Chapter 12

Model-Based Forecast Combination

In forecast accuracy comparison, we ask which forecast is best with respect to a particular loss function. Such "horse races" arise constantly in practical work. Regardless of whether one forecast is significantly better than the others, however, the question arises as to whether competing forecasts may be fruitfully combined to produce a composite forecast superior to all the original forecasts. Thus, forecast combination, although obviously related to forecast accuracy comparison, is logically distinct and of independent interest. We start with what one might call "model-based" forecast combination, and then we proceed to "survey-based" combination and "market-based" combination (financial markets, prediction markets, ...).

12.1   Forecast Encompassing

Whether there are gains from forecast combination turns out to be fundamentally linked to the notion of forecast encompassing, with which we now begin. We use forecast encompassing tests to determine whether one forecast incorporates (or encompasses) all the relevant information in competing forecasts. If one forecast incorporates all the relevant information, nothing can be gained by combining forecasts. For simplicity, let's focus on the case

of two forecasts, $y_{a,t+h,t}$ and $y_{b,t+h,t}$. Consider the regression

$$y_{t+h} = \beta_a y_{a,t+h,t} + \beta_b y_{b,t+h,t} + \varepsilon_{t+h,t}.$$

If $(\beta_a, \beta_b) = (1, 0)$, we'll say that model a forecast-encompasses model b, and if $(\beta_a, \beta_b = (0, 1)$, we'll say that model b forecast-encompasses model a. For other (β_a, β_b) values, neither model encompasses the other, and both forecasts contain useful information about y_{t+h}. In covariance stationary environments, encompassing hypotheses can be tested using standard methods.[1] If neither forecast encompasses the other, forecast combination is potentially desirable.

We envision an ongoing, iterative process of model selection and estimation, forecasting, and forecast evaluation. What is the role of forecast combination in that paradigm? In a world in which information sets can be instantaneously and costlessly combined, there is no role; it is always optimal to combine information sets rather than forecasts. That is, if no model forecast-encompasses the others, we might hope to eventually figure out what's gone wrong, learn from our mistakes, and come up with a model based on a combined information set that *does* forecast-encompass the others. But in the short run – particularly when deadlines must be met and timely forecasts produced – pooling of information sets is typically either impossible or prohibitively costly. This simple insight motivates the pragmatic idea of forecast combination, in which forecasts rather than models are the basic object of analysis, due to an assumed inability to combine information sets. Thus, forecast combination can be viewed as a key link between the short-run, real-time forecast production process, and the longer-run, ongoing process of model development.

---

[1]Note that $\varepsilon_{t+h,t}$ may be serially correlated, particularly if $h > 1$, and any such serial correlation should be accounted for.

12.2 Model-Based Combined Forecasts I: Variance-Covariance Forecast Combination

In forecast accuracy comparison, we ask which forecast is best with respect to a particular loss function. Such "horse races" arise constantly in practical work. Regardless of whether one forecast is significantly better than the others, however, the question arises as to whether competing forecasts may be fruitfully combined to produce a composite forecast superior to all the original forecasts. Thus, forecast combination, although obviously related to forecast accuracy comparison, is logically distinct and of independent interest.

Failure of each model's forecasts to encompass other model's forecasts indicates that both models are misspecified, and that there may be gains from combining the forecasts. It should come as no surprise that such situations are typical in practice, because forecasting models are *likely* to be misspecified – they are intentional abstractions of a much more complex reality. Many combining methods have been proposed, and they fall roughly into two groups, "variance-covariance" methods and "regression" methods. As we'll see, the variance-covariance forecast combination method is in fact a special case of the regression-based forecast combination method, so there's really only one method. However, for historical reasons – and more importantly, to build valuable intuition – it's important to understand the variance-covariance forecast combination, so let's begin with it.

12.2.1 Bivariate Case

Suppose we have two unbiased forecasts. First assume that the errors in y_a and y_b are uncorrelated. Consider the convex combination

$$y_C = \lambda \, y_a + (1 - \lambda) \, y_b,$$

where $\lambda \in [0, 1]$.[2] Then the associated errors follow the same weighting,

$$e_C = \lambda e_a + (1 - \lambda)e_b,$$

where $e_C = \mathrm{y} - y_C$, $e_a = \mathrm{y} - y_a$ and $e_b = \mathrm{y} - y_b$. Assume that both y_a and y_b are unbiased for y, in which case y_C is also unbiased, because the combining weights sum to unity.

Given the unbiasedness assumption, the minimum-MSE combining weights are just the minimum-variance weights. Immediately, using the assumed zero correlation between the errors,

$$\sigma_C^2 = \lambda^2 \sigma_a^2 + (1 - \lambda)^2 \sigma_b^2, \tag{12.1}$$

where $\sigma_C^2 = var(e_C)$, $\sigma_a^2 = var(e_a)$ and $\sigma_b^2 = var(e_b)$. Minimization with respect to λ yields the optimal combining weight,

$$\lambda^* = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_a^2} = \frac{1}{1 + \phi^2}, \tag{12.2}$$

where $\phi = \sigma_a/\sigma_b$.

As σ_a^2 approaches 0, forecast a becomes progressively more accurate. The formula for λ^* indicates that as σ_a^2 approaches 0, λ^* approaches 1, so that all weight is put on forecast a, which is desirable. Similarly, as σ_b^2 approaches 0, forecast b becomes progressively more accurate. The formula for λ^* indicates that as σ_b^2 approaches 0, λ^* approaches 0, so that all weight is put on forecast b, which is also desirable. In general, the forecast with the smaller error variance receives the higher weight, with the precise size of the weight depending on the disparity between variances.

Now consider the more general and empirically-relevant case of correlated

---

[2]Strictly speaking, we need not even impose $\lambda \in [0, 1]$, but $\lambda \notin [0, 1]$ would be highly nonstandard for two valuable and sophisticated y estimates such as y_a and y_b.

errors. Under the same conditions as earlier,

$$\sigma_C^2 = \lambda^2 \sigma_a^2 + (1-\lambda)^2 \sigma_b^2 + 2\lambda(1-\lambda)\sigma_{ab}, \tag{12.3}$$

so

$$\lambda^* = \frac{\sigma_b^2 - \sigma_{ab}}{\sigma_b^2 + \sigma_a^2 - 2\sigma_{ab}}$$

$$= \frac{1 - \phi\rho}{1 + \phi^2 - 2\phi\rho},$$

where $\sigma_{ab} = cov(e_a, e_b)$ and $\rho = corr(e_a, e_b)$.

The optimal combining weight is a simple function of the variances and covariances of the underlying forecast errors. The forecast error variance associated with the optimally combined forecast is less than or equal to the smaller of σ_a^2 and σ_b^2; thus, in population, we have nothing to lose by combining forecasts, and potentially much to gain. In practical applications, the unknown variances and covariances that underlie the optimal combining weights are unknown, so we replace them with consistent estimates; that is, we estimate λ^* by replacing unknown error variances and covariances with estimates, yielding

$$\hat{\lambda}^* = \frac{\hat{\sigma}_b^2 - \hat{\sigma}_{ab}^2}{\hat{\sigma}_b^2 + \hat{\sigma}_a^2 - 2\hat{\sigma}_{ab}^2}.$$

The full formula for the optimal combining weight indicates that the variances *and* the covariance are relevant, but the basic intuition remains valid. Effectively, we're forming a portfolio of forecasts, and as we know from standard results in finance, the optimal shares in a portfolio depend on the variances *and* covariances of the underlying assets.

12.2.2   General Case

The optimal combining weight solves the following problem:

$$\min_{\lambda} \lambda' \Sigma_t \lambda \qquad\qquad (12.4)$$
$$s.t.\ \lambda' \iota = 1.$$

where Σ is the $N \times N$ covariance matrix of forecast errors and ι is a $N \times 1$ vector of ones. The solution is

$$\lambda^* = \left(\iota' \Sigma_t^{-1} \iota \right)^{-1} \Sigma_t^{-1} \iota.$$

12.3   Model-Based Combined Forecasts II: Regression-Based Forecast Combination

Now consider the regression method of forecast combination. The form of forecast-encompassing regressions immediately suggests combining forecasts by simply regressing realizations on forecasts. This intuition proves accurate, and in fact the optimal variance-covariance combining weights have a regression interpretation as the coefficients of a linear projection of y_{t+h} onto the forecasts, subject to two constraints: the weights sum to unity, and the intercept is excluded.

In practice, of course, population linear projection is impossible, so we simply run the regression on the available data. Moreover, it's usually preferable *not* to force the weights to add to unity, or to exclude an intercept. Inclusion of an intercept, for example, facilitates bias correction and allows biased forecasts to be combined. Typically, then, we simply estimate the regression,

$$y_{t+h} = \beta_0 + \beta_a y_{a,t+h,t} + \beta_b y_{b,t+h,t} + \varepsilon_{t+h,t}.$$

Extension to the fully general case of more than two forecasts is immediate.

In general, the regression method is simple and flexible. There are many variations and extensions, because any regression tool is potentially applicable. The key is to use generalizations with sound motivation. We'll give four examples in an attempt to build an intuitive feel for the sorts of extensions that are possible: time-varying combining weights, dynamic combining regressions, shrinkage of combining weights toward equality, and nonlinear combining regressions.

12.3.1 Time-Varying Combining Weights

Relative accuracies of different forecasts may change, and if they do, we naturally want to weight the improving forecasts progressively more heavily and the worsening forecasts less heavily. Relative accuracies can change for a number of reasons. For example, the design of a particular forecasting model may make it likely to perform well in some situations, but poorly in others. Alternatively, people's decision rules and firms' strategies may change over time, and certain forecasting techniques may be relatively more vulnerable to such change.

We allow for time-varying combining weights in the regression framework by using weighted or rolling estimation of combining regressions, or by allowing for explicitly time-varying parameters. If, for example, we suspect that the combining weights are evolving over time in a trend-like fashion, we might use the combining regression

$$y_{t+h} = (\beta_0^0 + \beta_0^1 TIME) + (\beta_a^0 + \beta_a^1 TIME)y_{a,t+h,t}$$

$$+(\beta_b^0 + \beta_b^1 TIME)y_{b,t+h,t} + \varepsilon_{t+h,t},$$

which we estimate by regressing the realization on an intercept, time, each of the two forecasts, the product of time and the first forecast, and the product of time and the second forecast. We assess the importance of time variation

by examining the size and statistical significance of the estimates of β_0^1 , β_a^1 , and β_b^1.

12.3.2   Serial Correlation

It's a good idea to allow for serial correlation in combining regressions, for two reasons. First, as always, even in the best of conditions we need to allow for the usual serial correlation induced by overlap when forecasts are more than 1-step-ahead. This suggests that instead of treating the disturbance in the combining regression as white noise, we should allow for $MA(h-1)$ serial correlation,

$$y_{t+h} = \beta_0 + \beta_a y_{a,t+h,t} + \beta_b y_{b,t+h,t} + \varepsilon_{t+h,t}$$

$$\varepsilon_{t+h,t} \sim MA(h-1).$$

Second, and very importantly, the $MA(h-1)$ error structure is associated with forecasts that are optimal with respect to their information sets, of which there's no guarantee. That is, although the primary forecasts were designed to capture the dynamics in y, there's no guarantee that they do so. Thus, just as in standard regressions, it's important in combining regressions that we allow either for serially correlated disturbances or lagged dependent variables, to capture any dynamics in y not captured by the various forecasts. A combining regression with $ARMA(p,q)$ disturbances,

$$y_{t+h} = \beta_0 + \beta_a y_{a,t+h,t} + \beta_b y_{b,t+h,t} + \varepsilon_{t+h,t}$$

$$\varepsilon_{t+h,t} \sim ARMA(p,q),$$

with p and q selected using information criteria in conjunction with other diagnostics, is usually adequate.

12.3.3 Shrinkage of Combining Weights Toward Equality

Simple arithmetic averages of forecasts – that is, combinations in which the weights are constrained to be equal – sometimes perform very well in out-of-sample forecast competitions, even relative to "optimal" combinations. The equal-weights constraint eliminates sampling variation in the combining weights at the cost of possibly introducing bias. Sometimes the benefits of imposing equal weights exceed the cost, so that the MSE of the combined forecast is reduced.

The equal-weights constraint associated with the arithmetic average is an example of extreme shrinkage; regardless of the information contained in the data, the weights are forced into equality. We've seen before that shrinkage can produce forecast improvements, but typically we want to *coax* estimates in a particular direction, rather than to force them. In that way we guide our parameter estimates toward reasonable values when the data are uninformative, while nevertheless paying a great deal of attention to the data when they are informative.

Thus, instead of imposing a *deterministic* equal-weights constraint, we might like to impose a *stochastic* constraint. With this in mind, we sometimes coax the combining weights toward equality without forcing equality. A simple way to do so is to take a weighted average of the simple average combination and the least-squares combination. Let the shrinkage parameter γ be the weight put on the simple average combination, and let (1-γ) be the weight put on the least-squares combination, where γ is chosen by the user. The larger is γ, the more the combining weights are shrunken toward equality. Thus the combining weights are coaxed toward the arithmetic mean, but the data are still allowed to speak, when they have something important to say.

12.3.4   Nonlinear Combining Regressions

There is no reason to force linearity of combining regressions, and various of the nonlinear techniques that we've already introduced may be used. We might, for example, regress realizations not only on forecasts, but also on squares and cross products of the various forecasts, in order to capture quadratic deviations from linearity,

$$y_{t+h} = \beta_0 + \beta_a y_{a,t+h,t} + \beta_b y_{b,t+h,t}$$

$$+\beta_{aa}(y_{a,t+h,t})^2 + \beta_{bb}(y_{b,t+h,t})^2 + \beta_{ab} y_{a,t+h,t} y_{b,t+h,t} + \varepsilon_{t+h,t}.$$

We assess the importance of nonlinearity by examining the size and statistical significance of estimates of β_{aa}, β_{bb}, and β_{ab}; if the linear combining regression is adequate, those estimates should differ significantly from zero. If, on the other hand, the nonlinear terms are found to be important, then the full nonlinear combining regression should be used.

12.3.5   Regularized Regression for Combining Large Numbers of Forecasts

Another, related, approach, involving both shrinkage and selection, is lasso and other "regularization" methods. Lasso can be used to shrink and select, and it's a simple matter to make the shrinkage/selection direction "equal weights" rather than the standard lasso "zero weights."

12.4   Application: OverSea Shipping Volume Revisited

Now let's combine the forecasts. Both failed Mincer-Zarnowitz tests, which suggests that there may be scope for combining. The correlation between the two forecast errors is .54, positive but not too high. In Table 9 we show the results of estimating the unrestricted combining regression with

$MA(1)$ errors (equivalently, a forecast encompassing test). Neither forecast encompasses the other; both combining weights, as well as the intercept, are highly statistically significantly different from zero. Interestingly, the judgmental forecast actually gets *more* weight than the quantitative forecast in the combination, in spite of the fact that its $RMSE$ was higher. That's because, after correcting for bias, the judgmental forecast appears a bit more accurate.

Table 9
Shipping Volume Combining Regression

LS // Dependent Variable is VOL
Sample: 1/01/1988 7/18/1997
Included observations: 499
Convergence achieved after 11 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.181977	0.259774	8.399524	0.0000
VOLQ	0.291577	0.038346	7.603919	0.0000
VOLJ	0.630551	0.039935	15.78944	0.0000
MA(1)	0.951107	0.014174	67.10327	0.0000

R-squared	0.957823	Mean dependent var	19.80609
Adjusted R-squared	0.957567	S.D. dependent var	3.403283
S.E. of regression	0.701049	Akaike info criterion	-0.702371
Sum squared resid	243.2776	Schwarz criterion	-0.668603
Log likelihood-528.8088	F-statistic	3747.077
Durbin-Watson stat	1.925091	Prob(F-statistic)	0.000000

Inverted MA Roots   -.95

12.5   On the Optimality of Equal Weights

12.5.1   Under Quadratic Loss

In Figure 12.1 we graph λ^* as a function of ϕ, for $\phi \in [.75, 1.45]$. λ^* is of course decreasing in ϕ, but interestingly, it is only mildly sensitive to ϕ. Indeed, for our range of ϕ values, the optimal combining weight remains close to 0.5, varying from roughly 0.65 to 0.30. At the midpoint $\phi = 1.10$, we have $\lambda^* = 0.45$.

It is instructive to compare the error variance of combined y, σ_C^2, to σ_a^2 for a range of λ values (including $\lambda = \lambda^*$, $\lambda = 0$, and $\lambda = 1$).[3] From (12.1) we have:

$$\frac{\sigma_C^2}{\sigma_a^2} = \lambda^2 + \frac{(1-\lambda)^2}{\phi^2}.$$

In Figure 12.2 we graph σ_C^2/σ_a^2 for $\lambda \in [0, 1]$ with $\phi = 1.1$. Obviously the max-

---

[3]We choose to examine σ_C^2 relative to σ_a^2, rather than to σ_b^2, because y_a is the "standard" y estimate used in practice almost universally. A graph of σ_C^2/σ_b^2 would be qualitatively identical, but the drop below 1.0 would be less extreme.
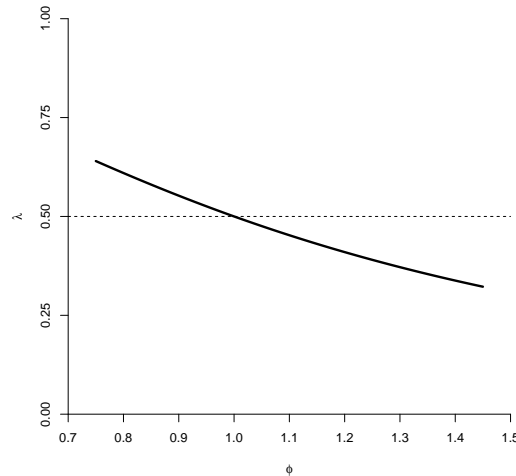
Figure 12.1: λ^* vs. ϕ. λ^* constructed assuming uncorrelated errors. The horizontal line for visual reference is at $\lambda^* = .5$. See text for details.

imum variance reduction is obtained using $\lambda^* = 0.45$, but even for nonoptimal λ, such as simple equal-weight combination ($\lambda = 0.5$), we achieve substantial variance reduction relative to using y_a alone. Indeed, a key result is that *for all* λ (except those very close to 1, of course) we achieve substantial variance reduction.

In Figure 12.3 we show λ^* as a function of ϕ for $\rho = 0, 0.3, 0.45$ and 0.6; in Figure 12.4 we show λ^* as a function of ρ for $\phi = 0.95, 1.05, 1.15$ and 1.25; and in Figure 12.5 we show λ^* as a bivariate function of ϕ and ρ. For $\phi = 1$ the optimal weight is 0.5 for all ρ, but for $\phi \neq 1$ the optimal weight differs from 0.5 and is more sensitive to ϕ as ρ grows. The crucial observation remains, however, that under a wide range of conditions it is optimal to put significant weight on both y_a and y_b, with the optimal weights not differing radically from equality. Moreover, for all ϕ values greater than one, so that less weight is optimally placed on y_a under a zero-correlation assumption, allowance for positive correlation further decreases the optimal weight placed on y_a. For a benchmark calibration of $\phi = 1.1$ and $\rho = 0.45$, $\lambda^* \approx 0.41$.

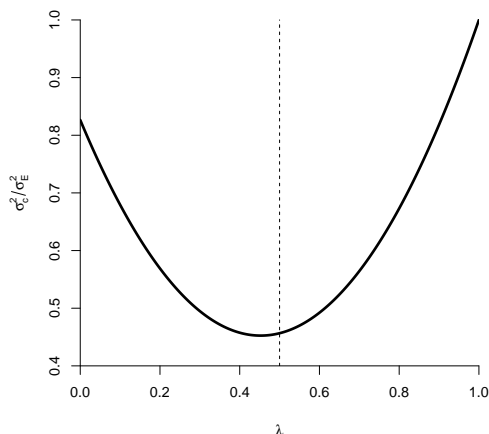Let us again compare σ_C^2 to σ_a^2 for a range of λ values (including $\lambda = \lambda^*$,

Figure 12.2: σ_C^2/σ_a^2 for $\lambda \in [0,1]$. We assume $\phi = 1.1$ and uncorrelated errors. See text for details.

$\lambda = 0$, and $\lambda = 1$). From (12.3) we have:

$$\frac{\sigma_C^2}{\sigma_a^2} = \lambda^2 + \frac{(1-\lambda)^2}{\phi^2} + 2\lambda(1-\lambda)\frac{\rho}{\phi}.$$

In Figure 12.6 we graph σ_C^2/σ_a^2 for $\lambda \in [0,1]$ with $\phi = 1.1$ and $\rho = 0.45$. Obviously the maximum variance reduction is obtained using $\lambda^* = 0.41$, but even for nonoptimal λ, such as simple equal-weight combination ($\lambda = 0.5$), we achieve substantial variance reduction relative to using y_a alone.

The "equal weights puzzle." It is clear from our analysis above that in realistic situations (similar variances, small or moderate correlations) the gains from optimally combining can be massive, and that the loss from combining with equal weights relative to optimal weights is small. That is, optimal weights are not generally equal, but combining with equal weights is often not far from the optimum, and *much* better than any primary forecast. Equal weights are fully optimal, moreover, in the equi-correlation case, or more generally, in the Elliott case. Also, from an estimation perspective, equal weights may be slightly biased, but they have no variance! So the equal weight puzzle is perhaps not such a puzzle.
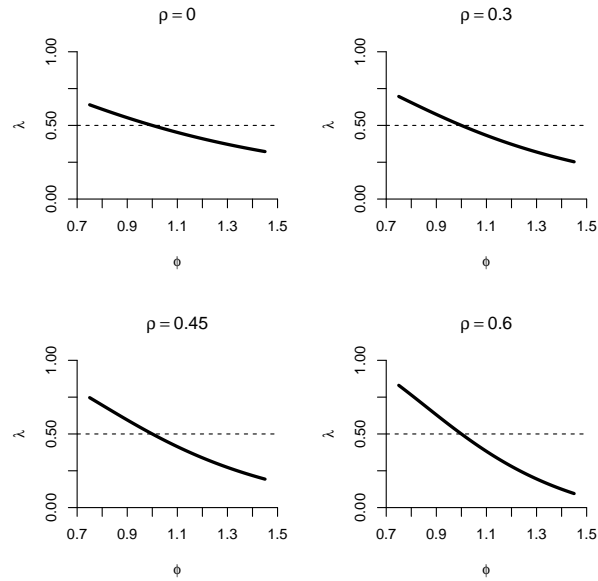
Figure 12.3: λ^* vs. ϕ for Various ρ Values. The horizontal line for visual reference is at $\lambda^* = .5$. See text for details.

12.5.2 Under Minimax Loss

Here we take a more conservative perspective on forecast combination, solving a different but potentially important optimization problem. We utilize the minimax framework of **?**, which is the main decision-theoretic approach for imposing conservatism and therefore of intrinsic interest. We solve a game between a benevolent scholar (the Econometrician) and a malevolent opponent (Nature). In that game the Econometrician chooses the combining weights, and Nature selects the stochastic properties of the forecast errors. The minimax solution yields the combining weights that deliver the smallest chance of the worst outcome for the Econometrician. Under the minimax approach knowledge or calibration of objects like ϕ and ρ is unnecessary, enabling us to dispense with judgment, for better or worse.

We obtain the minimax weights by solving for the Nash equilibrium in a two-player zero-sum game. Nature chooses the properties of the forecast errors and the Econometrician chooses the combining weights λ. For exposi-
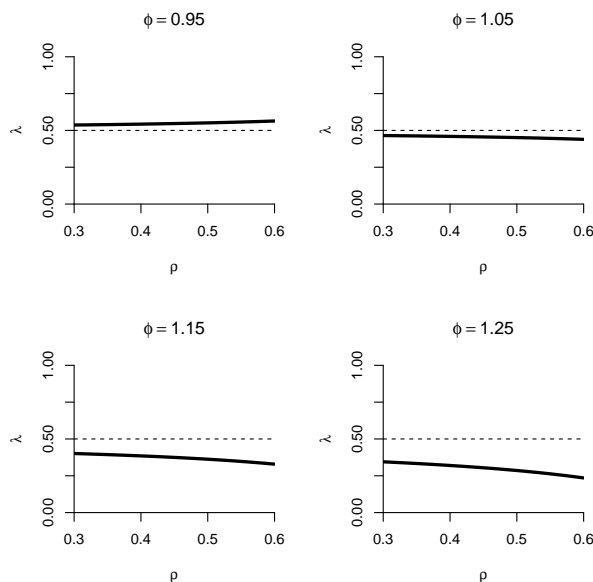
Figure 12.4: λ^* vs. ρ for Various ϕ Values. The horizontal line for visual reference is at $\lambda^* = .5$. See text for details.

tional purposes, we begin with the case of uncorrelated errors, constraining Nature to choose $\rho = 0$. To impose some constraints on the magnitude of forecast errors that Nature can choose, it is useful to re-parameterize the vector $(\sigma_b, \sigma_a)'$ in terms of polar coordinates; that is, we let $\sigma_b = \psi \cos \varphi$ and $\sigma_a = \psi \sin \varphi$. We restrict ψ to the interval $[0, \bar{\psi}]$ and let $\varphi \in [0, \pi/2]$. Because $\cos^2 \varphi + \sin^2 \varphi = 1$, the sum of the forecast error variances associated with y_a and y_b is constrained to be less than or equal to $\bar{\psi}^2$. The error associated with the combined forecast is given by
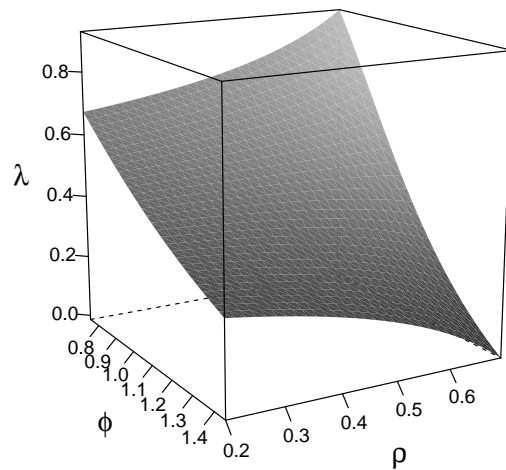
$$\sigma_C^2(\psi, \varphi, \lambda) = \psi^2 \left[\lambda^2 \sin^2 \varphi + (1 - \lambda)^2 \cos^2 \varphi \right]. \qquad (12.5)$$

so that the minimax problem is

$$\max_{\psi \in [0, \bar{\psi}], \, \varphi \in [0, \pi/2]} \, \min_{\lambda \in [0,1]} \, \sigma_C^2(\psi, \varphi, \lambda). \qquad (12.6)$$

The best response of the Econometrician was derived in (12.2) and can be expressed in terms of polar coordinates as $\lambda^* = \cos^2 \varphi$. In turn, Nature's

Figure 12.5: λ^* vs. ρ and ϕ. See text for details.

problem simplifies to

$$\max_{\psi \in [0, \bar{\psi}], \, \varphi \in [0, \pi/2]} \psi^2 (1 - \sin^2 \varphi) \sin^2 \varphi,$$

which leads to the solution

$$\varphi^* = \text{arc} \sin \sqrt{1/2}, \quad \psi^* = \bar{\psi}, \quad \lambda^* = 1/2. \tag{12.7}$$

Nature's optimal choice implies a unit forecast error variance ratio, $\phi = \sigma_a/\sigma_b = 1$, and hence that the optimal combining weight is $1/2$. If, instead, Nature set $\varphi = 0$ or $\varphi = \pi/2$, that is $\phi = 0$ or $\phi = \infty$, then either y_a or y_b is perfect and the Econometrician could choose $\lambda = 0$ or $\lambda = 1$ to achieve a perfect forecast leading to a suboptimal outcome for Nature.

Now we consider the case in which Nature can choose a nonzero correlation between the forecast errors of y_a and y_b. The loss of the combined forecast
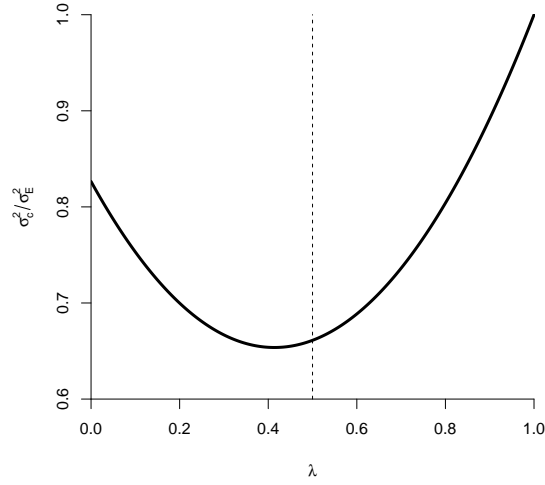
Figure 12.6: σ_C^2/σ_a^2 for $\lambda \in [0,1]$. We assume $\phi = 1.1$ and $\rho = 0.45$. See text for details.

can be expressed as

$$\sigma_C^2(\psi, \rho, \varphi, \lambda) = \psi^2 \left[\lambda^2 \sin^2 \varphi + (1-\lambda)^2 \cos^2 \varphi + 2\lambda(1-\lambda)\rho \sin \varphi \cos \varphi \right].$$
(12.8)

It is apparent from (12.8) that as long as λ lies in the unit interval the most devious choice of ρ is $\rho^* = 1$. We will now verify that conditional on $\rho^* = 1$ the solution in (12.7) remains a Nash Equilibrium. Suppose that the Econometrician chooses equal weights, $\lambda^* = 1/2$. In this case

$$\sigma_C^2(\psi, \rho^*, \varphi, \lambda^*) = \psi^2 \left[\frac{1}{4} + \frac{1}{2} \sin \varphi \cos \varphi \right].$$

We can deduce immediately that $\psi^* = \bar{\psi}$. Moreover, first-order conditions for the maximization with respect to φ imply that $\cos^2 \varphi^* = \sin^2 \varphi^*$ which in turn leads to $\varphi^* = \arc \sin \sqrt{1/2}$. Conditional on Nature choosing ρ^*, ψ^*, and φ^*, the Econometrician has no incentive to deviate from the equal-weights combination $\lambda^* = 1/2$, because

$$\sigma_C^2(\psi^*, \rho^*, \varphi^*, \lambda) = \frac{\bar{\psi}}{2} \left[\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) \right] = \frac{\bar{\psi}}{2}.$$

In sum, the minimax analysis provides a rational for combining y_a and y_b with equal weights of $\lambda = 1/2$. Of course it does not resolve the equal weights puzzle, which refers to quadratic loss, but it puts equal weights on an even higher pedestal, and from a very different perspective.

12.6 Interval Forecast Combination

12.7 Density Forecast Combination

12.7.1 Choosing Weights to Optimize a Predictive Likelihood

Has Bayesian foundations. Geweke-Amisano.

12.7.2 Choosing Weights Optimize Conditional Calibration

Maximize a test statistic for iid uniformity of the PIT.

12.8 Exercises, Problems and Complements

1. Combining Forecasts.

   You are a managing director at Paramex, a boutique investment bank in Los Angeles. Each day during the summer your two interns give you a 1-day-ahead forecast of the Euro/Dollar exchange rate. At the end of the summer, you calculate each intern's series of daily forecast errors. You find that the mean errors are zero, and the error variances and covariances are $\hat{\sigma}_{AA}^2 = 153.76$, $\hat{\sigma}_{BB}^2 = 92.16$, and $\hat{\sigma}_{AB}^2 = .2$.

   (a) If you were forced to choose between the two forecasts, which would you choose? Why?

   (b) If instead you had the opportunity to combine the two forecasts by forming a weighted average, what would be the optimal weights

according to the variance-covariance method? Why?

(c) Is it guaranteed that a combined forecast formed using the "optimal" weights calculated in part 1b will have lower mean squared prediction error? Why or why not?

2. The algebra of forecast combination.

Consider the combined forecast,

$$y_{t+h,t}^c = \lambda y_{t+h,t}^a + (1-\lambda)y_{t+h,t}^b.$$

Verify the following claims made in the text:

a. The combined forecast error will satisfy the same relation as the combined forecast; that is,

$$e_{t+h,t}^c = \lambda e_{t+h,t}^a + (1-\lambda)e_{t+h,t}^{b.}$$

b. Because the weights sum to unity, if the primary forecasts are unbiased then so too is the combined forecast.

c. The variance of the combined forecast error is

$$\sigma_c^2 = \lambda^2 \sigma_{aa}^2 + (1-\lambda)^2 \sigma_{bb}^2 + 2\lambda(1-\lambda)\sigma_{ab}^2,$$

where σ_{aa}^2 and σ_{bb}^2 are unconditional forecast error variances and σ_{ab}^2 is their covariance.

d. The combining weight that minimizes the combined forecast error variance (and hence the combined forecast error MSE, by unbiasedness) is

$$\lambda^* = \frac{\sigma_{bb}^2 - \sigma_{ab}^2}{\sigma_{bb}^2 + \sigma_{aa}^2 - 2\sigma_{ab}^2}.$$

e. If neither forecast encompasses the other, then

$$\sigma_c^2 < \min(\sigma_{aa}^2, \ \sigma_{bb}^2).$$

f. If one forecast encompasses the other, then

$$\sigma_c^2 = \min(\sigma_{aa}^2, \sigma_{bb}^2).$$

3. Quantitative forecasting, judgmental forecasting, forecast combination, and shrinkage.

Interpretation of the modern quantitative approach to forecasting as eschewing judgment is most definitely misguided. How is judgment used routinely and informally to modify quantitative forecasts? How can judgment be formally used to modify quantitative forecasts via forecast combination? How can judgment be formally used to modify quantitative forecasts via shrinkage? Discuss the comparative merits of each approach.

4. The empirical success of forecast combination.

In the text we mentioned that we have nothing to lose by forecast combination, and potentially much to gain. That's certainly true in population, with optimal combining weights. However, in finite samples of the size typically available, sampling error contaminates the combining weight estimates, and the problem of sampling error may be exacerbated by the collinearity that typically exists between $y_{t+h,t}^a$ and $y_{t+h,t}^b$. Thus, while we hope to reduce out-of-sample forecast MSE by combining, there is no guarantee. Fortunately, however, in practice forecast combination often leads to very good results. The efficacy of forecast combination is well-documented in a vast literature.

5. Regression forecasting models with expectations, or anticipatory, data.

A number of surveys exist of anticipated market conditions, investment intentions, buying plans, advance commitments, consumer sentiment, and so on.

(a) Search the World Wide Web for such series and report your results. A good place to start is the Resources for Economists page mentioned in Chapter **??**.

(b) How might you use the series you found in a regression forecasting model of y? Are the implicit forecast horizons known for all the anticipatory series you found? If not, how might you decide how to lag them in your regression forecasting model?

(c) How would you test whether the anticipatory series you found provide incremental forecast enhancement, relative to the own past history of y?

6. Crowd-sourcing via internet activity.

How, in a sense, are trends identified by search data (on Google, YouTube, ...), tweets, etc. "combined forecasts"?

7. Turning a set of point forecasts into a combined density forecast.

We can produce a combined density forecast by drawing from an estimate of the density of the combining regression disturbances, as we did in a different context in section 4.1.

12.9   Notes

The idea of forecast encompassing dates at least to Nelson (1972), and was formalized and extended by Chong and Hendry (1986) and Fair and Shiller (1990). The variance-covariance method of forecast combination is due to Bates and Granger (1969), and the regression interpretation is due to Granger

and Ramanathan (1984). Surveys of econometric forecast combination include Diebold and Lopez (1996) and Timmermann (2006). Surveys of survey-based combination include Pesaran and Weale (2006). Snowberg et al. (2013) (prediction markets) provide a nice review of prediction markets.