolute t-statistic was 0.84, which can be compared to a value of 0.67 for the normal distribution. Again, this suggests that a large fraction of the values of b in (3.1) are non-zero. The remaining rows of the table show the distribution across predictive regressions of the absolute t-statistics for specific regressors. These distributions show whether the particular regressor has marginal predictive power for the range of macroeconomics variables in this data set. The first four entries are for the variables that are typically used in small VAR models: Industrial Production (IP), consumer price inflation (PUNEW), the Federal Funds Rate (FYFF), and Commodity Prices (PMCP). With the exception of price inflation, these variables all perform better than the typical variable in the dataset. The Federal Funds rate is particularly noteworthy with a median absolute t-statistic that exceeds 1.5. The next ten variables listed in the table correspond to the variables that make up the Conference Board's Index of Leading Indicators (previously published by the U.S. Department of Commerce).[4] Six of the ten variables are useful predictors for a large fraction of the variables considered. Notably, the number of newly unemployed (LHU5) had the largest median t-statistic of all of the variables considered. This variable had an absolute t-statistics that exceeded 1.0 in 72% of the regressions. On the other hand, four of the Conference Department's indicators (the new orders and deliveries series MOCMQ, PMDEL, and MSONDQ and stock prices FSPCOM), had little predictive power, and t-statistic distributions that are very close to that of the standard normal. The remaining rows of the table show results for other variables that have been widely discussed as leading indicators. The public-private interest rate spread (PPSPR) and the unemployment rate (LHUR) have large t-statistics in a majority of the regressions. The nominal money supply (FM1 and FM2) are not very useful, although the real value of M2 (FM2DQ) was more useful. Unfilled orders and inventories have median t-statistics that were nearly equal to unity, suggesting some useful predictive power in a large number of estimated regressions. Two other measures of commodity prices (PSM99Q and PWCMSA) are not useful predictors.

[4] Because of the data set used here did not include all of the Conference Board's indicators, there are two changes from the Conference Board's list. The number of new unemployment insurance claims is replaced by the number unemployed for less than 5 weeks (LHU5) and the 5-Year Treasury Bond Rate is used in place of the 10-year rate in the term spread.

Of course, results based on these t-statistics may be misleading for several reasons. Most importantly, they assume a N(0,1) distribution for the sampling error in the t-statistics, and this may be a poor approximation in the setting with a large number of regressors. However, these results are suggestive and are consistent with the other empirical results presented below that favor the large-model approach.

3.4 Prediction R^2's using different numbers of predictors

While the empirical results in the last section suggest that "many" variables have marginal predictive content in (3.1), the results don't quantify the marginal gain from including, say 50 variables instead of 25, or 150 instead of 100. If we let $R^2(k)$ denote the value of the population R^2 from the regression of y onto k elements of x, then the question is" "How does $R^2(k)$ change as k increases"? The answer is important for forecast design, but more importantly, it provides information about the way macroeconomic variable interact. For example, suppose that each element of x contains important information about the future that cannot be gleaned from the other elements of x. (Housing starts in the Midwest contains some important information not contained in aggregate housing starts, the unemployment rate or any of the other variables.) In this case $R^2(k)$ will be a steadily increasing function, with the amount of information increasing in proportion to the number of predictors.

Alternatively, suppose that macro variables interact in the simple low-dimensional way suggested by the factor model. In this case, each regressor will contain useful information about the values of the factors (and hence useful information about future values of y), but the marginal value of particular regressor will depend on how many other variables have already been used. That is, if V_k denotes the variance of F_t conditional on the first k elements of x_t, then V_k-V_{k-1} decreases as k increases. To see this most easily, consider the simplest case when there is only one factor, the uniquenesses (u in (2.3)) are uncorrelated and homoskedastic, and the factor loadings (L in (2.3)) are all unity. In this case the evolution of V_k is particulary simple: $V_k=[V_{k-1}/(s_u^2+V_{k-1})]V_{k-1}$, so that the variance decreases at a sharply decreasing rate as k increases. This implies that

$R^2(k)$ will increase at a sharply decreasing rate: there can be large gains from increasing the number of predictors from 5 to 25, but negligible gains from increasing the number of predictors from 100 to 120.

Stock and Watson (2000b) discuss the problem of estimating the population R^2 in equations like (2.1) and (3.1) when the number of regressors is proportional to the sample size: $k = rT$ where $0 < r < 1$. In a classical version of (2.1) the usual degrees of freedom adjustment ($\bar{R}^2$) is all that is necessary to produce a consistent of the population R^2. However, things are more complicated in a model like (3.1) with predetermined but not exogenous regressors and an error terms that is serially correlated and/or conditionally heteroskedastic. In this case an alternative estimator must be used. Here I use a split sample estimator. Let $\hat{b}_1$ denote the least squares estimator of b using the first half of the sample, and $\hat{b}_2$ denote the estimator using the second half of the sample. Under general conditions, these two estimators will be approximately uncorrelated so that $n^{-1}\hat{b}_1'\hat{b}_2 \xrightarrow{p} b'b$, $T^{-1}\hat{b}_1'\sum_t x_t'x_t\hat{b}_2 \xrightarrow{p} E(b'x_t'x_tb)$, and $R_S^2 = (\hat{b}_1 \sum x_t'x_t \hat{b}_2 / \sum y_t^2)$ will provide a consistent estimator of the population R^2. Here, a partial R^2 version of this estimator is applied to (3.1) after controlling for the constant term and lagged values of y.

The results are summarized in Figure 3.2. This figure plots the average estimated R^2 as a function of k. These values were computed as follows. First, a random set of k regressors was selected from the 160 available regressors. Equation (3.1) was then estimated for each of the y-variables, and the split sample R^2 (as described above) was computed after controlling for the lagged values of y in the regression. This process was repeated 200 times, and the plot shows the average estimated R^2 as a function of k, where the average is computed across replications and across dependent variables.

The figure show a sharp increase in the R^2 as the number of regressors increases from $k=5$ to $k=50$ ($R^2(5)=.10$, $R^2(50)=.32$), but a much small increase as k increases from 100 to 150 ($R^2(100)=.35$, $R^2(150)=.37$). Thus, while many more variables appear to be useful than are typically used in small-scale models, the predictive component of the regressors is apparently common to many series in a way suggested by the factor model.

4. Simulated Out-of-Sample Empirical Evidence

This section uses a simulated real time forecasting experiment to compare the forecasting performance of several small and large model forecasting methods. The experiment is similar to the experiments reported in Stock and Watson (1999, 2000a) which studied forecasts of four measures of real activity and four measures of price inflation. Here forecasts for all 160 series in the data set are studied.

4.1 Experimental Design

The out-of-sample forecasting performance is simulated by recursively applying the forecasting procedures to construct 12-month ahead forecasts for each month of the sample beginning in T=1970:1 through T=1997:12. Small model forecasts were computed using regression models of the form

$$y_{t+12}^{12} = a + b(L)w_t + f(L)y_t + e_{t+12} \qquad (4.1)$$

where y_{t+12}^{12} and y_t were defined in section 3 and w_t is a (small) vector of predictors. Two versions of w_t were used. In the first w_t includes the four variables typically included in monthly VARs (industrial production (IP), CPI inflation (PUNEW), the federal funds interest rate (FYFF) and commodity prices (PMCP)). In the second, w_t includes a set of eleven leading indicators that previous researchers have identified as useful predictors for real activity (three labor force indicators (LPHRM, LHEL and LHNAPS), vendor performance (PMDEL), capacity utilization (IPXMCA), housing starts (HSBR), the public-private interest and long-short interest rate spreads (PPSPR and TBSPR), long term interest rates (FYGT10), the help wanted index, exchange rates (EXRATE), and unfilled orders (MDU82).) These are the leading indicators used in the Stock-Watson (1989) XLI and non-financial XLI. The forecasts from the first model will be referred to as the VAR forecasts (even though they are computed directly from (4.1) instead of iterating the 1-step VAR forward) and the second set of forecasts will be called leading indicator forecasts. Forecasts from a univariate autoregressive model were also computed from (4.1) after eliminating w_t.

In all of the models, the order of the lag polynomials $b(L)$ and $f(L)$ were determined recursively by BIC. In the VAR, $b(L)$ includes between 1 and 6 nonzero coefficients, and in the leading indicator model it includes between 1 and 4; in all of the models, $f(L)$ contains 0-6 non-zero coefficients.

To be specific about the recursive forecasting experiment, consider the forecast dated T=1970:1. To construct this forecast, equation (4.1) was estimated using data from t=1960:1 through t=1969:1, where the terminal date allows $y^{12}_{1970:1}$ to be used as the dependent variable in the last period, and values of data before 1960:1 are used as lags for the initial periods. The lag lengths were determined by BIC using these sample data, and the regression coefficients were estimated conditional on these lag lengths. The forecast for $y^{12}_{1971:1}$ was then constructed as $\hat{y}^{12}_{1971:1} = \hat{a} + \hat{b}(L)w_{1970:1} + \hat{f}(L)y_{1970:1}$. To construct the forecast in 1970:2, the process was repeated with data from 1970:2 added to the data set. This experiment differs from what could have been computed in real time only because of revisions made in the historical data. (The data used here are from a 1999 release of the DRI Basic Macroeconomic Database .)

The experimental design for the factor model forecasts is similar. These forecasts were constructed from (4.1), with $\hat{F}_t$ replacing w_t. The estimated factors, $\hat{F}_t$, were computed recursively (so that the factors for the forecast constructed in 1970:1 were estimated using data on x_t from 1959:3-1970:1, where the initial two observations were lost because of second differencing of some of the variables). The factors were estimated by principal components, with missing values handled by a least squares EM algorithm as described in Stock and Watson (1998). Two sets of forecasts were constructed: the first used only contemporaneous values of the factors in the forecasting equation with the number of factors selected recursively by BIC; in the second, lags of the factors were allowed to enter the equation (and BIC was used to determine the number of factors and the number of lags). Since lags of the factors were nearly never chosen by BIC, the forecasts that allowed for lags were nearly identical to the forecasts that did not include lags. Thus, only results for the no-lag model will be reported.

The forecasting methods will be evaluated using the sample mean square error over the simulated out-of-sample period: $MSE_{ij} = (336)^{-1}\sum_{T=1971:1}^{1998:12}(\hat{y}_{Tij}^{12} - y_{T,i}^{12})^2$ where i indexes the series forecast and j indexes the forecast method. Because the series differ in persistence and are measured in different units, the MSEs relative to the MSE from the univariate autoregression will be reported. The will be referred to as relative MSEs.

4.2 Forecasting Performance of the Models

The forecasting performance of the models is summarized in Figure 4.1. This figure shows the cumulative relative frequency distribution of the relative MSEs for the VAR, leading indicator and factor forecasting models. Each distributions summarizes the relative MSEs over the 160 variables that were forecast by each method. The figure shows that the factor forecasts clearly dominate the small model forecasts. The median relative MSE for the factor model is 0.81 (indicating an 19% MSE gain relative to the univariate autoregression) compared to median values of 0.96 for the VAR and 0.98 for the leading indicator model. For 84% of the series, the factor model improved on the univariate autoregression, compared to 66% for the VAR and 57% for the leading indicator forecasts. For 44% of the series, the factor model produced more than a 25% MSE improvement relative the univariate autoregression (so that the relative MSE was less than 0.75); only 22% of the VAR forecasts and 26% of the leading indicator forecasts achieved improvements this large.

Figure 4.2 presents a summary of the forecasting performance for the different categories of series. For each of the forecasting methods the figure presents box plots of the relative MSEs for each series category. (The outline of the boxes are the 25th and 75th percentiles of the distribution, the median is the horizontal line in the box and the vertical lines show the range of the distribution.) Looking first at the results for the factor model (panel c), there are important differences in the forecastability of the series across categories. For example (and not surprisingly), there are negligible forecasting gains for series in the last category which contains exchange rates and stock prices. On the other hand there are large gains for many of the other categories, notably the real activity variables in the production and employment categories, and the nominal variables making up the prices, wages, money and credit categories. In all series categories, the median relative MSE is less than 1.0, and in 5 of the 8 categories, the 75^{th} percentile is less than 1.0.

The forecasting performance of the VAR and leading indicator models are more mixed. For the real variables in the first four categories, these models perform much worse than the factor model, and on average they do not improve on the univariate autoregression (the relative MSE across the variables in these categories is 1.0 for the VAR and 0.99 for the leading indicator model). On the other hand, these models do yield improvements for wages, price, money and credit that are roughly equal to the improvements produced by the factor forecasts.

4.3 Additional Discussion of the Factor Model

What accounts for the strong performance of the factor model and the uneven performance of the other models? While I cannot provide a complete answer to this question, I can offer a few useful clues. Table 4.1 provides the first clue. It shows the relative MSE of several factor models: the model with the number of factors determined by BIC reported in Figures 4.1 and 4.2, and models with 1, 2 and 3 factors. Evidently, only a few factors are responsible for the model's forecasting performance. Two or three factors are useful for some categories of series, but only a single factor is responsible for the predictability of wages, prices, money and credit.

The second clue is provided by Figure 4.3. It plots the estimated first factor along with the index of capacity utilization, one the variables used in the leading indicator model. These series are remarkably similar over much of the sample period. Major peaks and troughs of the estimated factor correspond closely to NBER business cycle dates. Apparently, the first factor is an index of real economic activity, and the forecasting results say that this real activity index is an important predictor of future price and wage inflation. Of course, this is just the well known Phillips relation. Since both of the small models include good measures of the state of the real economy (industrial production in the VAR; capacity utilization, housing starts and several of the other variables in the leading indicator model) these models can exploit the Phillips relation to forecast future wage and price inflation, and this accounts for their good forecasting.

This raises the question: when forecasting future inflation, is a large model necessary, or should inflation be forecast using simple measures of real activity like the unemployment rate, industrial production or capacity utilization? The empirical evidence in Stock and Watson (1999a) provides a clear answer to this question, at least for typical measures of aggregate price inflation in the U.S. An index of real activity constructed from a large number of variables performs better than any single series representing real activity. There appear to be gains from using a large model approach even when the underlying forecasting relation is a simple Phillips relation.

While the first factor is easy to interpret, a complete understanding of the forecasting results requires an understanding of the second and third factors as well. Unfortunately, their interpretation and role in explaining future changes in the real variables is an open question.

5. Discussion

These empirical results raise several issues for economic forecasting and for macroeconometrics more generally. I will use this section to highlight a few of these issues and to discuss a few of the large number of open research questions that they suggest.

Evaluations of the accuracy of macroeconomic forecasts (e.g., Zarnowitz and Braun (1993) consistently find that "consensus" forecasts – averages of forecasts from many sources – are more accurate that individual forecasts. Averaging is a simple, but apparently very effective, large-model forecasting approach. How do the factor forecasts reported here compare to the consensus forecast benchmark? Differences in actual out-of-sample forecasting (which use real time data) and the simulated out-of-sample forecasting carried out here (which use revised data) makes a clean comparison difficult. But, a few calculations are suggestive. LaForte (2000) reports mean square errors for the consensus forecast from the Survey of Professional Forecasters maintained by the Philadelphia Federal Reserve Bank (Croushore (1993)), and computes relative mean square errors using univariate autoregressions recursively estimated using the real time data set constructed by Croushore and Stark (1993). Over the sample period 1969-1998 he reports relative mean square errors of roughly 0.40 for aggregate price inflation (measured by the GNP/GDP price deflator) and the unemployment rate. (The precise value of the relative MSE depends on particular assumptions about the dates that forecasts were constructed and the specification of the univariate autoregression.) This value of 0.40 is only slightly larger than values for price inflation and the unemployment rate that were found here for the simulated forecasts using the factor model. This crude comparison suggests that the information aggregation in the factor model is roughly comparable to current best practice of using consensus forecasts.

While these results are promising, a long list of open questions remain. Some are empirical and some are technical. Let me begin by listing two of the most obvious empirical questions. First, do the results reported here for the U.S. hold for other countries as well? Data limitations will make this question difficult to answer. For example, Marcellino, Stock and Watson (2000) study forecasts of the unemployment rate, inflation and short term interest rates for EMU countries using data on over 500 series from 1982-1998. They find that estimated factors are highly significant for in-sample regressions, but they find inconclusive out-of-sample forecast rankings because of the short sample period. The second question concerns the stability of the forecasting relations. This is important in the United States, where the late 1990's seem much different than the late 1970's, but is arguably more important for Europe which has experienced enormous changes in the past decade.

The list of open technical questions is long. The problems of efficient estimation and inference that have been solved in small models remains an open question in large models. For example, the estimated factors used in the forecasting exercise reported here were constructed by the simplest of methods – principal components. While this estimator is consistent, undoubtedly more efficient estimators can be constructed. The empirical results in Table 4.1 indicated some substantial improvement from using models with a fixed number of factors rather than BIC selected factors, and this suggests that model selection procedures can be improved.

The existing theoretical results cover I(0) models, but say nothing about integrated, cointegrated and cotrending variables. We know that common long-run factors are important for describing macroeconomic data, and theory needs to be developed to handle these features in a large model framework.

The difficult but important issues of nonlinearity and instability must also be addressed. While some may scoff at the notion that there are large gains from modeling nonlinearity in macroeconomic time series, much of this (well-founded) skepticism comes from experience with small models. But large-model results can be quite different. For example, in an experiment involving 12-month ahead forecasts for 215 macroeconomic time series, Stock and Watson (1999b) find that univariate autoregressions generally outperform standard nonlinear models (threshold autoregressions, artificial neural networks). Yet, in what can be interpreted as a large-model forecasting exercise, when forecasts from 50 nonlinear models were averaged, they outperformed any of the linear models, including combined models

Temporal instability has already been mentioned as an open question in the context of the empirical work, but there are important technical questions as well. For example, are large model methods more robust to instability than small model methods? That is, does the cross-sectional averaging in large model methods mitigate the effects of instability, and if so, what kinds of instability? There are already some results that relate to this question: Stock and Watson (1998) show that principal components estimators of factors remain consistent in the presence of some time variation in the factor loadings, but more general results are certainly possible and necessary.

While this paper has focused on the problem of macroeconomic forecasting, the empirical results have more general implications for macroeconometric models. One need only consider the role that expectations play in theoretical models to appreciate this. There is an unfortunate bifurcation in the care in which expectations are handled in macroeconometric models. When expectations are explicitly incorporated in the models, and when interest focuses on a few key parameters, then empirical researchers are careful to use methods (like instrumental variables) that are robust to the limited amount of information contained in small models. However, when expectations are implicitly included, as they are in most identified VARs, then researchers are much more cavalier, raising the possibility of large omitted variable biases. As Sims (1992) and Leeper, Sims and Zha (1996) argue, this can have disastrous effects on inference and on policy advice. For example, Leeper, Sims and Zha show that estimated effects of monetary policy on the macroeconomy decline sharply as they include more variables in their identified VAR to account for the central bank's forward looking behavior. The largest model that they consider includes 18 variables, which is very large by conventional standards, but the results reported above in section 3 suggest that there may be substantial increases in forecastability as the number of variables increases from, say 18 to 50 or to 100. Thus there may still be large biases even in as large a VAR as large as the one constructed in their analysis. The construction of large scale VAR or VAR-like models would seem to be a high priority for the large-model research program.

This paper has really been little more than a tease. It has pointed out important practical problems in the small-scale macroeconometric models that have been developed by researchers over the past twenty-five years. It has suggested that large models may solve many of these problems, so that formal statistical models can play a major role in economic forecasting and macroeconomic policy. A few theoretical results concerning large models were outlined. A set of empirical results were presented that suggest that these new models yield substantial improvements on small scale models, and indeed may perform as well as the current best practice of using consensus forecasts. My hope that others will find this tease intriguing enough to work on these problems and provide answers for the technical questions and empirical experience with the resulting new methods.

References

Chamberlain, G. and M. Rothschild (1983), "Arbitrage Factor Structure, and Mean-Variance Analysis of Large Asset Markets," *Econometrica*, Vol. 51, No. 5.

Connor, G. and R.A. Korajczyk (1986), "Performance Measurement with the Arbitrage Pricing Theory," *Journal of Financial Economics*, 15, 373-394.

Croushore, D. (1993)," Introducing The Survey of Professional Forecasters," Federal Reserve Bank of Philadelphia Business Review (November/December 1993), pp. 3-13.

Croushore, D. and T. Stark (1999), "A Real-Time Data Set for Macroeconomists," Philadelphia Federal Reserve Bank Working Paper 99-4.

Diggle, P.J. and P. Hall (1993), "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," *Journal of the Royal Statistical Society, Series B*, 55, 523-531.

Engle, R.F. and M.W. Watson (1981), "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association*, Vol. 76, No. 376, 1981, pp. 774-781.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (1998), "The Generalized Dynamic Factor Model: Identification and Estimation," manuscript, Universiteÿ08' Libre de Bruxelles.

Galambos, Janos (1987), *The Asymptotic Theory of Extreme Order Statistics*, 2nd Edition, Krieger Publishing: Malabar, Florida.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam, Ch. 19.

James, W. and C. Stein (1960), "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 361-379.

LaForte, J.P (2000), "The Relative Accuracy of the Survey of Professional Forecasters," Research Memorandum, Princeton University.

Lawley, D.N. and A.E. Maxwell (1971), *Factor Analysis as a Statistical Method*, American Elsevier Publishing: New York.

Leeper, E.M., C.A. Sims and T. Zha (1996), "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity*, 1996:2.

Marcellino, M., J.H. Stock and M.W. Watson (2000), "Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-Wide Information," manuscript.

Sargent, T.J, and C.A. Sims (1977), "Business Cycle Modeling without Pretending to have Too Much a-priori Economic Theory," in C. Sims et al. (eds.), *New Methods in Business Cycle Research*, Minneapolis: Federal Reserve Bank of Minneapolis.

Sims, C.A. (1992), "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 975-1001.

Stein, C. (1955), "Inadmissibility of the Usual Estimator for the Mean of Multivariate Normal Distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 197-206.

Stock, J.H. and M.W. Watson (1989), "New Indexes of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 351-393.

_________________ (1998), "Diffusion indexes," NBER Working Paper No. W6702.

_________________ (1999a), "Forecasting Inflation" *Journal of Monetary Economics*, (44) 2, pp. 293-335

_________________ (1999b), "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series," in (R.F. Engle and H.White eds.) *Cointegration, Causality and Forecasting, A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press.

_________________ (2000a), "Macroeconomic Forecasting Using Diffusion Indexes," manuscript.

_________________ (2000b), "The Limits of Predictability," in preparation.

Zarnowitz, V. and P. Braun (1993), "Twenty-two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance," in (J.H. Stock and M.W.Watson, eds.) *Business Cycles, Indicators, and Forecasting*, NBER Studies in Business Cycles, Vol. 28, University of Chicago Press, Chicago.

Appendix

This appendix lists the time series used in the empirical analysis is Section 4 and 5. The series were either taken directly from the DRI-McGraw Hill Basic Economics database, in which case the original mnemonics are used, or they were produced by authors' calculations based on data from that database, in which case the authors' calculations and original DRI/McGraw series mnemonics are summarized in the data description field. Following the series name is a transformation code and a short data description. The value of the transformation code indicates the transformation discussed in section 3.1. The transformations are (1) level of the series; (2) first difference; (3) second difference; (4) logarithm of the series; (5) first difference of the logarithm; (6) second difference of the logarithm. The following abbreviations appear in the data descriptions: SA = seasonally adjusted; NSA = not seasonally adjusted; SAAR = seasonally adjusted at an annual rate; FRB = Federal Reserve Board; AC = Authors' calculations.

<u>Output and income (Out)</u>

1. IP	5	INDUSTRIAL PRODUCTION: TOTAL INDEX (1992=100,SA)
2. IPP	5	INDUSTRIAL PRODUCTION: PRODUCTS, TOTAL (1992=100,SA)
3. IPF	5	INDUSTRIAL PRODUCTION: FINAL PRODUCTS (1992=100,SA)
4. IPC	5	INDUSTRIAL PRODUCTION: CONSUMER GOODS (1992=100,SA)
5. IPCD	5	INDUSTRIAL PRODUCTION: DURABLE CONSUMER GOODS (1992=100,SA)
6. IPCN	5	INDUSTRIAL PRODUCTION: NONDURABLE CONDSUMER GOODS (1992=100,SA)
7. IPE	5	INDUSTRIAL PRODUCTION: BUSINESS EQUIPMENT (1992=100,SA)
8. IPI	5	INDUSTRIAL PRODUCTION: INTERMEDIATE PRODUCTS (1992=100,SA)
9. IPM	5	INDUSTRIAL PRODUCTION: MATERIALS (1992=100,SA)
10. IPMD	5	INDUSTRIAL PRODUCTION: DURABLE GOODS MATERIALS (1992=100,SA)
11. IPMND	5	INDUSTRIAL PRODUCTION: NONDURABLE GOODS MATERIALS (1992=100,SA)
12. IPMFG	5	INDUSTRIAL PRODUCTION: MANUFACTURING (1992=100,SA)
13. IPD	5	INDUSTRIAL PRODUCTION: DURABLE MANUFACTURING (1992=100,SA)
14. IPN	5	INDUSTRIAL PRODUCTION: NONDURABLE MANUFACTURING (1992=100,SA)
15. IPMIN	5	INDUSTRIAL PRODUCTION: MINING (1992=100,SA)
16. IPUT	5	INDUSTRIAL PRODUCTION: UTILITIES (1992-=100,SA)
17. IPXMCA	1	CAPACITY UTIL RATE: MANUFACTURING,TOTAL(% OF CAPACITY,SA)(FRB)
18. PMI	1	PURCHASING MANA GERS' INDEX (SA)
19. PMP	1	NAPM PRODUCTION INDEX (PERCENT)
20. GMPYQ	5	PERSONAL INCOME (CHAINED) (SERIES #52) (BIL 92$,SAAR)
21. GMYXPQ	5	PERSONAL INCOME LESS TRANSFER PAYMENTS (CHAINED) (#51) (BIL 92$,SAAR)

<u>Employment and hours (Emp)</u>

22. LHEL	5	INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;SA)
23. LHELX	4	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF
24. LHEM	5	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)
25. LHNAG	5	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)
26. LHUR	1	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%,SA)
27. LHU680	1	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA)
28. LHU5	1	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)
29. LHU14	1	UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,SA)
30. LHU15	1	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)
31. LHU26	1	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)
32. LPNAG	5	EMPLOYEES ON NONAG. PAYROLLS: TOTAL (THOUS.,SA)
33. LP	5	EMPLOYEES ON NONAG PAYROLLS: TOTAL, PRIVATE (THOUS,SA)
34. LPGD	5	EMPLOYEES ON NONAG. PAYROLLS: GOODS-PRODUCING (THOUS.,SA)
35. LPMI	5	EMPLOYEES ON NONAG. PAYROLLS: MINING (THOUS.,SA)

36. LPCC 5 EMPLOYEES ON NONAG. PAYROLLS: CONTRACT CONSTRUCTION (THOUS.,SA)
37. LPEM 5 EMPLOYEES ON NONAG. PAYROLLS: MANUFACTURING (THOUS.,SA)
38. LPED 5 EMPLOYEES ON NONAG. PAYROLLS: DURABLE GOODS (THOUS.,SA)
39. LPEN 5 EMPLOYEES ON NONAG. PAYROLLS: NONDURABLE GOODS (THOUS.,SA)
40. LPSP 5 EMPLOYEES ON NONAG. PAYROLLS: SERVICE-PRODUCING (THOUS.,SA)
41. LPTU 5 EMPLOYEES ON NONAG. PAYROLLS: TRANS. & PUBLIC UTILITIES (THOUS.,SA)
42. LPT 5 EMPLOYEES ON NONAG. PAYROLLS: WHOLESALE & RETAIL TRADE (THOUS.,SA)
43. LPFR 5 EMPLOYEES ON NONAG. PAYROLLS: FINANCE,INSUR.&REAL ESTATE (THOUS.,SA
44. LPS 5 EMPLOYEES ON NONAG. PAYROLLS: SERVICES (THOUS.,SA)
45. LPGOV 5 EMPLOYEES ON NONAG. PAYROLLS: GOVERNMENT (THOUS.,SA)
46. LPHRM 1 AVG. WEEKLY HRS. OF PRODUCTION WKRS.: MANUFACTURING (SA)
47. LPMOSA 1 AVG. WEEKLY HRS. OF PROD. WKRS.: MFG.,OVERTIME HRS. (SA)
48. PMEMP 1 NAPM EMPLOYMENT INDEX (PERCENT)

<u>Consumption, manuf. and retail sales, and housing (RTS)</u>
49. HSFR 4 HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)(THOUS.,SA
50. HSNE 4 HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.
51. HSMW 4 HOUSING STARTS:MIDWEST(THOUS.U.)S.A.
52. HSSOU 4 HOUSING STARTS:SOUTH (THOUS.U.)S.A.
53. HSWST 4 HOUSING STARTS:WEST (THOUS.U.)S.A.
54. HSBR 4 HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)
55. HMOB 4 MOBILE HOMES: MANUFACTURERS' SHIPMENTS (THOUS.OF UNITS,SAAR)
56. MSMTQ 5 MANUFACTURING & TRADE: TOTAL (MIL OF CHAINED 1992 DOLLARS)(SA)
57. MSMQ 5 MANUFACTURING & TRADE:MANUFACTURING;TOTAL(MIL OF CH. 1992 DOLLARS)(SA)
58. MSDQ 5 MANUFACTURING & TRADE:MFG; DURABLE GOODS (MIL OF CH. 1992 DOLLARS)(SA)
59. MSNQ 5 MANUFACT. & TRADE:MFG;NONDURABLE GOODS (MIL OF CHAINED 1992$)(SA)
60. WTQ 5 MERCHANT WHOLESALERS: TOTAL (MIL OF CHAINED 1992 DOLLARS)(SA)
61. WTDQ 5 MERCHANT WHOLESALERS:DURABLE GOODS TOTAL (MIL OF CH. 1992 DOLLARS)(SA)
62. WTNQ 5 MERCHANT WHOLESALERS:NONDURABLE GOODS (MIL OF CHAINED 1992 $)(SA)
63. RTQ 5 RETAIL TRADE: TOTAL (MIL OF CHAINED 1992 DOLLARS)(SA)
64. RTNQ 5 RETAIL TRADE:NONDURABLE GOODS (MIL OF 1992 DOLLARS)(SA)
65. HHSNTN 1 U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)
66. GMCQ 5 PERSONAL CONSUMPTION EXPEND (CHAINED)-TOTAL (BIL 92$,SAAR)
67. GMCDQ 5 PERSONAL CONSUMPTION EXPEND (CHAINED)-TOTAL DURABLES (BIL 92$,SAAR)
68. GMCNQ 5 PERSONAL CONSUMPTION EXPEND (CHAINED)-NONDURABLES (BIL 92$,SAAR)
69. GMCSQ 5 PERSONAL CONSUMPTION EXPEND (CHAINED)-SERVICES (BIL 92$,SAAR)
70. GMCANQ 5 PERSONAL CONS EXPEND (CHAINED)-NEW CARS (BIL 92$,SAAR)

<u>Real inventories and inventory-sales ratios (Inv)</u>
71. IVMTQ 5 MANUFACTURING & TRADE INVENTORIES:TOTAL (MIL OF CHAINED 1992)(SA)
72. IVMFGQ 5 INVENTORIES, BUSINESS, MFG (MIL OF CHAINED 1992 DOLLARS, SA)
73. IVMFDQ 5 INVENTORIES, BUSINESS DURABLES (MIL OF CHAINED 1992 DOLLARS, SA)
74. IVMFNQ 5 INVENTORIES, BUSINESS, NONDURABLES (MIL OF CHAINED 1992 DOLLARS, SA)
75. IVWRQ 5 MANUFACTURING & TRADE INV:MERCHANT WHOLESALERS (MIL OF CH. 1992 $)(SA)
76. IVRRQ 5 MANUFACTURING & TRADE INV:RETAIL TRADE (MIL OF CHAINED 1992 DOLLARS)(SA)
77. IVSRQ 2 RATIO FOR MFG & TRADE: INVENTORY/SALES (CHAINED 1992 DOLLARS, SA)
78. IVSRMQ 2 RATIO FOR MFG & TRADE:MFG;INVENTORY/SALES (87$)(S.A.)
79. IVSRWQ 2 RATIO FOR MFG & TRADE:WHOLESALER;INVENTORY/SALES(87$)(S.A.)
80. IVSRRQ 2 RATIO FOR MFG & TRADE:RETAIL TRADE;INVENTORY/SALES(87$)(S.A.)
81. PMNO 1 NAPM NEW ORDERS INDEX (PERCENT)
82. PMDEL 1 NAPM VENDOR DELIVERIES INDEX (PERCENT)
83. PMNV 1 NAPM INVENTORIES INDEX (PERCENT)
84. MOCMQ 5 NEW ORDERS (NET)-CONSUMER GOODS & MATERIALS, 1992 DOLLARS (BCI)
85. MDOQ 5 NEW ORDERS, DURABLE GOODS INDUSTRIES, 1992 DOLLARS (BCI)
86. MSONDQ 5 NEW ORDERS, NONDEFENSE CAPITAL GOODS, IN 1992 DOLLARS (BCI)
87. MO 5 MFG NEW ORDERS: ALL MANUFACTURING INDUSTRIES, TOTAL (MIL$,SA)

88. MOWU 5 MFG NEW ORDERS: MFG INDUSTRIES WITH UNFILLED ORDERS(MIL$,SA)
89. MDO 5 MFG NEW ORDERS: DURABLE GOODS INDUSTRIES, TOTAL (MIL$,SA)
90. MDUWU 5 MFG NEW ORDERS:DURABLE GOODS INDUST WITH UNFILLED ORDERS(MIL$,SA)
91. MNO 5 MFG NEW ORDERS: NONDURABLE GOODS INDUSTRIES, TOTAL (MIL$,SA)
92. MNOU 5 MFG NEW ORDERS: NONDURABLE GDS IND.WITH UNFILLED ORDERS(MIL$,SA)
93. MU 5 MFG UNFILLED ORDERS: ALL MANUFACTURING INDUSTRIES, TOTAL (MIL$,SA)
94. MDU 5 MFG UNFILLED ORDERS: DURABLE GOODS INDUSTRIES, TOTAL (MIL$,SA)
95. MNU 5 MFG UNFILLED ORDERS: NONDURABLE GOODS INDUSTRIES, TOTAL (MIL$,SA)
96. MPCON 5 CONTRACTS & ORDERS FOR PLANT & EQUIPMENT (BIL$,SA)
97. MPCONQ 5 CONTRACTS & ORDERS FOR PLANT & EQUIPMENT IN 1992 DOLLARS (BCI)

<u>Prices and Wages (PWG)</u>

98. LEHCC 6 AVG HR EARNINGS OF CONSTR WKRS: CONSTRUCTION ($,SA)
99. LEHM 6 AVG HR EARNINGS OF PROD WKRS: MANUFACTURING ($,SA)
100. PMCP 1 NAPM COMMODITY PRICES INDEX (PERCENT)
101. PWFSA 6 PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)
102. PWFCSA 6 PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)
103. PWIMSA 6 PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)
104. PWCMSA 6 PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)
105. PSM99Q 6 INDEX OF SENSITIVE MATERIALS PRICES (1990=100)(BCI-99A)
106. PUNEW 6 CPI-U: ALL ITEMS (82-84=100,SA)
107. PU83 6 CPI-U: APPAREL & UPKEEP (82-84=100,SA)
108. PU84 6 CPI-U: TRANSPORTATION (82-84=100,SA)
109. PU85 6 CPI-U: MEDICAL CARE (82-84=100,SA)
110. PUC 6 CPI-U: COMMODITIES (82-84=100,SA)
111. PUCD 6 CPI-U: DURABLES (82-84=100,SA)
112. PUS 6 CPI-U: SERVICES (82-84=100,SA)
113. PUXF 6 CPI-U: ALL ITEMS LESS FOOD (82-84=100,SA)
114. PUXHS 6 CPI-U: ALL ITEMS LESS SHELTER (82-84=100,SA)
115. PUXM 6 CPI-U: ALL ITEMS LESS MIDICAL CARE (82-84=100,SA)
116. GMDC 6 PCE,IMPL PR DEFL:PCE (1987=100)
117. GMDCD 6 PCE,IMPL PR DEFL:PCE; DURABLES (1987=100)
118. GMDCN 6 PCE,IMPL PR DEFL:PCE; NONDURABLES (1987=100)
119. GMDCS 6 PCE,IMPL PR DEFL:PCE; SERVICES (1987=100)

<u>Money and credit quantity aggregates (Mon)</u>

120. FM1 6 MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)
121. FM2 6 MONEY STOCK:M2(M1+ON RPS,ER$,G/P&B/D MMMFS&SAV&SM TM DEP(B$, SA)
122. FM3 6 MONEY STOCK: M3(M2+LG TIME DEP,TERM RP'S&INST ONLY MMMFS)(BIL$,SA)
123. FM2DQ 5 MONEY SUPPLY-M2 IN 1992 DOLLARS (BCI)
124. FMFBA 6 MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL$,SA)
125. FMRRA 6 DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL$,SA)
126. FMRNBC 6 DEPOSITORY INST RESERVES:NONBOR.+EXT CR,ADJ RES REQ CGS(MIL$,SA)
127. FCLNQ 5 COMMERCIAL & INDUSTRIAL LOANS OUSTANDING IN 1992 DOLLARS (BCI)
128. FCLBMC 1 WKLY RP LG COM'L BANKS:NET CHANGE COM'L & INDUS LOANS(BIL$,SAAR)

<u>Interest rates (Int)</u>

129. FYFF 2 INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (% PER ANNUM,NSA)
130. FYCP90 2 INTEREST RATE: 90 DAY COMMERCIAL PAPER, (AC) (% PER ANN,NSA)
131. FYGM3 2 INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(% PER ANN,NSA)
132. FYGM6 2 INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(% PER ANN,NSA)
133. FYGT1 2 INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(% PER ANN,NSA)
134. FYGT5 2 INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)
135. FYGT10 2 INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)
136. FYAAAC 2 BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)
137. FYBAAC 2 BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)

138. FYFHA	2	SECONDARY MARKET YIELDS ON FHA MORTGAGES (% PER ANNUM)
139. SFYGM3	1	Spread FYGM3 - FYFF
140. SFYGM6	1	Spread FYGM6 - FYFF
141. SFYGT1	1	Spread FYGT1 - FYFF
142. SFYGT5	1	Spread FYGT5 - FYFF
143. SFYAAAC	1	Spread FYAAAC - FYFF
144. SFYBAAC	1	Spread FYBAAC - FYFF
145. SFYFHA	1	Spread FYFHA - FYFF
146. PPSPR	1	Public-Private Spread FYCP90-FYGM3
147. TBSPR	1	Term Spread FYGT10-FYGT1

Exchange rates, stock prices and volume (ESP)

148. FSNCOM	5	NYSE COMMON STOCK PRICE INDEX: COMPOSITE (12/31/65=50)
149. FSPCOM	5	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
150. FSPIN	5	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)
151. FSPCAP	5	S&P'S COMMON STOCK PRICE INDEX: CAPITAL GOODS (1941-43=10)
152. FSPUT	5	S&P'S COMMON STOCK PRICE INDEX: UTILITIES (1941-43=10)
153. FSDXP	1	S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)
154. FSPXE	1	S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%,NSA)
155. EXRUS	5	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)
156. EXRGER	5	FOREIGN EXCHANGE RATE: GERMANY (DEUTSCHE MARK PER U.S.$)
157. EXRSW	5	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.$)
158. EXRJAN	5	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.$)
159. EXRUK	5	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
160. EXRCAN	5	FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)

Table 3.1
Distribution of Absolute Value of t-Statistics

	Percentile		
	25th	**50th**	**0.75**
Standard Normal	*0.31*	*0.67*	*1.12*
All Indicators	0.40	0.84	1.46
Indicator			
IP	0.66	1.16	1.65
PUNEW	0.28	0.63	1.06
FYFF	0.82	1.51	2.29
PMCP	0.45	1.02	1.82
LPHRM	0.70	1.22	1.93
LHU5	0.90	2.32	3.92
MOCMQ	0.32	0.66	1.12
PMDEL	0.30	0.68	1.19
MSONDQ	0.26	0.51	0.80
HSBR	0.63	1.27	2.10
FSPCOM	0.28	0.66	1.39
FM2DQ	0.54	1.12	1.79
SFYGT5	0.71	1.55	2.24
HHSNTN	0.69	1.44	2.36
PPSPR	0.64	1.39	2.21
LHUR	1.18	2.26	3.49
GMCNQ	0.40	0.75	1.31
FM1	0.34	0.80	1.33
FM2	0.30	0.64	1.11
PSM99Q	0.29	0.60	0.99
PWCMSA	0.26	0.54	0.90
MDU	0.50	0.97	1.46
IVMTQ	0.48	0.95	1.43

Notes: The table shows the 25th, 50th and 75th percentiles of the absolute values of the t-statistics for the regressors shown in the first column across 160 forecasting regressions as described in Section 3.

Table 4.1
Median of Relative MSEs

	Number of Factors			
Series Category	**BIC Choice**	**1**	**2**	**3**
Overall	0.81	0.88	0.77	0.81
Output and Income	0.67	0.95	0.73	0.72
Employment and hours	0.75	0.78	0.69	0.74
Consumption and Sales	0.90	0.98	0.85	0.88
Inventories and Orders	0.90	0.92	0.81	0.86
Prices and Wages	0.46	0.45	0.45	0.48
Money and Credit	0.63	0.63	0.63	0.64
Interest Rates	0.81	0.85	0.83	0.82
Exch. Rates and Stock Prices	0.97	0.96	0.96	0.94

Note: This table shows the median of the mean squared error of the factor models for the series in the category listed in the first column. The mean squared errors are relative to the MSE for the univariate autoregression.

Figure 3.1 t-statistic NonCentrality Parameter CDF
(Absolute Value)

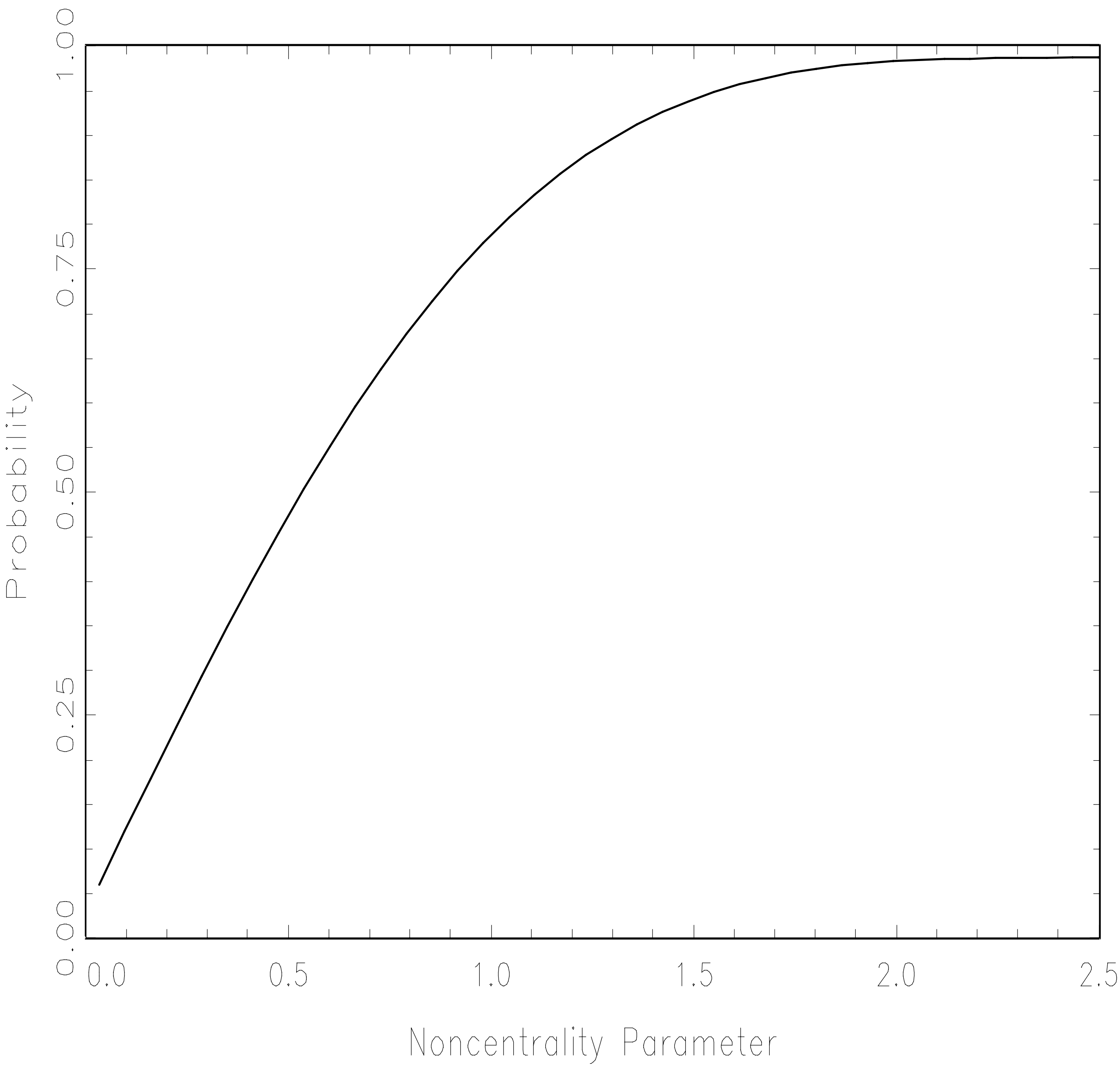

Figure 3.2 Predictive R^2

Estimated Average Value

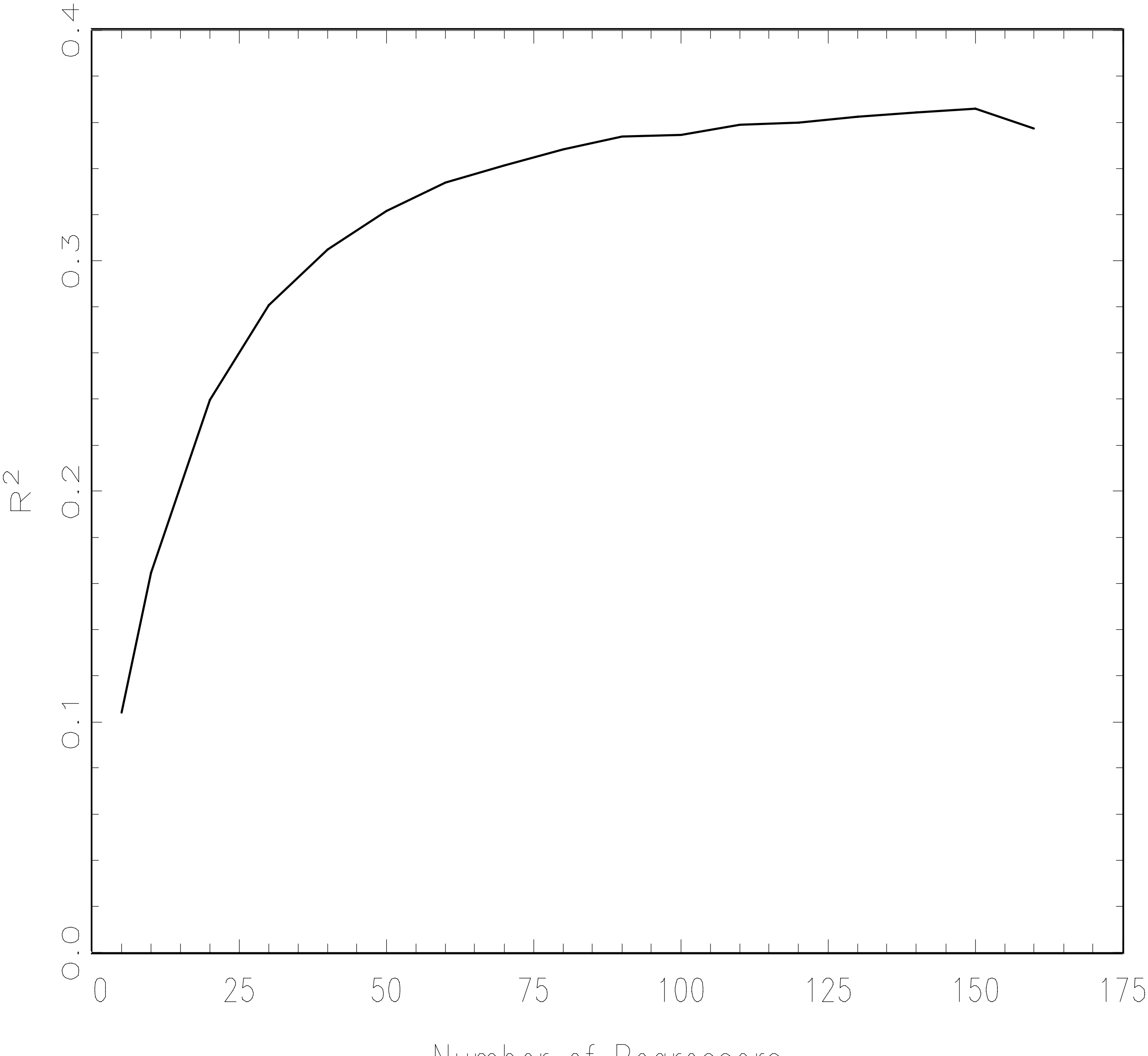

Figure 4.1 Relative MSE
Cummulative Frequency Distribution

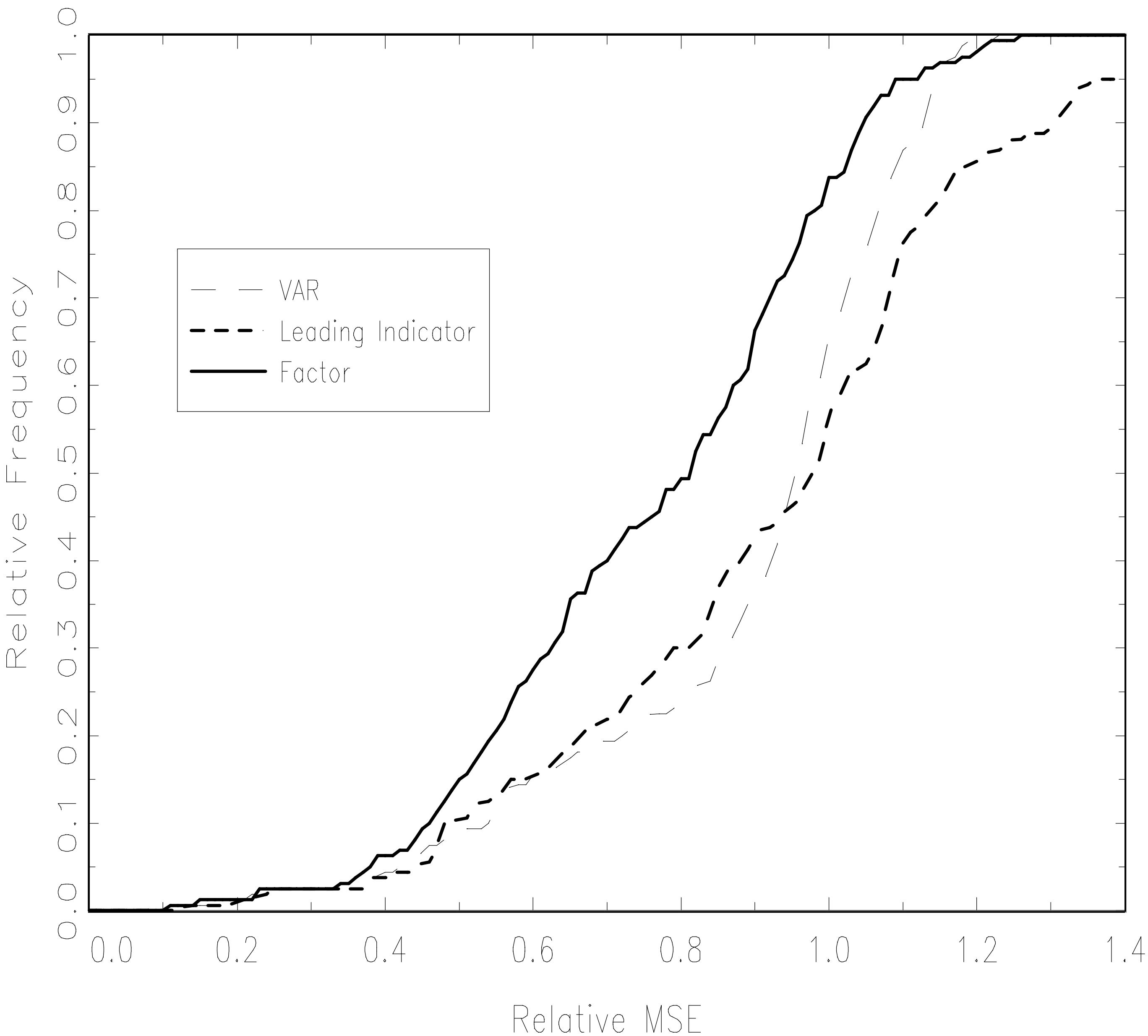

Figure 4.2 Relative MSE by Forecast Category

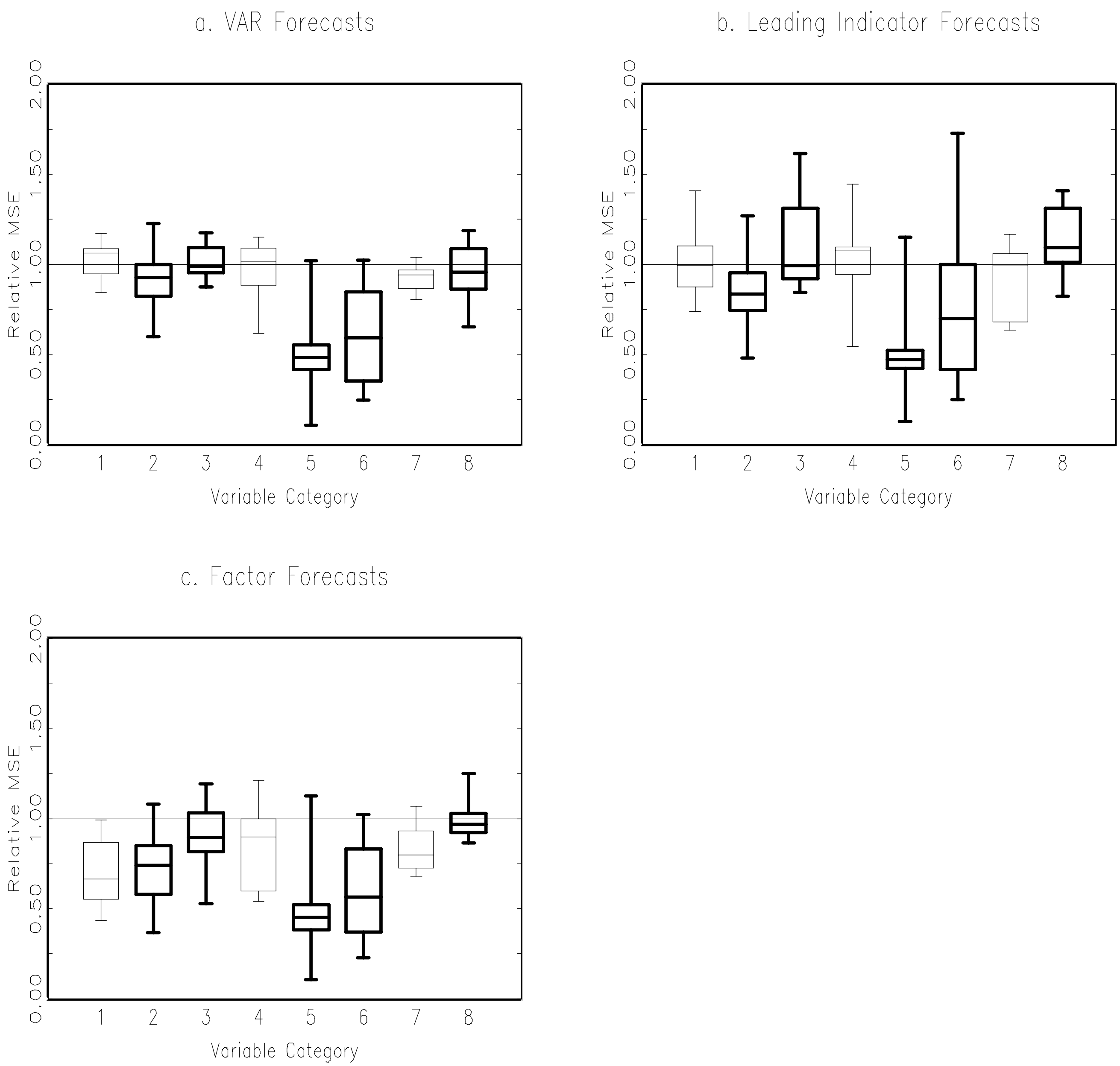

Figure 4.3
Factor 1 and Index of Capacity Utilization