We denote the time- information set by . As first pass it seems most natural to think of the information set as containing the available past history of the series, where for theoretical purposes we imagine history as having begun in the infinite past So long as y is covariance stationary, however, we can just as easily express the information available at time T in terms of current a nd past shocks, Suppose, for example, that the process to be forecast is a covariance station- ary AR(1), Assembling the discussion thus far, we can view the time-T information set as containing the current and past values of either (or both) and

Based upon that information set, we want to find the optimal forecast of at some future time . The optimal forecast is the one with the smallest loss on average, that is, the forecast that minimizes expected loss. It turns out that under reasonably weak conditions the optimal forecast is the conditional mean,

the expected value of the future value of the series being forecast, conditional upon available information. In general, the conditional mean need not be a linear function of the elements of the information set. Because linear functions are particularly tractable, we prefer to work with linear forecasts – forecasts that are linear in the elements of the information set – by finding the best linear approximation to the conditional mean, called the linear projection, denoted

This explains the common term “linear least squares forecast.” The linear projection is often very useful and accurate, because the conditional mean is often close to linear. In fact, in the Gaussian case the conditional expectation is exactly linear, so that

Our forecasting method is always the same: we write out the process for the future time period of interest, , and project it on what’s known at time T , when the forecast is made. This process is best learned by example.

Consider an MA(2) process,

Suppose we’re standing at time and we want to forecast . First we write out the process for ,

Then we project on the time-T information set, which simply means that all future innovations are replaced by zeros. Thus

To forecast 2 steps ahead, we note that

and we project on the time-T information set to get

Continuing in this fashion, we see that

for all $ h>2$.

Now let’s compute the corresponding forecast errors. We have:

for all .

Finally, the forecast error variances are:

or all h > 2. Moreover, the forecast error variance for is just the unconditional variance of $y t $.

Now consider the general M A(q) case. The model is

First, consider the forecasts. If , the forecast has the form

whereas if the forecast is

Thus, an M A(q) process is not forecastable (apart from the unconditional mean) more than steps ahead. All the dynamics in the M A(q) process, which we exploit for forecasting, “wash out” by the time we get to horizon q, which reflects the autocorrelation structure of the M A(q) process. (Recallthat, as we showed earlier, it cuts off at displacement q.) Second, consider the corresponding forecast errors. They are

for and

for h > q. The h-step-ahead forecast error for is just the process itself, minus its mean. Finally, consider the forecast error variances. For h ≤ q,

whereas for ,

In summary, we’ve thus far studied the M A(2), and then the general M A(q), process, computing the optimal h-step-ahead forecast, the corresponding forecast error, and the forecast error variance. As we’ll now see, the emerging patterns that we cataloged turn out to be quite general.

## Optimal Point Forecasts for Infinite-Order Moving Averages

By now you’re getting the hang of it, so let’s consider the general case of an infinite-order M A process. The infinite-order moving average process may seem like a theoretical curiosity, but precisely the opposite is true. Any covariance stationary process can be written as a (potentially infinite-order) moving average process, and moving average processes are easy to understand and manipulate, because they are written in terms of white noise shocks, which have very simple statistical properties. Thus, if you take the time to understand the mechanics of constructing optimal forecasts for infinite moving-average processes, you’ll understand everything, and you’ll have some powerful technical tools and intuition at your command. Recall that the general linear process is Recall that the general linear process is

Then we project on the time- information set. The projection yields zeros for all of the future ’s (because they are white noise and hence unforecastable), leaving

It follows that the -step ahead forecast error is serially correlated; it follows an MA(h − 1) process,

with mean 0 and variance

A number of remarks are in order concerning the optimal forecasts of the general linear process, and the corresponding forecast errors and forecast error variances. First, the 1-step-ahead forecast error is simply . is that part of that can’t be linearly forecast on the basis of (which, again, is why it is called the innovation). Second, although it might at first seem strange that an optimal forecast error would be serially correlated, as is the case when h > 1, nothing is awry. The serial correlation can’t be used to improve forecasting performance, because the autocorrelations of the M A(h−1) process cut off just before the beginning of the time-T information set $ T , T −1 , . $This is a general and tremendously important property of the errors associated with optimal forecasts: errors from optimal forecasts can’t be forecast using information available when the forecast was made. If you can forecast the forecast error, then you can improve the forecast, which means that it couldn’t have been optimal. Finally, note that as h approaches infinity approaches zero, the unconditional mean of the rocess, and , the unconditional variance of the process, which reflects the fact that as h approaches infinity the conditioning information on which the forecast is based becomes progressively less useful. In other words, the distant future is harder to forecast than the near future! Now we construct interval and density forecasts. Regardless of whether the moving average is finite or infinite, we proceed in the same way, as follows. The definition of the h-step-ahead forecast error is

Equivalently, the h-step-ahead realized value, equals the forecast plus the error,

If the innovations are normally distributed, then the future value of the series of interest is also normally distributed, conditional upon the information set available at the time the forecast was made, and so we have the 95% h-step- ahead interval forecast . In similar fashion, we construct the h-step-ahead density forecast as

. The mean of the conditional distribution of is , which of course must be the case because we constructed the point forecast as the conditional mean, and the variance of the conditional distribution is , the variance of

As an example of interval and density forecasting, consider again the M A(2) process,

Assuming normality, the 1-step-ahead 95% interval forecast is

and the 1-step-ahead density forecast is

## Making the Forecasts Operational

So far we’ve assumed that the parameters of the process being forecast are known. In practice, of course, they must be estimated. To make our forecasting procedures operational, we simply replace the unknown parameters in our formulas with estimates, and the unobservable innovations with residuals. Consider, for example, the M A(2) process,

As you can readily verify using the methods we’ve introduced, the 2-step ahead optimal forecast, assuming known parameters, is

with corresponding forecast error

and forecast-error variance

To make the forecast operational, we replace unknown parameters with estimates and the time-T innovation with the time-T residual, yielding

and forecast error variance

Then, if desired, we can construct operational 2-step-ahead interval and den- sity forecasts, as

and

The strategy of taking a forecast formula derived under the assumption of known parameters, and replacing unknown parameters with estimates, is a natural way to operationalize the construction of point forecasts. However, using the same strategy to produce operational interval or density forecasts involves a subtlety that merits additional discussion. The forecast error variance estimate so obtained can be interpreted as one that ignores parameter estimation uncertainty, as follows. Recall once again that the actual future value of the series is

and that the operational forecast is

Thus the exact forecast error is

the variance of which is very difficult to evaluate. So we make a convenient approximation: we ignore parameter estimation uncertainty by assuming that estimated parameters equal true parameters. We therefore set to zero, which yields

with variance

which we make operational as

## Forecasting Cycles From an Autoregressive Perspective: Wold’s Chain Rule

### Point Forecasts of Autoregressive Processes

Because any covariance stationary AR(p) process can be written as an infinite moving average, there’s no need for specialized forecasting techniques for autoregressions. Instead, we can simply transform the autoregression into a moving average, and then use the techniques we developed for forecasting moving averages. It turns out, however, that a very simple recursive method for computing the optimal forecast is available in the autoregressive case. The recursive method, called the chain rule of forecasting, is best learned by example. Consider the AR(1) process,

First we construct the optimal 1-step-ahead forecast, and then we construct the optimal 2-step-ahead forecast, which depends on the optimal 1-step-ahead forecast, which we’ve already constructed. Then we construct the optimal 3-step-ahead forecast, which depends on the already-computed 2-step-ahead forecast, which we’ve already constructed, and so on. To construct the 1-step-ahead forecast, we write out the process for time ,

Then, projecting the right-hand side on the time-T information set, we obtain

Now let’s construct the 2-step-ahead forecast. Write out the process for time T + 2,

Then project directly on the time-T information set to get

Note that the future innovation is replaced by 0, as always, and that we have directly replaced the time T +1 value of y with its earlier-constructed optimal forecast. Now let’s construct the 3-step-ahead forecast. Write out the process for time T + 3,

Then project directly on the time-T information set,

The required 2-step-ahead forecast was already constructed. Continuing in this way, we can recursively build up forecasts for any and all future periods. Hence the name “chain rule of forecasting.” Note that, for the AR(1) process, only the most recent value of y is needed to construct optimal forecasts, for any horizon, and for the general AR(p) process only the p most recent values of y are needed.

## Point Forecasts of ARMA processes

Now we consider forecasting covariance stationary ARMA processes. Just as with autoregressive processes, we could always convert an ARMA process to an infinite moving average, and then use our earlier-developed methods for forecasting moving averages. But also as with autoregressive processes, a simpler method is available for forecasting ARMA processes directly, by combining our earlier approaches to moving average and autoregressive fore- casting. As always, we write out the ARM A(p, q) process for the future period of interest,

On the right side we have various future values of y and , and perhaps also past values, depending on the forecast horizon. We replace everything on the right-hand side with its projection on the time-T information set. That is, we replace all future values of y with optimal forecasts (built up recursively using the chain rule) and all future values of with optimal forecasts (0), yielding

When evaluating this formula, note that the optimal time-T “forecast” of any value of y or dated time T or earlier is just y or itself. As an example, consider forecasting the ARM A(1, 1) process,

Let’s find y\_{T +1,T} . The process at time T + 1 is

Projecting the right-hand side on yields

Now let’s find . The process at time is

Projecting the right-hand side on yields

Substituting our earlier-computed 1-step-ahead forecast yields

Continuing, it is clear that

for all .

## Interval and Density Forecasts

The chain rule, whether applied to pure autoregressive models or to ARMA models, is a device for simplifying the computation of point forecasts. Interval and density forecasts require the h-step-ahead forecast error variance, which we get from the moving average representation, as discussed earlier. It is

which we operationalize as

Note that we don’t actually estimate the moving average representation; rather, we solve backward for as many b’s as we need, in terms of the original model parameters, which we then replace with estimates. Let’s illustrate by constructing a 2-step-ahead 95% interval forecast for the ARM A(1, 1) process. We already constructed the 2-step-ahead point forecast, ; we need only compute the 2-step-ahead forecast error variance. The process is

Substitute backward for y t−1 to get

We need not substitute back any farther, because the 2-step-ahead forecast error variance is

where is the coefficient on in the moving average representation of the ARMA(1,1) process, which we just calculated to be . Thus the 2-step-pahead interval forecast is , or . We make this operational as