OPTIMAL SURVEY AGGREGATION:

APPLICATION TO PRICES AND

ECONOMIC ACTIVITY

Joel Hasbrouck

Working Paper No. 6-82

RODNEY L. WHITE CENTER
FOR FINANCIAL RESEARCH

University of Pennsylvania
The Wharton School
Philadelphia, PA 19104

I am indebted to Stephen Figlewski, Robert Stambaugh and members of the Wharton Finance Workshop for comments on an earlier version. The manuscript was prepared by Claudia McNeill and Stephanie Hogue. The Rodney L. White Center provided partial financial support. All errors are my own responsibility.

Optimal Survey Aggregation:

Application to Prices and Economic Activity

ABSTRACT

This paper addresses econometric issues involved in using past prediction accuracy to derive optimal weights for a set of survey respondents' current forecasts of economic variables. Both univariate and multivariate forecast aggregation techniques are discussed, and a general factor-analytic technique is presented to facilitate use of the methods in situations in which the number of forecasters is large relative to the length of performance track records. Several of these techniques are applied in a rolling fashion to the Livingston surveys of consumer price index and industrial production expectations. The performance of these forecasts, as measured by average bias and root-mean-square prediction error, indicates significant improvement over the naive forecast formed as a simple average over all respondents.

I. Introduction and Summary of Results.

In the search for accurate forecasts of economic variables, the collection and aggregation of survey data has become commonplace. Usually a consensus forecast is formed by averaging the individual respondents' guesses, and in large samples where there is little or no continuity, this may be the only practical procedure. The situation frequently arises, however, in which the survey sample consists of a relatively small number of forecasters, for which a history of predictions exists. This paper addresses the econometric issues in the use of these track records to compute optimal weights for current individual predictions, and develops techniques of general applicability. These algorithms are then applied to the Livingston forecasts of the consumer price index and industrial production.

The principal prior contributions in this area are due to Figlewski (1979), who developed procedures for optimal univariate forecasting under certain individual rationality assumptions using maximum likelihood techniques, and applied the procedures to the Livingston inflation forecasts.[1] This study refines and extends Figlewski's work in several respects. First, the problem is cast in the framework of a general linear model, an approach that facilitates the derivations by permitting direct application of results familiar from regression analysis. This approach also yields results under a continuum of assumptions about individual rationality somewhat broader than that considered by Figlewski. Finally, the method easily generalizes to multivariate forecasting situations. This is of great potential practical importance since the predictions solicited of survey respondents often comprise several variables or span varying horizons. The

[1] Figlewski's techniques were applied to survey data on money supply forecasts by Figlewski and Uhrich (1981).

algorithms discussed here permit computation of optimal forecast weights based on past accuracy assessed over a multivariate prediction history.

Figlewski's original contribution was also innovative in the application of a one-factor "market" model to obtain a parsimonious representation of the interdependence of forecasters' prediction errors. In many practical situations, this necessity arises from data limitations, as when the number of forecasters exceeds the length of the forecast history. The present piece improves on Figlewski's technique through the use of a formal factor-analytic model. This generalization facilitates use of the optimizing techniques when the forecast heterogeneity exists in more than one dimension, a consideration of particular importance in the analysis of multivariate forecasts.

In applying these techniques to the Livingston forecasts of the consumer price index six-months hence, the results are quite similar to Figlewski's in that the univariate optimal forecasts tend to yield some improvement in forecasting accuracy over the naive procedure of taking the arithmetic average. It was also possible, however, to closely match the performance of the optimal forecasts by merely correcting the arithmetic average forecast by the historically-estimated additive bias. As one step in the implementation of the optimal aggregation strategies, factor analyses were performed on past prediction errors for the set of forecasters. The results revealed a high degree of homogeneity among the respondents. Even when a relatively large number of factors was employed, the first factor accounted for a large proportion of the explained variance.

Although prior researchers have concentrated on the CPI forecasts, the Livingston surveys comprise a diverse set of variables. Among these, the industrial production index forecasts constitute a particular interesting source, a real variable in contrast with the CPI, which is generally speaking monetary. When the optimal aggregation techniques were applied to the industrial production data, the results were somewhat similar to the CPI application in that an improvement was found relative to the naive averaging, but similar performance could be obtained by a simple bias-corrected version of the average. Relative to the CPI forecasts, accuracy in the industrial production predictions was somewhat poorer, and homogeneity of the predictions seemed higher. The contemporaneous correlation between price and production forecast errors was found to be insignificantly negative.

The balance of this paper is divided into sections on theory and application. In the theoretical section, the univariate and multivariate forecast techniques are discussed and the factor-analytic model for dealing with data insufficiency problems is presented. The third section describes the application of the procedures to the Livingston data on consumer prices and industrial production, and concludes with some reflections on the efficacy of the techniques.

II. The Optimal Aggregate Forecast.

The Univariate Case

Let f_t denote a row vector of n individuals' predictions about the subsequent realization of some stochastic economic variable, and denote the actual value taken on by this variable as y_t. Throughout, the time subscript will refer to the point at which predictions are made, and thus the realization associated with time t actually occurs subsequent to that point. The central problem considered here is the determination of a predictor, $y_t^*(f_t)$, a function of the individual predictions, which is in some sense an optimal forecast of y_t. To impose some structure on this very general problem, we adopt:

(A.1) At time t, the predictions and the subsequent realization, (f_t, y_t), may be considered random variables drawn from a multivariate normal population.

An alternative assumption will be presented later, but preliminary use of a normality assumption confers some desirable statistical and expository features. In particular, the predictor function may be written:

$$y_t^* = E[y_t | f_t] = \beta_0 + f_t \beta \tag{1}$$

where β_0 is an intercept and β is an (n×1) weighting vector. The joint normality assumption implies linearity, unbiasedness, and the property that y_t^* is optimal in the sense of minimizing the expected squared prediction error, $E[(y_t^* - y_t)^2]$.

Since it is unlikely that the parameters of the distribution describing (f_t, y_t) are known, we now consider the problem of estimating (1). Suppose we have at our disposal a set of forecasts and realizations (f_t, y_t) for $t = 1, \ldots, T$, with $T > n$, and furthermore

(A.2) (f_t, y_t) for $t = 1, \ldots, T$ are identically and independently distributed.

We may then write

$$y = [\underline{1} \;\; F] \begin{bmatrix} \beta_0 \\ \beta \end{bmatrix} + u \tag{2}$$

where F is the (T×n) matrix of forecasts ($F' = [f_1' \; f'_2 \; \ldots \; f_T']$); y is a (1×T) column vector of realizations ($y' = [y_1 \; y_2 \; \ldots \; y_T]'$); $\underline{1}$ is a conformable column vector of one's; and u a (1×T) column vector of disturbances. Under the above assumptions, $E[u] = \underline{0}$ and $E[uu'] = \sigma^2 I$, and the coefficients and disturbance variance of (2) may be estimated using ordinary least squares:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \underline{1}'\underline{1} & \underline{1}'F \\ F'\underline{1} & F'F \end{bmatrix}^{-1} \begin{bmatrix} \underline{1}' \\ F' \end{bmatrix} y$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T - n - 1} \tag{3}$$

where $\hat{u}$ is the estimated residual vector.

Assessing Significance

Statistical significance of this regression may be assessed by the usual methods, but in this application, one particular question predominates. Specifically, is the least-squares construction of an optimal forecast a significant improvement over simply averaging the respondents' predictions? That is, can we reject the null hypothesis that $\beta_0 = 0$ and $\beta = (1/n)\underline{1}$? In addition to this most restrictive null hypothesis, however, we may also be interested in variants that permit limited biases of a systematic nature which do not involve differential weighting of the respondents. Such biases may be classed as either additive or proportional. The former suggests $\beta_0 \neq 0$, and the latter suggests $\beta \neq (1/n)\underline{1}$, while maintaining the assumption of constant β_i for $i = 1, . . ., n$. The combinations are summarized in Figure 1; the Roman numerals denote the alternative null hypotheses. (The particular null hypothesis described above is denoted "IV.")

Proportional Bias/Additive Bias	No ($\beta_0 = 0$)	Yes ($\beta_0 \neq 0$)
No ($\beta_i = 1/n, i = 1, . . ., n$)	IV	II
Yes (β_i = constant, $i = 1, . . ., n$)	III	I

Figure 1. Alternative Null Hypotheses.

It is clear that each variant represents a (possibly restricted) linear transformation of the simple average forecast. The forecasts computed in such ways are of particular interest because they ignore any correlations between forecaster errors and so may be computed at low cost. In addition, deviations from IV may indicate the nature of aggregate forecast "irrationality."

Testing the validity of any of these null hypotheses is tantamount to testing for specification of a partial or complete regression coefficient vector. These tests are well-known (see, for example, Theil (1971)), and so a complete description will not be presented here. In any case, the validity of these tests is vitiated by the factor-analytic strategy discussed later and actually applied to the Livingston data.

Implications of Individual Rationality

To this point, no assumptions beyond stable joint normality have been utilized in describing the individual forecasts. Additional structure may be imposed on the problem by considering the implications of individual rationality and systematic deviations therefrom. The treatment will closely parallel that used in considering the various null hypotheses for the aggregate forecasts. Under the joint normality assumption, the conditional expectation of y_t is a linear function of a particular individual's forecast.

$$E[y_t | f_{it}] = b_i + c_i f_{it} \tag{4}$$

where b_i and c_i are functions of the parameters of the distribution of (f_t, y_t). In this representation, b_i is an additive bias and c_i is a multiplicative or proportional bias. Assuming c_i to be nonzero, we may take unconditional expectations of (4) and rearrange to obtain $E[f_{it}] = (m_t - b_i)/c_i$, where m_t is the unconditional mean $E[y_t]$. In matrix notation, $E[f_t] = m_t[1/c_i] - [b_i/c_i]$, where $[1/c_i]$ and $[b_i/c_i]$ are the $(1\times n)$

vectors with elements $1/c_i$ and c_i/b_i. Since the optimal aggregate forecast is unbiased,

$$m_t = E[y_t^*] = \beta_0 + (m_t[1/c_i] - [b_i/c_i])\beta \quad (5)$$

For (5) to hold, it is sufficient to require

$$[1/c_i]\beta = 1 \text{ and } \beta_0 = [b_i/c_i]\beta \quad (6)$$

Thus, restrictions placed on the individual conditional expectations lead to restrictions on the optimal aggregate forecast weights, and we turn now to a systematic discussion of these. As a starting point, denote as model I the system with no restrictions on c or b, with corresponding weights estimated by (3). This is analogous to null hypothesis I in Figure 1, in which both additive and proportional biases are permitted.

Strict rationality requires an absence of conditional bias in the following sense. The i^{th} individual is assumed fully cognizant of the joint marginal distribution (f_{it}, y_t) and therefore his expectation conditional on his own forecast must be that given in (4) with $b_i = 0$ and $c_i = 1$. The restrictions on the optimal forecast weights derived from (6) are $\beta_0 = 0$ and $\underline{1}'\beta = 1$, a zero intercept and normalization constraint. The modifications to the OLS estimates necessitated by this rationality requirement are as follows. Denote the (1×n) vector of individual forecast errors as d_t, with $f_t = y_t\underline{1}' + d_t$. Then

$$\begin{aligned} y_t = y_t^* + u_t &= \beta_0 + (y_t\underline{1}' + d_t)\beta + u_t \\ &= d_t\beta + u_t \ . \end{aligned} \quad (7)$$

Extending this to the T-period sample, define D as the (T×n) matrix with $D' = [d_1' \ d_2' \ . \ . \ . \ d_T']$. Then the best estimate of β is provided by estimating $\underline{U} = D\beta + u$ subject to the constraint that $\underline{1}'\beta = 1$. This is easily shown

(Theil (1971), p. 44) to be:

$$\hat{\beta} = (D'D)^{-1}\underline{1}[\underline{1}'(D'D)^{-1}\underline{1}]^{-1} \ . \tag{8}$$

This system, characterized by a polar assumption of total rationality, will be denoted model IV.

The two intermediate cases (models II and III) arise when we selectively permit either a multiplicative or additive bias. Under a postulate of no multiplicative bias, but (possibly) nonzero additive bias ($c = \underline{1}'$, $b \neq \underline{0}'$), the restrictions (6) imply $\underline{1}'\beta = 1$ and $\beta \neq 0$, normalization and a nonzero intercept. To clarify the modifications to the estimator, decompose the forecast deviation vector as $d_t = -b + e_t$, where e_t is a (1×n) vector of zero-mean errors. Then

$$\begin{aligned} y_t = y_t^* + u_t &= \beta_0 + (y_t\underline{1}' - b + e_t)\beta + u_t \\ &= e_t\beta + u_t \ . \end{aligned} \tag{9}$$

Extending this to the T-period sample, we may estimate $\underline{0} = E\beta + u$, where $E' = [e_1' \ e_2' \ . \ . \ . \ e_T']$, subject to the constraint that $\underline{1}'\beta = 1$. This constrained coefficient estimate is

$$\hat{\beta} = (E'E)^{-1}\underline{1}[\underline{1}'(E'E)^{-1}\underline{1}]^{-1} \ . \tag{10}$$

The intercept may then be computed as $\beta_0 = b'\beta$. This is the general case considered by Figlewski and Figlewski and Uhrich, and here it will be denoted model II.[1]

The final case follows from permitting a multiplicative bias and suppressing the additive bias ($b = \underline{0}'$, $c \neq \underline{1}'$). The implied requirement that

---

[1]Model IV may of course be viewed as a special case of model II, with zero additive biases, "bias restricted" in Figlewski's terminology.

$\beta_0 = 0$ may be met by suppressing the intercept in (3):

$$\hat{\beta} = (F'F)^{-1}F'y \ . \tag{11}$$

This will be denoted model III.

Model selection among the various alternatives may be effected by joint tests on the coefficients of regressions of actuals on predictions corresponding to (4). Alternatively, if one is only concerned with the implied final coefficient restrictions on β_0 or β, these may be tested directly. The extent of the aggregate additive bias may be ascertained by examining the t-statistic on the intercept in the coefficient vector (3), and a likelihood ratio test on the normalization restriction will disclose that of the aggregate multiplicative bias. While these tests are sufficient for choosing among the aggregate models implied by the rationality postulates, they are obviously insufficient for establishing the extent of individual rationality. Once a model has been chosen and estimated, it is then natural to evaluate it relative to one or more of the null hypotheses in Figure 1.

When the normalization restriction $\underline{1}'\beta = 1$ is present, then the optimal aggregate forecast is a weighted average of the individual forecasts with an optional constant. This weighted average interpretation is intuitively convenient, but it should be emphasized that depending on the correlations among the forecasts, the weights may vary considerably in sign and magnitude. Suppose, for example, that there are only two forecasters and that their individual forecasts are unbiased (model IV). By the application of (8),

$$\hat{\beta} = \frac{1}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \begin{bmatrix} \sigma_2^2 - \sigma_{12} \\ \sigma_1^2 - \sigma_{12} \end{bmatrix} \tag{12}$$

where the σ's are the variances and covariances of the forecast deviations for the two forecasters.

If the forecasters tend to be equally accurate, i.e., $\sigma_1 = \sigma_2$, then the optimal weights will be one half on each forecaster. If the forecasters are not equally proficient, the relative weights are also determined by the correlation between their errors. Suppose that the first is more accurate than the second, $\sigma_1 < \sigma_2$. If the correlation is zero, more weight will be placed on the first and this weight will increase as the correlation increases. If $\sigma_1^2 < \sigma_{12}$, then the weight on the second forecaster will be driven negative. In practice, negative weights are frequently obtained.

Alternative Statistical Assumptions

The assumptions (A.1) and (A.2) which mandate identical and independent multivariate normal distributions for (f_t, y_t) are convenient for readily obtaining many of the above results, but much of the development remains intact under somewhat weaker postulates. If the joint normality assumption (A.1) is dropped, then the resulting OLS predictor (3) is only optimal within the class of linear predictors and the significance tests are invalid. An intermediate assumption supporting the OLS results is, of course

(A.1') The distribution of (f_t, y_t) is such that the conditional expectation of y_t is linear: $E[y_t | f_t] = \beta_0 + f_t\beta$.

This merely transforms (1) into an assumption. In addition if the residuals $u_t \equiv y_t - \beta_0 - f_t\beta$ are i.i.d. normal, then the significance tests are valid.

The stability requirement on the distribution specified in (A.2) may also be relaxed. The OLS estimators will be efficient as long as the linear relationship specified in (1) is stable. If any of the coefficient restrictions derived from (4) are imposed, then (4) must be assumed stable as well.

In applying OLS estimation to the predictions and realizations, we are relying on the assumption that the residuals u_t for $t = 1, \ldots, T$ are identically and independently distributed: $E[uu'] = \sigma^2 I$. This assumption may be deemed too restrictive in many practical applications. It may be the case that the residuals are heteroscedastic, reflecting perhaps varying levels of prediction uncertainty over the sample period. Alternatively, serial correlation may be present in the forecast residuals.

Concerning this last point, it should be noted that residual autocorrelation need not necessarily vitiate forecast rationality for the simple reason that past residuals may not be observable. This case arises in the frequently encountered instance of long-range forecasts being made at relatively brief prediction intervals. It will be recalled that time subscript in the present notation refers to the time at which the predictions are made. Suppose predictions are made at monthly intervals of the level of the consumer price index ten years hence. The June 1982 and July 1982 forecast residuals will both reflect, except under certain unrealistic assumptions concerning the stochastic behavior of the CPI, events subsequent to the latter date, and hence may exhibit autocorrelation. The ten-year residual associated with the June 1982 forecast is not, however, known as of July.

When the residuals are not i.i.d., estimation may proceed using generalized least squares, either constrained or unconstrained as warranted by the individual rationality postulates employed. Let $E[uu'] = \Omega$ be the residual covariance matrix. Under rationality postulate II, for example (possible additive bias, no multiplicative bias), an efficient estimate of β analogous to that presented in (10) is given by the constrained GLS estimator

$$\hat{\beta} = (E'\Omega^{-1}E)^{-1}\underline{1}[\underline{1}'(E'\Omega^{-1}E)^{-1}\underline{1}]^{-1} . \qquad (13)$$

In practical estimation situations, the actual covariance matrix Ω will not be known, of course, and must be estimated. Estimation in the presence of autocorrelated residuals is a familiar problem, and procedures are discussed in most texts. Less frequently encountered, but still adequately treated in the literature are techniques for specification and estimation in the presence of heteroscedasticity. In this application, the dispersion of respondents' predictions may be a useful proxy for the relative residual variance.

Multivariate Forecasts

Very often the predictions we wish to aggregate cover more than one economic variable. In the Livingston surveys, for example, predictions are solicited for some thirteen variables. Furthermore, predictions are made over a variety of horizons, placing the effective total number of variables predicted somewhat in excess of thirty. In this part, we discuss ways in which multivariate predictions may meaningfully be aggregated.

For simplicity, the discussion will be confined to the two variable case. Analogous to the notation in the first part, denote by f_{jt} and y_{jt} the $(1 \times n)$ prediction vector and subsequent realization of the variable for $j = 1, 2$. Corresponding to assumption (A.1) we have

> (A.1'). The set of predictions and realizations $(y_{1t}, f_{1t}, y_{2t}, f_{2t})$ may be considered random variables drawn from a multivariate normal population.

As before, the joint normality assumption is sufficient to establish the optimality of the linear predictor given by

$$\begin{aligned} y^*_{jt} &= E[y_{jt} | f_{1t}, f_{2t}] \\ &= \beta_{j0} + f_{1t}\beta_{j1} + f_{2t}\beta_{j2} \qquad \text{for } j = 1, 2 \end{aligned} \tag{14}$$

where β_{j1} and β_{j2} are $(n \times 1)$ weighting vectors. Note that this most general formulation permits an individual's forecast of both variables to convey information concerning the likely realization of either one.

If a stability assumption corresponding to (A.2) is made, and if a history of past forecasts and realizations is available, (13) may be estimated by least squares. Suppose we have a set of forecasts and realizations, (y_{1t}, f_{1t}, f_{2t}) for $t = 1, \ldots T_1$, and y_{2t} for $t = 1, \ldots, T_2$, where we order the variables so that $T_1 > T_2$. The inequality in length of the histories may arise when the two variables in question have different horizons. In attempting to construct an optimal joint forecast of the CPI 6 and 12 months ahead, for example, we will typically have at our disposal one more observation for the 6-month forecast/realization pairs than we have for the 12-month. The system may then be described:

$$y_j = [\underline{1} \quad F_1 \quad F_2] \begin{bmatrix} \beta_{j0} \\ \beta_{j1} \\ \beta_{j2} \end{bmatrix} + u_j \text{ for } j = 1, 2 \tag{15}$$

where y_j is the $(1 \times T_j)$ vector of actual values; u_j is the $(1 \times T_j)$ vector of residuals; F_1 and F_2 are the $(T_j \times n)$ forecast matrices.

The estimation procedure depends on assumptions made concerning the residuals and individual rationality. If $\Omega \equiv E\left[\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} [u_1' \quad u_2']\right]$ is block diagonal, i.e. the residuals are uncorrelated across variables, single equation OLS or GLS estimation will suffice. Frequently, however, such independence is a poor assumption. Forecast errors in inflation and economic activity would ordinarily be assumed dependent, for example, as would errors in predictions over six and twelve month horizons. In such cases, joint GLS procedures (such as a seemingly unrelated regression model) will provide the most efficient parameter estimates.

As in the univariate case, additional structure may be imposed on the problem by appealing to varying degrees of individual rationality. Corresponding to (4) we may write the i^{th} individual's conditional expectation as a linear function of his predictions:

$$E[y_{jt} | f_{1it}, f_{2it}] = b_{ji} + c_{1ji} f_{1it} + c_{2ji} f_{2it} \text{ for } j = 1, 2 . \qquad (16)$$

Now in addition to the additive bias and multiplicative bias in the "own" forecast, there is also the possibility of biases resulting from the forecast of the other variable. For example, an individual's prediction of the CPI twelve months hence might conceivably contain information relevant to the best prediction of the CPI six months hence that is not reflected in that individual's six-month forecast. Such a condition might arise in a situation in which a forecaster was ignorant of the true structure of the underlying economic system, but had access to superior information.

If a restrictive rationality postulate is deemed appropriate, implications may easily be derived for restrictive conditions on the final coefficient estimates, corresponding to those conditions present in the univariate analysis. In the multivariate case, however, it may also be desirable to impose restrictions relating the coefficients (weights) for different forecast variables. For example, if forecasting performance is essentially homogeneous across variables (but not forecasters), it may make sense to require $\beta_{12} = \beta_{21} = 0$ and $\beta_{11} = \beta_{22}$ in (14) and (15). With two variables, exhaustive consideration of all the permutations of the various rationality postulates leads to a large number of models. Generation of these models is quite straightforward, however, and little additional insight may be gleaned from individual consideration of these diverse cases.

Factor-Analytic Approaches to Data Limitation

The main obstacle to the direct implementation of the above forecasting procedures is likely to be data availability, a situation in which there are more forecasters than periods over which to evaluate performance ($T < n$). In this case $(F'F)$ will not be of full rank, rendering indeterminant the β coefficients in (3) or (11). Similar problems exist for $(D'D)$ and $(E'E)$ in (8) and (10). Certainly one way of handling this problem is to arbitrarily discard forecasters, but this may result in the discarding of information as well. A better procedure is found in the application of factor analysis.

For simplicity the discussion will initially focus on the univariate situations in which the forecast residuals may be assumed i.i.d.. A simple factoral approach to this problem was first utilized by Figlewski. Briefly, Figlewski suggested the following model for an individual forecast deviation:

$$d_{it} = a_i + b_i z_t + e_{it}$$

where a_i and b_i are forecaster-specific bias and co-movement weights and e_{it} is a residual disturbance. The single factor, z, was taken by Figlewski as the mean deviation over all forecasters. This is an intuitively reasonable procedure, but it suffers from two defects. First, the selection of the single factor used is imprecise: the mean deviation is not likely to be the best single factor available. Second, it is difficult to envision a second factor which might reasonably be employed if this model were to be generalized.

These defects motivate the utilization of a formal multifactor model. The $(1\times n)$ forecast vector f_t might be expressed as

$$f_t' = \mu + \Lambda + v_t \tag{17}$$

where μ is an (n×1) mean vector, z_t is a (p×1) vector of orthogonal standard normal variates, Λ is an (n×p) matrix of factor loadings and v_t is a vector of mutually uncorrelated residual disturbances. Analogous factor structures may be used for d_t, or (with $\mu = 0$), for e_t.

Using this model to represent the (T×n) data matrix F, we have

$$F' = \mu\underline{1}' + \Lambda Z + V \tag{18}$$

where $Z = [z_1 \; z_2 \; . \; . \; . \; z_T]$ and $V = [v_1 \; v_2 \; . \; . \; . \; v_T]$. The expected value for the required cross-product matrix may easily be computed as

$$E\left[F'F\right] = T\left[\mu\mu' + \Lambda\Lambda' + \Psi\right] \tag{19}$$

where Ψ is the diagonal matrix of specific variances $E[vv']$. This may then be substituted into (3) to obtain estimates of the optimal weights. Since $\Lambda\Lambda'$ is invariant to orthogonal rotations of the factor model, the estimated β will be unique.

There are a number of approaches to the factor-estimation problem. Most preferable on theoretical grounds are iterative techniques that provide maximum likelihood estimates of Ψ and Λ. For reasons of computational economy, however, an approximate solution method called the principal factor technique was used. The sample covariance matrix may be expressed as sRs where R is the (n×n) sample correlation matrix and s is the (n×n) diagonal matrix of sample standard deviations. In the principal factor method, R is factored as $\left(\Gamma\sqrt{\delta}\right)\left(\Gamma\sqrt{\delta}\right)'$ where δ is the diagonal matrix of eigenvalues and Γ is the matrix of eigenvectors. Denote as Q the (n×p) submatrix of $\left(\Gamma\sqrt{\delta}\right)$ formed by retaining those columns corresponding to the p largest eigenvalues. The loading matrix Λ is then estimated as sQ and the matrix of specific variances Ψ is estimated as $s(I - QQ')s$. The Kaiser-Gutmann criterion was applied to determine the appropriate number of factors: a factor was retained if the corresponding eigenvalue was greater than unity. Note, however, that the rank of the sample covariance matrix in this problem is generally equal to the number of time periods employed. An (n×n) positive semidefinite matrix of factors obtained should be judged relative to the number of time periods used, not the number of forecasters, which will generally be somewhat larger.[1] It should also be noted that the validity of the Kaiser-Gutman criterion is asymptotic and the small sample properties have not been established.

The factor analytic solution to the data limitation problem involves two primary drawbacks. First, a factor model may necessitate stronger stability requirements. The OLS procedure in (3) for example will yield a linearly optimal predictor as long as $E[y_t|f_t]$ is a stable linear function of f_t. Stability of the distribution of f_t is not necessary here, but it is an obvious assumption of the factor model (17). Second, while the Kaiser-Gutmann criterion may be an effective guideline for judging the adequacy of a given factor model, determination of the appropriate number of degrees of freedom to use in computing the variance estimate in (3) is unclear. This variance estimate is needed in computing the F-statistics for testing the null hypotheses previously described, and there appear to be no obvious alternative procedures.

The factor analysis strategy may be extended to univariate and multivariate situations involving GLS estimation, such as (14), but with an additional complication. It is clear that an attempt to construct a product of the form $(E'\Omega^{-1}E)$ (as required in (13) for example) from a factoral model

---

[1]A discussion of the principal factor solution to the factor analysis problem is given in Press (1972), and a discussion of the Kaiser-Gutmann criterion is provided in Lindeman, Merenda and Gold (1980). The IMSL routine OFPRI was used for the computation.

for the residuals E will result in a multiplicity of results: while the constructed product (E'E) is invariant to factor rotation, the product $(E'\Omega^{-1}E)$ is not. The solution to this dilemma lies in transforming the data prior to factor analysis. Factor the error covariance matrix $\Omega^{-1} = \omega'\omega$, and perform factor analysis on the product $(E'\omega')$. The reconstructed product $(E'\omega')(\omega E)$ will be invariant to factor rotation, and estimation may proceed.[1]

III. Application to the Livingston Data.

This section describes the use and evaluation of the optimal aggregation techniques applied to the Livingston survey data. Semiannually since 1946, Joseph Livingston of the Philadelphia Inquirer has polled a group of professional economists soliciting their predictions of key economic variables. Over the years, some 250 economists have been surveyed, but the number of respondents to any given survey is generally about fifty. Forecasts of twenty-two variables have at one time or another been solicited, and the forecast horizons are generally six and twelve months ahead. Later surveys include other horizons, but these are not considered here.[2]

Among the survey variables, the most extensively studied have been the consumer price index (CPI) predictions, usually with a view toward ascertaining the rationality of the implied inflation forecasts.[3] In

[1]Another problem in applying the factor analytic procedure involves missing data. This condition may arise in a multivariate case when the forecast variables span different horizons. An algorithm for factor model estimation with missing data has been suggested by Dempster, Laird and Rubin (1977). Figlewski has suggested applying this procedure to the univariate case when the missing data is due to gaps in a forecaster's track record.

[2]The Livingston data were kindly provided by the Federal Reserve Bank of Philadelphia.

[3]In addition to Figlewski (1979), see Carlson (1977), Wachtel (1977), Mulineaux (1978) and (1980), and Pearce (1980).

addition, the CPI forecasts were studied by Figlewski in his initial treatment of the optimal forecast problem. This prior attention makes the CPI a natural first choice in testing the optimal aggregation techniques. In contrast to the bulk of the earlier work, however, the immediate concern is whether or not the optimal aggregation procedures seem to produce superior forecasts. Ascertaining the rationality of these forecasts lies beyond the scope of the present work.

In addition to the CPI forecasts, the techniques are also applied to the Livingston forecasts of industrial production. Several considerations strongly favor this course. First, since a primary purpose of this paper is the exposition of general techniques to enhance forecast accuracy, application of the methods to at least one additional variable seems desirable. Second, with the first variable belonging primarily to the monetary sector, considerations of diversity suggest selection of a real variable as an alternative. The interrelationships between forecasts of real and monetary variables are of some interest in their own right, and furthermore, prices and industrial production constitute a sensible pair of variables with which to investigate the potential efficacy of the bivariate optimal aggregation methods.

The choice of industrial production for the second variable is in some respects, however, less than ideal. As a measure of economic activity, it excludes production of services, a component which is large and only partially correlated with other measures. In addition, the bulk of macroeconomic literature, particularly that addressing the linkages between real and monetary sectors, has tended to use the unemployment rate as a convenient summary measure of real activity. Although an unemployment rate forecast is now present on the Livingston surveys, it has only been requested since 1960, and a similar problem exists for the forecasts of constant-dollar GNP, which were first requested in 1971. In contrast, forecasts of industrial production have been present on the survey from its inception.

Having selected the consumer price and industrial production indices as the variables of interest, it is then necessary to address the question of what transformations, if any, are to be used. The forecast variables actually requested were either the level of the CPI or industrial production (the latter formally being a rate of production), and to these responses, two types of transformations were applied. The first transformation is minimal in nature, preserving the level character of the forecasts and made for reasons of computational convenience. This involved rebasing all the forecasts to a common base year and taking the natural logarithm. The last step was taken to minimize distortions caused by differences in scale of the series over the sample period.

The second transformation was a rate-of-change computation similar to that usually employed in the Livingston inflation studies. At the time the forecast was formed, a base value from two months earlier was assumed to be available.[1] The logarithm of the ratio of the predicted value to the base value was taken, and this result multiplied by (8/6) to place the resulting growth rate on a six-month basis. Values for the series subsequent to the second prior month may, of course, have been available to the respondents and this uncertainty may introduce error.[2]

---

[1]The base values were taken in most cases from the Survey of Current Business issue in the month of the forecast, and hence are preliminary figures. The base values so obtained are quite close to those actually given by Livingston on the questionnaires.

[2]See Carlson (1977).

The sample period over which forecasts were computed and evaluated was June, 1952 through December, 1979. While Livingston started the surveys in 1946, the early periods are consumed in the estimation process. In evaluating a forecast, accuracy was assessed relative to the logarithm of the actual revised series, or the rate of change in the actual revised series, depending on which transformation was used.[1] Forecasts were analyzed by comparing the predictions to the actual values and assessing mean bias, standard deviation of the forecast errors, and root-mean-square forecasting error (RMSFE), defined as the square root of the average squared forecast deviations over the sample period.[2] These statistics are reported for the log-level forecasts of consumer prices and industrial production in Table 1, and for the rate of change forecasts in Table 2.

Line 1 of the top panels gives performance statistics for the forecast formed as the simple arithmetic average of all respondents to a given survey. Over the sample period, the forecasts are characterized by a downward bias, which with 56 observations appears to be statistically significant. This feature, noted by other researchers already cited, is a reflection of persistent underestimation of the accelerating rate of price increase commencing in the late 1960's, and constitutes prima facie evidence of the "irrationality" of the forecasts.

---

[1] It should be noted that the cumulative revisions in the CPI and industrial production index series have in some cases persistently shifted the time paths of these variables to one side of the path prior to revision. These revisions do not materially affect the rate-of-change forecasts, but the resulting discrepancies in the log-level forecasts may constitute a troublesome source of error.

[2] An alternative forecast evaluation procedure not used here involves computation of the inequality coefficients proposed by Theil (1966). This method, however, is predicated on the assumption that the predictions and realizations are drawn from a stable distribution, which is probably not the case with the price and production levels considered here.

With the exception of the simple average forecast, all of the techniques discussed in this paper require determination of a performance history. The length of this history was set at five years: forecasters were included who had responded to the current survey and the nine previous ones. This determination was made in order to strike a balance between a desire for a long track record and availability of forecasters who had such histories. This criterion resulted in an initial sample (June, 1952) of six forecasters. The size of the sample grew quite rapidly after that, however, reaching twenty in December, 1956 and generally remaining at that number or above through the end of the 1960's. In the beginning of the 1970's, the sample size began to wane, perhaps in consequence of an increase in economic uncertainty and an attendant reluctance to make predictions. In the June 1974 survey, although there were over fifty respondents, only six of these had responded continuously for the requisite five years. This constitutes the low point: by June, 1976, the sample increased to fifteen and stayed at that level or above through the rest of the decade.

Again with the exception of the arithmetic average, all of the forecasting techniques were implemented in a rolling fashion. For the June, 1960 forecast, for example, estimations were performed using the (nine) forecasts and realizations from the December, 1955 through December, 1959 surveys. The objection may be raised that this information was not fully available to an analyst aggregating the forecasts in June, 1960: the last observation includes the actual revised value for June, 1960, which would not have been available for several months afterward. In a follow-up analysis, however, suppressing this point from the data set did not materially affect the results.

Line 2, the top panels of Tables 1 and 2, gives performance statistics on the forecast computed as the average of all those respondents with the necessary 5-year history. The results are essentially identical to those obtained when the averaging was over all of the respondents to a given survey. Line 3 reports performance statistics when the forecast was formed in conformance with the null hypothesis H_0^{II}: aggregate additive bias was estimated by averaging over the nine periods prior to the forecast and this bias was then subtracted from the average forecast used in line 2. The results indicate a dramatic reduction in bias and mean square forecasting error.

The optimal linear models, it will be recalled from the previous section, resulted from a two-step procedure of factor analysis followed by a least-squares procedure. It was indicated that the factor model could only be applied to variates with stable distributions, and this requirement is probably not met by the forecast levels (F) used in models I and III. Attempts to fit these models using the factor analysis technique yielded wildly unstable final forecasts and were therefore judged unsatisfactory. Models II and IV, on the other hand, are estimated from the forecast deviations, the distribution of which may be stable even when that of the levels is not.

Due to space limitations, the complete results of the factor analysis step are not presented, but a few general characteristics of the procedure will be discussed. A condensed description of the factor analysis results from the estimation of Model II for the log-level forecasts is presented in Table 3. For the price forecasts, even when a relatively large number of factors was used, the vast bulk of total variance was explained using the first two. This suggests a high degree of uniformity in the information available to the forecasters. It is reemphasized that the number of factors should be judged relative to the number of time periods employed (nine), due to rank considerations discussed in the last section.

Using the factor-analytic regression strategies, forecasts were computed using the univariate models II and IV, and performance statistics are presented in lines 4 and 5. Relative to the simple arithmetic average, a large improvement in the bias is visible, even in model IV which restricts the additive correction to zero. Mean square prediction errors are slightly improved. Nevertheless, these forecasts seem to be dominated by the bias-corrected average (line 3).

The performance patterns in the univariate CPI forecasts is in most respects similar to that found by Figlewski in his analysis of the inflation forecasts. Figlewski estimated forecasts corresponding to the average, bias-corrected average, model II and model IV, and found reductions in RMS forecasting error generally greater than those reported here and reductions in bias that were generally less. Figlewski's estimation procedure employed the single factor model previously indicated and a slightly different selection procedure, however, and this probably accounts for the difference in results.

Performance of the industrial production forecasts generally follows the pattern found in the CPI forecasts, although the size of the forecast errors tends to be much larger. As indicated by the factor analysis results from the model II estimation, there appears to be greater homogeneity in the forecast errors than was encountered in the CPI analysis: fewer factors were generally needed and the first factor tended to explain a very high proportion of the variance. As with the CPI forecasts, optimal forecasts from models II and IV yielded biases smaller than that of the average forecast, but the bias-corrected average forecast was the best performer.

As a precursor to attempting bivariate estimation of optimal forecasts based on the six-month forecasts of industrial production and consumer prices, the correlation between forecast errors in these two variables was examined. Irrespective of what forecast model was used, the forecast errors between prices and production were negatively, but insignificantly correlated. Thus, the implementation of a joint forecasting procedure was not deemed worthwhile.[1]

Reflections on the Efficacy of the Procedures.

While implementation of the optimal aggregation methods seems to yield improvement in forecasting performance relative to naive averaging, a simple correction using historically-estimated bias (and ignoring differential weighting) does about as well. This may be a consequence of an estimation period (nine observations) too brief to yield reliable estimates, or it may be the case that predictive ability is not sufficiently stable to ensure the validity of the procedure. Indeed, given the high degree of homogeneity suggested by the intermediate factor analyses, the very existence of useful forecast relationships in the Livingston sample must be questioned. The defects, however, characterize the data and not the statistical methods

---

[1]Pairwise correlations were also computed between price and production forecasts across forecasters in a given survey. Throughout the 1950's and 1960's, these cross-sectional correlations were generally significantly positive: if a forecaster predicted higher-than-average for the CPI, he was also likely to have made a higher-than-average industrial production prediction. This tendency broke down in the 1970's, however. In that decade, a wide range of correlations was observed, including several significantly negative ones.

The economic significance of this correlation is unclear. An individual's forecast of a variable may be thought of as consisting of his guess at the "consensus" prediction, plus a residual which is specific to his forecast alone. Thus, pairwise correlations between individual forecasts reflect complex relationships between individual estimates of consensus forecasts and individual-specific components.

employed, the value of which will be determinable only as these methods are applied to other bodies of survey data.

References

1. John Carlson, "A Study of Price Forecasts," Annals of Economic and Social Measurement 6, (1977), pp. 27-56.

2. Laird, Rubin and Dempster, Journal of the Royal Statistical Society, 1977.

3. Stephen Figlewski, "Optimal Price Forecasting Using Survey Data," Solomon Brothers Center Working Paper No. 167, 1979, revised 1982, forthcoming in Review of Economics and Statistics.

4. Stephen Figlewski and Thomas Urich, "Optimal Aggregation of Money Supply Forecasts: Accuracy, Profitability and Market Efficiency," mimeo, New York University, 1981.

5. Richard H. Lindeman, Peter F. Merenda and Ruth Z. Gold, Introduction to Multivariate Statistics (Scott, Foresman and Co., Glenview, Illinois), 1980.

6. Donald Mullineaux, "On Testing for Rationality: Another Look at the Livingston Price Data," Journal of Political Economy, April 1978.

7. __________, "Inflation Expectations and Money Growth in the United States," American Economic Review, March, 1980.

8. Douglas Pearce, "Comparing Survey and Rational Measures of Expected Inflation," JMCB, 1979, pp. 447-56.

9. S. James Press, Applied Multivariate Analysis, (Holt, Rinehart and Wilson, New York), 1972, pp. 318-319.

10. Henri Theil, Applied Economic Forecasting (North Holland, Amsterdam), 1966.

11. Henri Theil, Principles of Econometrics, (John Wiley and Sons, New York), 1971.

12. Paul Wachtel, "Survey Measures of Expected Inflation and Their Potential Usefulness," in Joel Popkin, ed., Analysis of Inflation: 1965-1974 (National Bureau of Economic Analysis), 1977, pp. 361-395.

Table 1.

Log Consumer Price Index Forecasts.

No.	Description	Bias	St. Dev.	RMSFE
1.	Average (H_0^{IV}, all respondents).	.009836	.010560	.014362
2.	Average (H_0^{IV}, respondents with 5-year history).	.010191	.010721	.014723
3.	Average corrected for additive bias (H_0^{II}).	-.000217	.011966	.011860
4.	Model II.	-.000796	.014047	.013944
5.	Model IV.	.002196	.013818	.013869

Log Industrial Production Forecasts.

No.	Description	Bias	St. Dev.	RMSFE
1.	Average (H_0^{IV}, all respondents).	.018351	.041358	.044908
2.	Average (H_0^{IV}, respondents with 5-year history).	.020316	.042853	.047078
3.	Average corrected for additive bias (H_0^{II}).	.000622	.043529	.043143
4.	Model II.	.004826	.050571	.050349
5.	Model IV.	.008148	.050856	.051055

Table gives bias, root-mean-square forecasting error and standard deviation for aggregate forecasts from 1952 I (June) to 1979 II (December) for a total of 56 points. Details of calculations are given in the text.

Table 2.

Rate of Change Consumer Price Index Forecasts.

No.	Description	Bias	St. Dev.	RMSFE
1.	Average (H_0^{IV}, all respondents).	.007328	.007952	.010761
2.	Average (H_0^{IV}, respondents with 5-year history).	.007954	.008064	.011024
3.	Average corrected for additive bias (H_0^{II}).	.000105	.008710	.008633
4.	Model II.	-.000818	.010545	.010482
5.	Model IV.	.001734	.010034	.010094

Rate of Change Industrial Production Forecasts.

No.	Description	Bias	St. Dev.	RMSFE
1.	Average (H_0^{IV}, all respondents.	.005667	.029201	.029488
2.	Average (H_0^{IV}, resondents with 5-year history).	.007141	.030683	.031235
3.	Average corrected for additive bias (H_0^{II}).	-.000780	.031138	.030869
4.	Model II.	.001716	.037154	.036861
5.	Model IV.	-.000497	.037165	.036835

See notes to Table 1.

Table 3.

Factor Analysis Results.

Consumer Price Index:

Number of Factors	Number of Periods	Proportion of Variance Explained					
		1	2	3	4	5	6
1	4	0.82					
2	12	0.73	0.11				
3	12	0.66	0.13	0.08			
4	12	0.63	0.13	0.09	0.06		
5	15	0.59	0.14	0.09	0.07	0.05	
6	1	0.51	0.15	0.11	0.07	0.07	0.05

Industrial Production:

Number of Factors	Number of Periods	Proportion of Variance Explained			
		1	2	3	4
1	26	0.87			
2	17	0.84	0.07		
3	8	0.77	0.10	0.05	
4	5	0.67	0.14	0.07	0.05

Appropriate number of factors for a given period was decided by the Kaiser-Gutmann criterion. Table entries represent proportion of total (diagonal) variance explained by factors, averaged over all periods using the specified number of factors.