Augmenting our AR(4) Model of Inflation

- Adding lagged unemployment to our model of inflationary change, we get:

$$\begin{array}{llllll}
\widehat{\Delta Inf}_t = & 1.28 & -(0.31)\Delta Inf_{t-1} & -(0.39)\Delta Inf_{t-2} & +(0.09)\Delta Inf_{t-3} \\
 & (0.53) & (0.09) & (0.09) & (0.08) \\
 & & -(0.08)\Delta Inf_{t-4} & -(0.21) Unemp_{t-1} & \\
 & & (0.09) & (0.09) &
\end{array}$$

- The $\bar{R}^2 = 0.21$.
- The corresponding forecast for 2005:I is 3.9%, with a forecast error of -1.5%
- Stock and Watson consider including 2nd through 4th lags for unemployment as well and find the additional regressors are jointly significant at a 1% level.

The Autoregressive Distributed Lag (ADL) Model

- The models from the previous slide are known as Autoregressive Distributed Lag (ADL) Model.
- Specifically, with
  - p lags of the dependent variable and
  - q lags of the additional explanatory variable

  the model is call an ADP(p,q) model.
- The unemployment model on the previous slide becomes an ADL(4,1).
- Formally, the model becomes:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} \\ &+ \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + u_t \end{aligned}$$

- An underlying assumption justifying OLS in this case is that:

$$E(u_t | Y_{t-1}, Y_{t-2}, \cdots, X_{t-1}, X_{t-2}, \cdots) = 0 \qquad (13)$$

- One can obviously add multiple regressors with differing lags.

Stationarity

- A key assumption in most time series models is that of stationarity:

  A time series Y_t is stationary if its probability distribution does not change over time; i.e., if the joint distribution of $(Y_{s+1}, Y_{s+2}, \ldots, Y_{s+T})$ does not depend upon s.

- Intuitively, since we are wanting to use past values of Y_t to predict future values of Y_t, it helps if the historical distribution for Y_t carries into the future.

The Key Assumptions Underlying OLS in the AR(p,q) model

Our more general model becomes:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} \\ &+ \delta_{11} X_{1,t-1} + \cdots + \delta_{1q_1} X_{1,t-q} \\ &+ \delta_{k1} X_{k,t-1} + \cdots + \delta_{kq_k} X_{k,t-q} + u_t \end{aligned}$$

The assumptions underlying OLS are:

1. $E(u_t | Y_{t-1}, Y_{t-2}, \cdots, X_{1,t-1}, X_{1,t-2}, \cdots X_{k,t-1}, X_{k,t-2},) = 0.$
2. Has two parts:
   - $(Y_t, X_{1,t}, \ldots, X_{k,t})$ has a stationary distribution.
   - $(Y_t, X_{1,t}, \ldots, X_{k,t})$ and $(Y_{t-j}, X_{1,t-j}, \ldots, X_{k,t-j})$ become independent at j gets large (known as weak dependence).
3. $(Y_t, X_{1,t}, \ldots, X_{k,t})$ have nonzero and finite fourth moments.
4. The is no perfect multicollinearity.

Granger Causality

- With the $ADL(p, q)$ model, we are using lagged values of one variable (i.e., the X_t's) to aid in predicting another (i.e., the Y_t's).
- The Granger causality test essentially is testing the validity of this assumption using an F-test, with the null hypothesis being: $H_0 : \delta_{k1} = \cdots = \delta_{kq_k} = 0$.
- As the book notes, the terminology here is different from that used in the previous chapter. We are not saying that lagged values of the X_t's cause changes in the Y_t's.
- Instead, we are saying that "...past values of the X_t's are useful in predicting Y_t, beyond the information contained in the lagged values of Y_t.

Forecast Uncertainty and Forecast Intervals

- Consider the simple $ADL(1, 1)$ model with a forecasted value for Y_{T+1} given by
$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\delta}_1 X_T. \tag{14}$$
- The corresponding forecast error is given by

$$\begin{aligned}
\text{Forecast error} &= Y_{T+1} - \hat{Y}_{T+1|T} \\
&= [\beta_0 + \beta_1 Y_T + \delta_1 X_T + u_{T+1}] \\
&- \left[\hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\delta}_1 X_T\right] \\
&= u_{T+1} + \left[\left(\beta_0 - \hat{\beta}_0\right)\right. \\
&+ \left.\left(\beta_1 - \hat{\beta}_1\right) Y_T + \left(\delta_1 - \hat{\delta}_1\right) X_T\right]
\end{aligned}$$

The Mean Squared Forecast Error (MSFE)

- The error term u_{T+1}
  - has a mean of zero
  - is homoskedastic with variance σ_u^2 due to the stationarity assumption
  - and is uncorrelated with the OLS estimator.
- As a result:

$$\begin{aligned} MSFE &= E\left[\left(Y_{T+1} - \hat{Y}_{T+1|T}\right)^2\right] \\ &= \sigma_u^2 + var\left[\left(\beta_0 - \hat{\beta}_0\right)\right. \\ &+ \left.\left(\beta_1 - \hat{\beta}_1\right) Y_T + \left(\delta_1 - \hat{\delta}_1\right) X_T\right] \end{aligned}$$

- The former can be estimated using the square of the SER, while the latter requires computing the variance of a weighted average of the parameter estimates (using “test” in STATA).

Forecast Intervals

- One period ahead forecast intervals are formed in the usual way,

$$\hat{Y}_{T+1|T} \pm z_\alpha \times \sqrt{MSFE}. \tag{15}$$

- If we are forecasting more than one period ahead, the uncertainty becomes larger.
- Consider forecasting two periods ahead in an AR(1) Model. We know that

$$\begin{aligned} Y_{T+2} &= \beta_0 + \beta_1 Y_{T+1} + u_{T+2} \\ &= \beta_0 + \beta_1 \left[\beta_0 + \beta_1 Y_T + u_{T+1}\right] + u_{T+2} \\ &= \beta_0 + \beta_1 \beta_0 + \beta_1^2 Y_T + \beta_1 u_{T+1} + u_{T+2} \end{aligned}$$

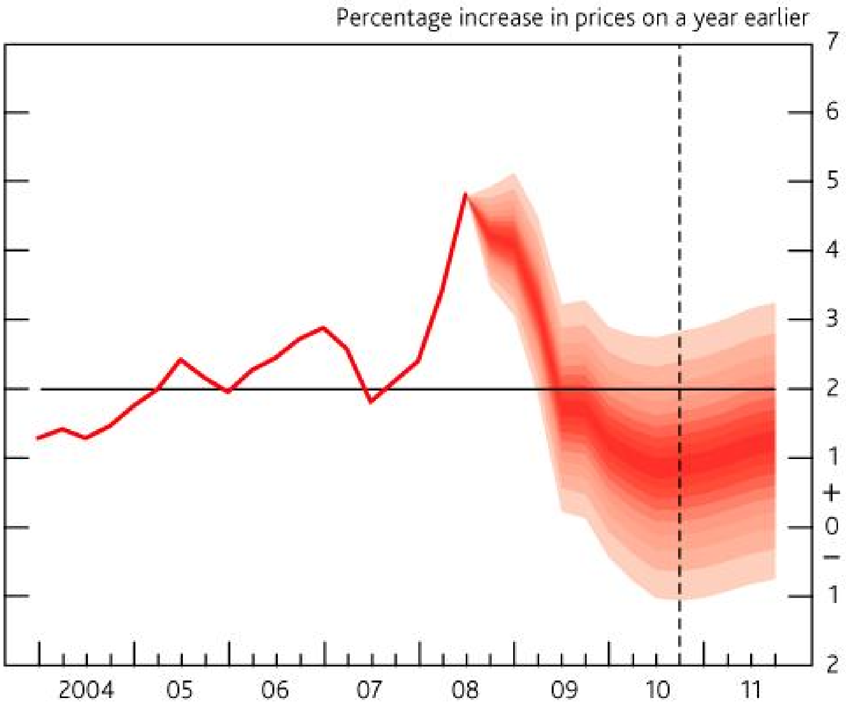
- For a large sample

$$\begin{aligned} RMSFE_2 &= \sqrt{E[(Y_{T+2} - \hat{Y}_{T+2|T})^2]} \\ &\approx \sqrt{E[\beta_1^2 u_{T+1}^2 + u_{T+2}^2]} = \sqrt{(1+\beta_1^2) var(u_t)} \approx (1+\beta_1^2) SER \end{aligned}$$

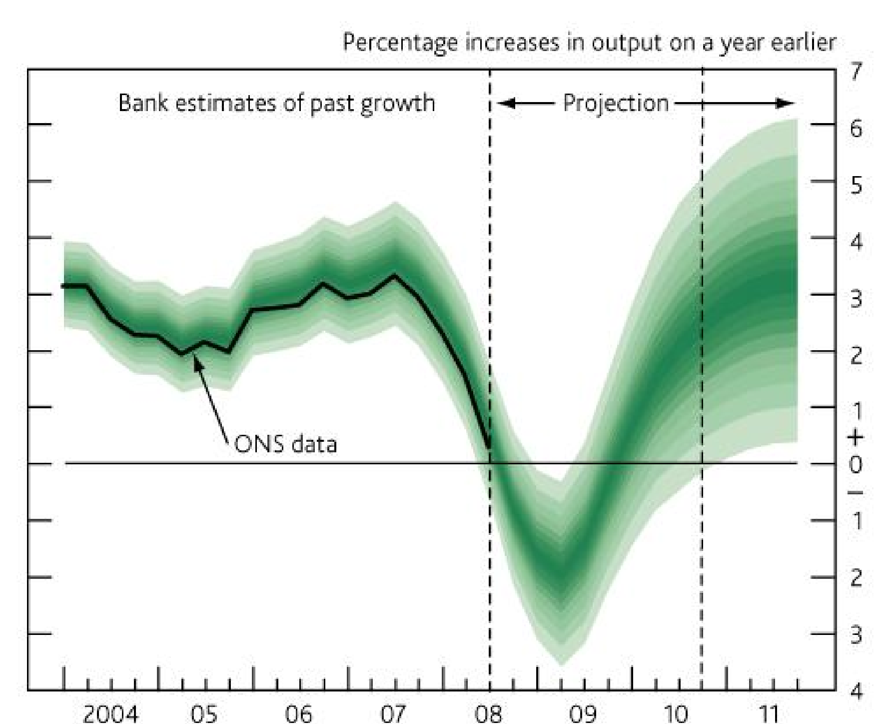
The River of Blood Graphs

- Stock and Watson note that professional forecasters "...often report confidence intervals that are tighter than 95%..." because otherwise the intervals are of "...limited use in decision making."
- This, of course, is nonsense, since tighter looking 68% intervals (one standard errors) are no more informative, only appearing to be more precise.
- The Bank of England uses a gradated chart, indicating reduced certainty as you consider light shaded regions of the projections (similar to hurricane landing forecasts).

Example #1: River of Blood Chart for Inflation



Example #2: River of Blood (Actually Green) Chart for GDP Growth



Selecting Lag Length in an AR(p) Model

- Choosing the order p of an AR model requires balancing the information lost from omitting useful past lags against estimation error.
- There are a variety of approaches
  1. F-statistic: Start with a high value of p, dropping last lag if it is statistically insignificant at a given level. $p - value$ will be incorrect.
  2. Bayes information criterion (BIC:

$$BIC(p) = ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{ln(T)}{T} \tag{16}$$

     This tradesoff the reduction in SSR versus number of lags.
  3. Akaike Information criterion (AIC):

$$AIC(p) = ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T} \tag{17}$$

     AIC tends to overestimate the number of lags.

BIC for US Inflation Case

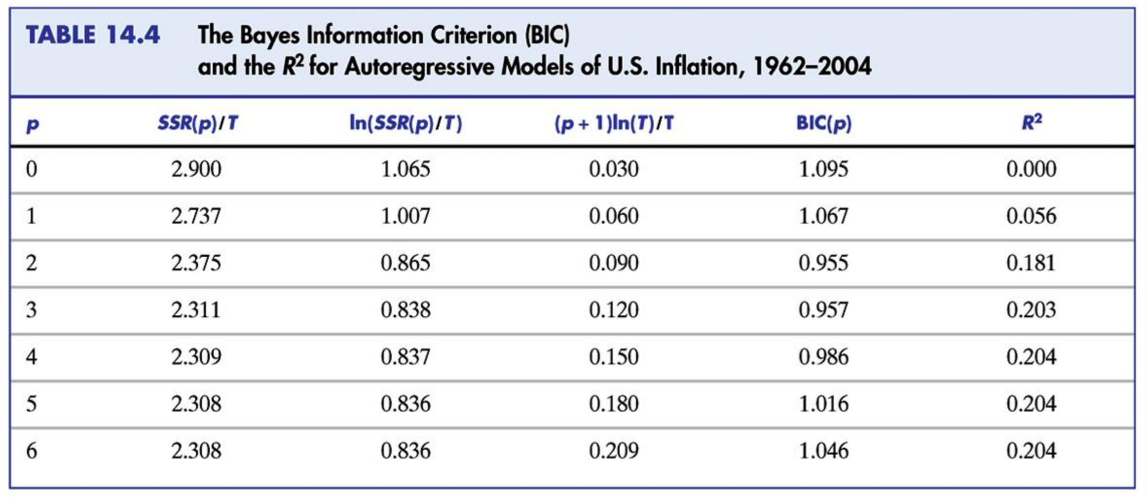


TABLE 14.4 The Bayes Information Criterion (BIC) and the R^2 for Autoregressive Models of U.S. Inflation, 1962–2004

p	$SSR(p)/T$	$\ln(SSR(p)/T)$	$(p+1)\ln(T)/T$	$BIC(p)$	R^2
0	2.900	1.065	0.030	1.095	0.000
1	2.737	1.007	0.060	1.067	0.056
2	2.375	0.865	0.090	0.955	0.181
3	2.311	0.838	0.120	0.957	0.203
4	2.309	0.837	0.150	0.986	0.204
5	2.308	0.836	0.180	1.016	0.204
6	2.308	0.836	0.209	1.046	0.204

Lag Length with Multiple Regressors

- A similar set of issues arise when there are multiple regressors, as in a ADL(p,q) setting.
- Again, an F-statistic approach can be used, but will tend to overstate the various lag lengths.
- There are counterpart expressions for the BIC and AIC. For example, for the BIC, we have

$$BIC(K) = ln\left(\frac{SSR(p)}{T}\right) + K\frac{ln(T)}{T} \tag{18}$$

where K denotes the number of coefficients including the intercept.

Trends

- A trend is a persistent, long-term movement of a variable over time, with the series fluctuating around this trend.
- In our earlier figures, we saw
  - Figure 14.1a: An upward trend in inflation prior to 1982 and a downward trend thereafter.
  - Figure 14.1c: An upward trend in Japanese GDP, or alternatively a downward trend in the growth rate of its GDP.
- Trends can be:
  - deterministic: a nonrandom function of time, such as a linear trend equal to $a \times t$ where a is a constant.
  - stochastic: a random function varying over time. These are typically viewed as more realistic, allowing for fluctuations or "surprises" over time.

The Random Walk

- The random walk is a particular stochastic trend, with

$$Y_t = Y_{t-1} + u_t \tag{19}$$

  where the u_t's (i.e., the "steps") are $i.i.d.$.
- In the case of a random walk, the best forecast of tomorrow's value for the variable is today's value.
- A slightly more general version of the random walk is the random walk with drift; i.e.,

$$Y_t = \beta_0 + Y_{t-1} + u_t \tag{20}$$

  where β_0 is the drift term.