Chapter 3

Prediction and model selection

Based on Peña et al (2000), Chapter 5.

CHAPTER 3. CONTENTS.

INTRODUCTION

- Traditionally, statistical evaluation of an Econometric model focuses on “in sample” analysis of the residuals of a fitted model.
- However, a number of studies find that models that seem to fit well by conventional in-sample criteria do poorly at out-of-sample prediction.
- Klein (1992) argue that “ability to make useful ex-ante forecasts is the real test of a model”.

Literature review

- While use of out-of-sample forecasts for model evaluation is less common than use of in-sample evidence, forecast evaluation has a long and distinguished history in economics.
- Early references include Christ (1956) and Goldberger (1959).
- Most tests that are used in practice can be grouped into one of five categories:

- Forecast encompassing
- Forecast efficiency
- Zero forecast bias
- Sign predictability
- Equal forecast accuracy between two predictive models.

- We assume an observed sample:

$$Z_T = (z_1, z_2, \ldots, z_T)'$$

- and want to generate predictions of future values given the observations,

$$z_{T+1}, z_{T+2}, \ldots, z_{T+k}$$

- T is the forecast origin and k the forecast horizon.

Three components

- 1. Estimation of new values: prediction.
- 2. Measure of the uncertainty: prediction intervals.
- 3. Arrival of new data: updating.

Chapter 3. Prediction and model selection.

3.1. The basic of forecast.

THE BASIC OF FORECAST

- Basic Questions
  - What needs to be forecast?
  - What is the forecast horizon and interval?
  - Should one develop a model?
  - What type of Model should be constructed

What needs to be forecasts

- For monetary policy decisions, inflation and output growth are the obvious candidates, but aggregate demand, future wage development, and the global outlook may be the others.

- The list is rarely exhaustive because intermediate forecast of different variables is needed very often.

- Disaggregated variables.

What is the forecast horizon and interval

- In the case of monetary policy the relevant period is generally set at between six to eight quarters, although sometimes longer-term forecasts are made.

- How often a forecast is to be made?
  - Not a clear answer. Mainly depends on the frequency of publications of national accounts and how often the institutions require them.

SHOULD ONE DEVELOP A MODEL

- If the forecast period is relatively short, there is often accurate surveys available on the plans of economic agents over its length.

- Even if the forecast horizon is relatively long, a large number of forecasts relating to a country for one- to two- year period now emanate from many sources: The economist, Consensus forescast, IMF, OECD, different documents from banks.

- Some reasons to specify a model:
  - To improve the available forecast, for instance, by using more information.
  - Transparency.

What type of models should be constructed

- Statistical vs economic models.

- Having a large number of different opinions may lead to seemingly unconsistent relations.

- Models should be simple and flexible.

What type of models should be constructed

- The "core" model: most policy-setting institutions have a "core" model that summarized the main relationships within the macroeconomy. These models usually contain about 30 or so stochastic equations and determine another 100-150 variables through identities.
- Vector autoregressions. Difficulty to use then for policy analysis: a user of VAR forecast has to accept that the policy instrument will vary continuously over the forecast horizon, something that is not easy to explain to policy-makers who are considering whether to change a policy instrument that they feel will be sustained over the forecast horizon.
- Small forward-looking models. The high degree of aggregation and the tendency to have a simplified dynamic structure means they may not be very useful for short term forecasting.
- Single equation regression models
- Dynamic optimizing models

CHAPTER 3. PREDICTION AND MODEL SELECTION.

3.2. Forecast accuracy

Forecast Accuracy

Let forecast be based upon a function $\hat{z}_{t+1}(\theta)$ such that the time $t = R, \ldots, T-1$ forecast of z_{t+1} is $\hat{z}_{t+1}(\theta) \equiv \hat{z}_{t+1}(\hat{\theta}_t)$. If we let P denote the number of one-step-ahead forecast error $\hat{e}_{t+1} = z_{t+1} - \hat{z}_{t+1}$, then

$$R + P = T + 1$$

For l-step-ahead forecast errors $\hat{e}_{t+l} = z_{t+l} - \hat{z}_{t+l}$

$$R + P = T + l$$

Measures of accuracy

$$Average\ error = \frac{1}{P-l}\sum_{i=1}^{P-l} \hat{e}_{t+i-1}(l)$$

$$MAE = \frac{1}{P-l}\sum_{i=1}^{P-l} |\hat{e}_{t+i-1}(l)|$$ (Mean Absolute Error)

$$MAPE = \frac{1}{P-l}\sum_{i=1}^{P-l} \frac{|\hat{e}_{t+i-1}(l)|}{z_{t+i+l-1}}$$ (Mean Absolute Percentage Error)

$$RMSE = \sqrt{\frac{1}{P-l}\sum_{i=1}^{P-1} \hat{z}_{t+i-1}^2(l)}$$ (Root Mean Square Error)

Forecast Combination

- The test of forecast encompassing is related to the literature on the combination of forecasts introduced by Bates and Granger (1969).

- The idea of forecast combination is that if two forecasting models are available, then taking a weighted combination of the available forecasts may generate a better forecast.

- Granger and Ramanathan (1984) suggest doing this using regression-based methods. Suppose that $\hat{z}_{1,t+1}$ and $\hat{z}_{2,t+1}$ are competing forecasts for z_{t+1}.

- They construct future forecasts using the weights α_0, α_1 and α_2 estimated by the OLS regression

$$z_{t+1} = \alpha_0 + \alpha_1 \hat{z}_{1,t+1} + \alpha_2 \hat{z}_{2,t+1} + error\ term$$

- Chong and Hendry (1986) observe that under the null that forecast from model 1 encompass the forecasts from model 2, α_1 and α_2 should be 1 and 0 respectively.

- If this is the case, then a test for forecast encompassing can be constructed, using the t-statistic associated with α_2 in the OLS estimated regression

$$\hat{e}_{1,t+1} = \alpha_0 + \alpha_2 \hat{z}_{2,t+1} + error\ term$$

- Under the null that the model 1 forecast encompasses model 2, they argue that the t-statisc should be asymptotically standard normal and hence normal tables can be used to conduct a test of encompassing.

- Ericsson (1992) notes that if the population level forecast errors are I(0) but the population level forecasts are I(1), then the equation is unbalanced. Then he suggests using the t-statistic associated with $(\hat{z}_{2,t+1} - \hat{z}_{1,t+1})$ when it and a constant are regressed on $\hat{e}_{1,t+1}$.

Mincer and Zarnowitz (1969): test of **forecast efficiency**. The test for forecast efficiency is based upon the observation that if the forecast is constructed using all available information, then the optimal forecast and the forecast error should be uncorrelated. If this is the case, then a test of efficiency may be constructed using the t-statistic associated with the estimate of α_1 in the OLS estimated regression

$$\hat{e}_{t+1} = \alpha_0 + \alpha_1 \hat{z}_{t+1} + error\ term$$

- The test for zero forecast bias is sometimes referred to as a test for zero-mean prediction error. Mincer and Zarnowitz (1969) introduce this test in the context of tests of efficiency. They note that if forecasts are unbiased then the intercept term should be zero.

- Ericsson and Marquez (1993) also give this argument in the context of regression-based test of encompassing like that in

$$\hat{e}_{1,t+1} = \alpha_0 + \alpha_2 \hat{z}_{2,t+1} + error\ term$$

- Because of this, test for zero bias are commonly reported along with tests of efficiency and encompassing.

- Berger and Krane (1985) use an F-test to construct a joint test for both efficiency and zero bias.
- Ericsson and Marquez (1993) do the same but for encompassing and zero bias.
- It is also possible to take a more direct approach when testing for zero bias. Stock and Watson (1993) and Oliner, Redebusch and Sichel (1995) construct a text for zero bias by regressing the forecast error on a constant. They then use the t-statistic associated with the intercept term to test for zero bias.

- Perhaps the most well known test of sign predictability was developed in a series of papers by Merton (1981) and Henriksson and Merton (1981).

- These authors are interested in evaluating whether decisions based upon $\hat{z}_{t+1}$ are useful in the absence of knowing the actual value of z_{t+1}.

- The context that the authors had in mind was one where decisions had to be made to either buy or sell an asset.

- For example, let z_{t+1} denote the return on an asset and let $\hat{z}_{t+1}$ denote the forecast return.

- We denote the sign of $\hat{z}_{t+1}$ by $sgn(\hat{z}_{t+1})$.

- We are interested in whether the sign of $\hat{z}_{t+1}$ is useful as a predictor of the sign of z_{t+1}. Henriksson and Merton (1981) suggest a test using the t-statistic associated with α_1 in the following OLS estimated regression:

$$1(\hat{z}_{t+1} \geq 0) = \alpha_0 + \alpha_1 1(z_{t+1} \geq 0) + error\ term,$$

where the function $1(.)$ takes the value one if the argument is true and zero otherwise.

Pesaran and Timmermann (1992) focus on the sign of the predictand y_t, and a nonparametric test is developed based on the number of correct predicted signs in the forecast series of size T. Let

$$p = sample\ proportion\ of\ times\ that\ the\ sign\ of\ y_t\ is\ correctly\ predicted,$$

$$\pi_1 = \Pr(y_t > 0),$$

$$\pi_2 = \Pr(\hat{y}_t > 0),$$

$$p_1 = sample\ proportion\ of\ times\ that\ actual\ y\ is\ positive,$$

$$p_2 = sample\ proportion\ of\ times\ that\ forecast\ y\ is\ positive$$

,

Under the null hypothesis that $\hat{z}_t$ and z_t are independently distributed of each other (so that the forecast values have no ability to predict the sign of z_t), then the number of correct sign predictions in the sample has a binomial distribution with T trials and success probability equal to

$$\pi^* = \pi_1 \pi_2 + (1 - \pi_1)(1 - \pi_2)$$

If π_1 and π_2 are known (for example, if the distributions are symmetric around zero), the test statistic is simply

$$PTK = (p - \pi^*)/[\pi^*(1 - \pi^*)/T]^{1/2}$$

When π_1 and π_2 are not known, they can be estimated by the sample proportions p_1 and p_2, so that π^* can be estimated by

$$p^* = p_1 p_2 + (1 - p_1)(1 - p_2)$$

The test statistic in this case, which Pesaran and Timmermann show to converge in distribution to N(0,1) under the null hypothesis is

$$PTNK = (p - p^*)[\widehat{var}(p) - \widehat{var}(p^*)]^{\frac{1}{2}},$$

where

$$\widehat{var}(p) = p^*(1 - p^*)/T,$$

$$\widehat{var}(p^*) = \frac{(2p_1 - 1)^2 p_2(1 - p_2)}{T} + \frac{(2p_2 - 1)^2 p_1(1 - p_1)}{T} + \frac{4p_1 p_2(1 - p_1)(1 - p_2)}{T^2}$$

- Pesaran and Timmermann generalize this test to situations where there are two or more meaningful categories for the actual and forecast values of the predictand.

- They also remark that, in the case of two categories, the square of PTNK is asymptotically equal to the chi-squared statistic in the standard goodness-of-fit test using the 2x2 contingency categorizing actual and forecast values by sign.

- A very general test of predictive ability is one that test for equal forecast accuracy across two or more models.
- To construct a test of this type one need to select a measure of accuracy.
- McCulloch and Rossi (1990), Leitch and Tanner (1991) and West, Edison and Cho (1993) compare predictive ability using economic measures of accuracy.
- In particular, the most common comparison is whether two predictive models have the same mean squared error (MSE).

A general test for equal forecast accuracy is that suggested by Diebold and Mariano (1995).

They base their test for equal MSE upon the loss differential,
$d_{t+1} = \hat{u}^2_{1,t+1} - \hat{u}^2_{2,t+1}$.
Let $\gamma_j = E d_{t+1} d_{t+1-j}$ and let $\hat{S}_{dd}$ denote a consisten estimate of
$S_{dd} = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j$.
The authors argue that the statistic $P^{-1/2} \sum_{t=R}^{R+P-1} d_{t+1}/\hat{S}_{dd}^{1/2}$ converges in distribution to a standard normal variable and hence normal tables can be used to test the null of equal forecast accuracy.

NON PARAMETRIC TESTS

Standard sign test is based on the assumption that the loss differential series is independent and identically distributed, and test the null hypothesis that the mediam of the loss-differential distribution is equal to zero.

Under the null hypothesis, the number N, of positive loss-differentials in a sample size T has a binomial distribution with a number of trials equal to T and success probability equal to ½.

The test statistic in this case is

$$SIGN = (N - 0.5T)/\left(0.5T^{1/2}\right)$$

which is asymptotically N(0,1) as sample size increases.

Wilcoxon's signed rank test also has been used. This test considers the sum of the ranks of the absolute values of positive forecast loss differentials:

$$SR = \sum I(d_t > 0) rank(|d_t|)$$

where $I(d_t > 0)$ is the indicator function, taking the value one when $d_t > 0$.

For loss-differentials that are independent, identically distributed with a symmetric distribution around zero, the exact critical values of SR are tabulated in standard texts on nonparametric methods.

When T tends to infinite

$$[SR - T(T+1)/4][(T(T+1)2T+1)/24]^{1/2} \rightarrow N(0,1)$$

- This statistic has been frequently used in recent years.
- Engel (1994), Chinn and Meese (1995) and Blomberg and Hess (1997) test wheter regime-shifting medels, error correction models, and models of political behavior, respectively, outperform the random walk in the prediction of exchange rates.
- Mark (1995) and Kilian (1999) use a bootstrapped version of this statistic to compare the preditive ability of long-horizon models of exchange rates to the random walk.
- An advantage of the technique proposed by Diebold and Mariano (1995) is that it can be extended to other measures of loss.

Chapter 3. Prediction and model selection.

3.3. Properties of MMSE (minimum mean-square error of prediction).

Properties of MMSE of prediction

Prediction by the conditional expectation

- We have T observations of a zero mean stationary time series , Z_T , and we want to forecast the value of z_{T+k} .

- In order to compare alternative forecasting procedures, we need a criterion of optimality.

PROPERTIES OF MMSE OF PREDICTION

- Minimum Mean Square Error Forecasts (MMSF). Forecasts that minimize this criterion can be computed as follows.
- Let $g_T(k)$ be the forecast we want to generate, this forecast must minimize

$$MSE(z_{T+k}, g) = E[z_{T+k} - g_T(k)\ ]^2$$

where the expected value is taken over the joint distribution of z_{T+k} and Z_T

Properties of MMSE of prediction

- Using the well-known property of

$$E(y) = E_x E_{y/x}(y)$$

- we obtain

$$MSE(z_{T+k} / Z_T) = E[z_{T+k}^2 / Z_T] + \\ + g_T(k)^2 - 2g_T(k)E[z_{T+k} / Z_T]$$

Properties of MMSE of prediction

- and taking the derivative, we obtain

$$g_T(k) = E[z_{T+k} / Z_T] = \hat{z}_T(k)$$

- This result indicates that, conditioning to the observed sample, the MMSEF is obtained by computing the conditional expectation of the random variable given the available information.

PROPERTIES OF MMSE OF PREDICTION

- <u>Linear predictions</u>

- Conditional expectations can be, in some cases, difficult to compute.
  - Restrict our search to forecasting functions that are linear functions of the observations.
- General equation for a linear predictor

$$\hat{z}_T(k) = b_{ko} z_T + \dots + b_{k(T-1)} z_1 = b_k^{'} Z_T$$

PROPERTIES OF MMSE OF PREDICTION

- calling MSEL to the mean square error of a linear forecast

$$MSEL(z_{T+k}/Z_T) = E[z_{T+k} - b_k{}'Z_T]^2$$

- minimizing this expression with respect to the parameters, we have

$$E[(z_{T+k} - b_k{}'Z_T)Z_T]$$

PROPERTIES OF MMSE OF PREDICTION

- Which implies that the best linear forecast must be such that the forecast error is uncorrelated with the set of observed variables.

- This property suggests the interpretations of the linear predictor as projections.

Properties of MMSE of prediction

- that is, finding the coefficients of the best linear predictor is equivalent to regress,

$$z_{T+k} \text{ on } Z_T$$

- then,

$$\boxed{b_k = \Gamma_T^{-1}\gamma_k}$$

where Γ_T^{-1} is the covariance matrix of Z_T and γ_k is the covariance vector between z_{T+k} and Z_T

Chapter 3. Prediction and model selection.

3.4. The computation of ARIMA forecasts.

THE COMPUTATION OF ARIMA FORECASTS

- Suppose we want to forecast a time series that follows an ARIMA(p,d,q) model. First, we will assume that the parameters are known and the prediction horizon is 1 (k=1)

$$z_{T+1} = \phi_1 z_T + \cdots + \phi_h z_{T-h+1} + a_{T+1} - \theta_1 a_T - \cdots - \theta_q a_{T-q+1}$$

where h=p+d

THE COMPUTATION OF ARIMA FORECASTS

- The one-step-ahead forecast will be,

$$\hat{z}_T(1) = E[z_{T+1}/Z_T]$$

and because the expected value for the observed sample data or the errors are themselves, and the only unknown is a_{T+1}

$$\hat{z}_T(1) = \phi_1 z_T + \cdots + \phi_h z_{T-h+1} - \theta_1 a_T - \cdots - \theta_q a_{T-q+1}$$

- Therefore, the one-step prediction error is

$$a_{T+1} = z_{T+1} - \hat{z}_T(1)$$

remember this is considering that the parameters are known, and therefore, the innovations are also known because we can compute them recursively from the observations

The computation of ARIMA forecasts

- Multiple steps ahead forecast.

$$\hat{z}_T(k) = \phi_1 \hat{z}_T(k-1) + \ldots + \phi_h \hat{z}_T(k-h) -$$

$$-\theta_1 \hat{a}_T(k-1) - \ldots - \theta_q \hat{a}_T(k-q)$$

- where

$$\hat{z}_T(j) = E[z_{T+j} / Z_T] \quad j = 1,2,\ldots k$$

$$\hat{a}_T(j) = E[a_{T+j} / Z_T] \quad j = 1,2,\ldots,k$$

The computation of ARIMA forecasts

- This expression has two parts:
  - The first one, which depends on the AR coefficients, will determine the form of the long run forecast (eventual forecast equation).
  - The second one, which depends on the moving average coefficients, will dissapear for $k>q$

The computation of ARIMA forecasts

- <u>AR(1) model</u>

$$\hat{z}_T(1) = \phi z_T$$

$$\hat{z}_T(2) = \phi \hat{z}_T(1) = \phi^2 z_T$$

$$\hat{z}_T(k) = \phi \hat{z}_T(k-1) = \phi^k z_T$$

- for large k, the term $\phi^k z_T$, and therefore, the long-run forecast (for any ARMA(p,q)) will go to the mean of the process.

THE COMPUTATION OF ARIMA FORECASTS

- Random walk with constant.

$$\hat{z}_T(1) = c + z_T$$
$$\hat{z}_T(2) = c + \hat{z}_T(1) = 2c + z_T$$
$$\hat{z}_T(k) = c + \hat{z}_T(k-1) = kc + z_T$$

- The forecasts follow a straight line with slope c. If c=0, all forecasts are equal to the last observed value.

CHAPTER 3. PREDICTION AND MODEL SELECTION.

3.3. Interpreting the forecasts from ARIMA models.

Interpretation of the forecasts

Nonseasonal models.

- The eventual forecast function of a nonseasonal ARIMA model verifies for k>q

$$\phi(B)(\nabla^d \hat{z}_T(k) - \mu) = 0$$

- where

$$\mu = mean(\nabla^d z_t)$$

- Espasa and Peña (1995) proved that the general solution for this equation can be written as,

$$\hat{z}_T(k) = P_T(k) + t_T(K) \quad k > \max(0, q - d - p)$$

- where, the permanent component is,

$$\nabla^d P_T(k) = \mu$$

INTERPRETATION OF THE FORECASTS

- and the transitory component is,

$$\phi(B)t_T = 0$$

- Permanent component will be given by

$$P_T(k) = \beta_0^{(T)} + \beta_1^{(T)}k + \cdots + \beta_d^{(T)}k^d$$

with $\beta_d = {}^{\mu}/_{d!}$ determined by the mean of the stationary process

Interpretation of the forecasts

- whereas the rest of the parameters, $\beta_i^{(T)}$, depend on the initial values and change with the forecast origin.

- <u>Examples:</u>

$$P_T(k) = \mu \quad d = 0$$

$$P_T(k) = \beta_0^{(T)} + \mu k \quad d = 1$$

$$P_T(k) = \beta_0^{(T)} + k\beta_1^{(T)} + \mu k^2/2 \quad d = 2$$

INTERPRETATION OF THE FORECASTS

- 1. will be constant for all horizons.

- 2. deterministic linear trend with slope μ , if $\mu = 0$, then the permanent component is just a constant.

- 3. the solution is a quadratic trend with the leading term determined by μ. If $\mu = 0$ the equation reduces to a linear trend, but now the slope depends on the origin of the forecast.

INTERPRETATION OF THE FORECASTS

- In summary, the long-run forecast from an ARIMA model is the mean if the series is stationary and a polynomial for nonstationary models.

  - In this last case, the leading term of the polynomial is a constant (when the mean is different from zero), whereas it depends on the forecast origin (adaptative) if the mean is zero.

Interpretation of the forecasts

It is interesting to compare

1) the straight-line forecasts generated by a I(1) model with a constant term

2) A model I(2) without constant

Interpretation of the forecasts

Let β_1 be the slope in the straight line generated by the forecasts from the I(1) model

$$\hat{\beta}_1 = \frac{1}{n-1}\sum_{t=2}^{n} \nabla z_t$$

Interpretation of the forecasts

Let us compare this forecast with the one generated by

$$\nabla^2 z_T = (1 - \theta B) a_T$$

that also generates the following straight line

$$\hat{z}_T(1) = 2z_T - z_{T-1} - \theta a_T = z_T + \hat{\beta}_2$$

where

$$\hat{\beta}_2 = z_T - z_{T-1} - \theta a_T$$

Interpretation of the forecasts

For the next period

$$\hat{z}_T(2) = 2\hat{z}_T(1) - z_T = z_T + 2\hat{\beta}_2$$

In general, for any $k > 0$ the forecasts are

$$\hat{z}_T(k) = z_T + k\hat{\beta}_2$$

The forecast will follow a straight line with slope $\hat{\beta}_2$.

Interpretation of the forecasts

This slope incorporates the sample information. As $a_T = (1-\theta B)^{-1}\nabla^2 z_T$, then the slope can be written as

$$\hat{\beta}_2 = \nabla z_T - \theta(1-\theta B)^{-1}(\nabla z_T - \nabla z_{T-1})$$

which can be written as

$$\hat{\beta}_2 = (1-\theta)\sum_{i=0}^{i=T-1}\theta^i \nabla z_{T-i}$$

Interpretation of the forecasts

The slope is a weighted mean of the observed growths with weights decreasing with the lag.

I(2) models compute the slope, β_2, as a weighted average of the observed growth values but given more weight to the last observed growths and less to the most remote ones.

Important difference between a I(1) model with a constant and a I(2) model

- The I(1) model with constant makes a simple average: past growths are as relevant as the latest to forecast the next growth.

- The I(2) model makes a weighted average with weights that decrease exponentially with time, so that past growths have smaller weights that do the latest growth.

Interpretation of the forecasts

Important advantage

Models without constant are more adaptive than models that include a constant

Implication: when in doubt, it is better to differentiate in order to have a model without a constant to make the model more robust and flexible.

In general integrated models incorporate time information in an intuitively sensible way.

This is an important difference with respect to nondynamic models, as linear resgression

Interpretation of the forecasts

- Transitory component. Can be given by

$$t_T(k) = \sum_{i=1}^{p} A_i G_i^k$$

where G_i^{-1} are the roots of the AR polynomial and A_i are coefficientes depending on the forecast origin.

INTERPRETATION OF THE FORECASTS

- Example. Consider the model,

$$(1-\phi B)\nabla z_t = a_t$$

- then $G_1 = \phi$, and the forecasts must have the form,

$$\hat{z}_T(k) = c_T + A_1\phi^k$$

where c_T ,the constant that appears as the solution of $\nabla P_T(k) = 0$ and A_1 , the constant in the transitory equation must be determined by the initial conditions

INTERPRETATION OF THE FORECASTS

- The transitory component can be obtained by

$$\hat{z}_T(1) = c_T + A_1\phi = z_T + \phi(z_T - z_{T-1})$$

$$\hat{z}_T(2) = c_T + A_1\phi^2$$
$$= z_T + \phi(z_T - z_{T-1}) + \phi^2(z_T - z_{T-1})$$

- and the solution of these two equations is

$$c_T = z_T + \frac{\phi(z_T - z_{T-1})}{1-\phi}$$

Interpretation of the forecasts

- and,

$$A_1 = -\frac{\phi(z_T - z_{T-1})}{1-\phi}$$

- these results indicate that the forecasts are slowly approaching the long run forecast. Note that as $A_1\phi^k$ goes to zero, the adjustment made by the transitory decreases exponentially.

Interpretation of the forecasts

Seasonal models

For seasonal processes the forecast will satisfy the equation

$$\Phi(B^s)\phi(B)\left(\nabla_s^D \nabla^d \hat{z}_t(k) - \mu\right) = 0$$

This equation can also be decomposed into a term associated to the nonstationary part and another linked to the stationary part.

Assume that $D = 1$. Then, the seasonal difference can be written as

$$(1 - B^s) = (1 + B + B^2 + \cdots + B^{s-1})(1 - B)$$

INTERPRETATION OF THE FORECASTS

and using the following notation

$$S_s(B) = 1 + B + B^2 + \cdots + B^{s-1}$$

the forecast equation can be written as

$$\Phi(B^s)\phi(B)\left(S_s(B)\nabla^{d+1}\hat{z}_t(k) - \mu\right) = 0$$

which has the property that all the operators involved do not share roots in common.

Interpretation of the forecasts

The solution of this equation for $k > max(0, q + sQ - d - s - p - SP)$ is given by

$$\hat{z}_T(k) = T_T(k) + E_T(k) + t_T(k)$$

where now the permanent component has been split into two terms

INTERPRETATION OF THE FORECASTS

The first one is the trend component, and it is the solution of

$$\nabla^{d+1} T_T(k) = \frac{\mu}{s}$$

the second is the seasonal component that is the solution of

$$S_s(B) E_T(k) = 0$$

Finally the transitory component is now the solution of

$$\Phi(B^s)\phi(B) t_T(k) = 0$$

and will die out for large horizon.

INTERPRETATION OF THE FORECASTS

The seasonal component will be given by

$$\sum_{j=1}^{s} E_T(j) = \sum_{s+1}^{2s} E_T(j) = 0$$

the solution of this equation is a function of period s and values summing zero each s lags.

The coefficients of this function are called seasonal coefficients, and they will be changing over time because they depend on the forecast origin.

INTERPRETATION OF THE FORECASTS

Example: One of the models most often used with economic and business monthly data.

$$\nabla\nabla_{12} z_t = (1-\theta B)(1-\Theta B^{12})a_t$$

The equation of the forecast generated by this model is

$$\hat{z}_t(k) = \hat{z}_t(k-1) + \hat{z}_t(k-12) - \hat{z}_t(k-13) - \theta\hat{a}_t(k-1) - \Theta\hat{a}_t(k-12) + \theta\Theta\hat{a}_t(k-13)$$

INTERPRETATION OF THE FORECASTS

this equation can be written for $k > 0$ as

$$\hat{z}_t(k) = \beta_0^{(t)} + \beta_1^{(t)} k + S_k^{(t)}$$

that is, a linear trend plus a seasonal component with coefficients that are changing over time.

In order to determine the parameters, we need to know the initial conditions.

Interpretation of the forecasts

we can write, for $j = 1, \dots, 13$

$$\hat{z}_t(j) = \hat{\beta}_0^{(t)} + \hat{\beta}_1^{(t)} j + S_j^{(t)}$$

with $S_j^{(t)} = S_{j+12}^{(t)}$,

Interpretation of the forecasts

It is interesting to compare the previous model with a model with deterministic seasonality

The deterministic seasonality model has the form

$$\varphi(L)\nabla y_t = \alpha_1\delta_{1t} + \alpha_2\delta_{2t} + \alpha_3\delta_{3t} + \alpha_4\delta_{4t} + u_t$$

δ_{it} for i=1,...,4 are seasonal dummies which take the value one when the observation t falls in the quarter i and are otherwise zero.

The polynomial $\varphi(L)$ is assumed to be of order r, stationary and hence to have all roots outside the unit circle. In order to capture stationary stochastic seasonality, $\varphi(L)$ may include important nonzero coefficients at the annual lags of 4,8, etc.

Interpretation of the forecasts

The deterministic seasonality is indicated by the fact that

$$\mathrm{E}(\nabla y_t) = E(\nabla y_{t-4})$$

But this mean varies with the quarter.

For example, if r=1 and t falls in quarter 1

$$\nabla y_t = \varphi_1 \nabla y_{t-1} + \alpha_1 \delta_{1t} + \alpha_2 \delta_{2t} + \alpha_3 \delta_{3t} + \alpha_4 \delta_{4t} + u_t$$

Interpretation of the forecasts

If we substitute recursively

$$\nabla y_t = \varphi_1(\varphi_1 \nabla y_{t-1} + \alpha_1\delta_{1t-1} + \alpha_2\delta_{2t-1} + \alpha_3\delta_{3t-1} + \alpha_4\delta_{4t-1} + u_t) + \alpha_1\delta_{1t} + \alpha_2\delta_{2t} + \alpha_3\delta_{3t} + \alpha_4\delta_{4t} + u_t$$

$$\mathrm{E}(\nabla y_t) = \varphi_1(\varphi_1 \nabla y_{t-1} + \alpha_4) + \alpha_1$$

Then successively substituting

$$\mathrm{E}(\nabla y_t) = \frac{\alpha_1 + \varphi_1\alpha_4 + \varphi_1^2\alpha_3 + \varphi_1^3\alpha_2}{1 - \varphi_1^4}$$

INTERPRETATION OF THE FORECASTS

However, if the annual growth is of interest, then the identity

$$\nabla_4 y_t = \nabla_1 y_t + \nabla_1 y_{t-1} + \nabla_1 y_{t-2} + \nabla_1 y_{t-3}$$

For the case of general r implies

$$E(\nabla_1 y_t) = \frac{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}{1 - \varphi_1 - \cdots - \varphi_r},$$

so that the expected annual growth is constant over t.

Interpretation of the forecasts

- In contrast to the four unit-root processes implicitly present in the seasonally integrated model, the model here contains only one unit-root process.
- This unit-root process implies that y_t will wander widely and smoothly over time, but in these meanderings y_t is linked to y_{t-1}.
- Put a different way, the change in y_t from one season to the next has a constant variance over time and hence the values for any two values cannot deviate “too far” from each other over time.
- Thus, unlike the seasonality integrated process, here “summer cannot become winter”.

Chapter 3. Prediction and model selection.

3.7. Prediction confidence intervals

PREDICTION CONFIDENCE INTERVALS

Let's consider the MA representation from a general ARIMA model

$$z_t = \psi(B)a_t$$

the ψ_i parameters are obtained by using the relationship

$$\phi(B)(1-B)^d\psi(B) = \theta(B)$$

Then, we can write

$$z_{T+k} = \sum_{0}^{\infty} \psi_i a_{T+k-i}\ , \quad \psi_0 = 1$$

Prediction confidence intervals

and taking expected values conditional to the observed data, we have that

$$\hat{z}_T(k) = \sum_{0}^{\infty} \psi_{k+j} a_{T-j}$$

The forecast error is

$$\begin{aligned} e_T(k) &= z_{T+k} - \hat{z}_T(k) \\ &= a_{T+k} + \psi_1 a_{T+k-1} + \cdots + \psi_{k-1} a_{T+1} \end{aligned}$$

with variance

$$Var\big(e_T(k)\big) = \sigma^2 (1 + \psi_1^2 + \cdots + \psi_{k-1}^2)$$

Prediction confidence intervals

This equation indicates that the uncertainty of the long-run forecasts is different for stationary and nonstationary models

Stationary model

$$\psi_k^2 \to 0 \text{ when } k \to \infty$$

For example, for an AR(1) model

$$\psi_k = \phi^k$$

and

$$Var\big(e_T(k)\big) = \frac{\sigma^2}{(1-\phi^2)}$$

Prediction confidence intervals

The long-run forecast goes to the mean and the uncertainty of this forecast is finite.

Although the uncertainty can be much larger than σ^2, the uncertainty of the one-step-ahead forecast remains bounded.

However, when ϕ goes to one the variance of the forecast grows without bounds

This means that we cannot make useful long-run forecasts for nonstationary models because the uncertainty will go to infinite.

Prediction confidence intervals

If the distribution of the forecast error is known, we can compute confidence intervals for the forecast or prediction confidence intervals.

For instance, assuming normal errors, the 95% confidence interval for the random variable z_{T+k} is

$$\hat{z}_T(k) \pm 1.96\sigma(1 + \psi_1^2 + \cdots + \psi_{k-1}^2)^{1/2}$$

Prediction confidence intervals

- Unknown parameter values. It can be shown that the uncertainty introduced in the forecast for this additional source is small for moderate sample size, and can be ignored in practice.

- Suppose an AR(1) model,

Prediction confidence intervals

then the forecast is

$$\hat{z}_T(1) = \hat{\phi} z_T$$

and the true forecast error, $e_T(1) = a_T$, is related to the observed forecast error, $e_T^*(1) = z_{T+1} - \hat{\phi} z_T$, by

$$e_T(1) = e_T^*(1) + (\hat{\phi} - \phi) z_T$$

Prediction confidence intervals

Assuming to simplify that z_T is fixed, and using that

$$Var(\hat{\phi}) = \frac{\sigma^2}{\sum z_{t-1}^2}$$

we have that

$$Var\left(e_T^*(1)\right) = \sigma^2(1 + z_T^2/ns_z^2)$$

where $ns_z^2 = \sum z_{t-1}^2$.

PREDICTION CONFIDENCE INTERVALS

The forecast error has two components

The first one, σ^2, is the uncertainty due to the random behavior of the observation we want to forecast

The second component measures the parameter uncertainty because the parameters are estimated from the sample. This second term is or order 1/n, and it can be safely ignored for medium or large sample size.

PREDICTION CONFIDENCE INTERVALS

In general

- slope change

- model specification

- forecast-origin uncertainty

- slope estimation

CHAPTER 3. PREDICTION AND MODEL SELECTION.

3.8. Forecasting updating.

FORECASTING UPDATING.

How forecast are adapted when new observations become available

$$\hat{z}_T(k) = \psi_k a_T + \psi_{k+1} a_{T-1} + \cdots$$

$$\hat{z}_{T+1}(k-1) = \psi_{k-1} a_{T+1} + \psi_k a_T + \cdots$$

which leads to

$$\hat{z}_T(k-1) = \hat{z}_{T+1}(k) + \psi_{k-1} a_{T+1}$$

Testing for model stability

Box and Tiao (1976)

If the model is correct and we call $\hat{a}_{T+j}$ to the one-step-ahead forecast errors computed from the estimated parameter values, we have the statistic

$$Q = \frac{\sum_{j=1}^{h} \hat{a}_{T+j}^2}{\sigma^2}$$

will be distributed as a χ^2 distribution with h degrees of freedom.

Forecasting updating.

As σ^2 will be estimated by the sample residual variance $\hat{\sigma}^2$, the statistic

$$Q^* = \frac{\sum_{j=1}^{h} \hat{a}_{T+j}^2 / h}{\hat{\sigma}^2}$$

will be distributed as an F distribution with h and $n - p - q$ degrees of freedom.

CHAPTER 3. PREDICTION AND MODEL SELECTION.

3.9. Model selection criteria.

MODEL SELECTION CRITERIA.

Suppose we want to select the order of an AR(p) model in such a way that the out-of-sample one-step-ahead prediction mean-squared error is minimized.

This MSE is given by

$$MSE(z_{T+1}) = E\left[z_{T+1} - \hat{\phi}' Z_p\right]$$

Now

$$Z_p = \begin{pmatrix} z_T \\ \vdots \\ z_{T-p} \end{pmatrix}$$

and the expectation is taken with respect to the joint distribution of the variables $\left(z_{T+1}, \hat{\phi}', Z_p\right)$

The forecast error can be decomposed as

$$e_{T+1} = z_{T+1} - \phi' Z_p + (\phi - \hat{\phi})' Z_p$$

and so

$$MSE(z_{T+1}) = \sigma^2 + E\left[(\phi - \hat{\phi})' Z_p Z_p{}' (\phi - \hat{\phi})\right]$$

which decompose the forecast error as the sum of the variable uncertainty and the parameter uncertainty.

We can assume, to simplify, that the variables $\left(z_{T+1}, Z_p\right)$ are independent of $\hat{\phi}$

Model selection criteria.

Under this assumption, we can compute the expectation

$$E\left[(\phi - \hat{\phi})' Z_p Z_p{}' (\phi - \hat{\phi})\right]$$

with respect to $\left(Z_0, Z_p\right)$

$$MSE(z_{T+1} \backslash Z_0) = \sigma^2 + E\left[(\phi - \hat{\phi})' \Gamma_p (\phi - \hat{\phi})\right]$$

where $\Gamma_p = E\left(Z_p Z_p{}'\right)$

Model selection criteria.

This function depends on the sample only through $\hat{\phi}$

As

$$\sqrt{n}\left(\phi - \hat{\phi}\right) \sim N\left(0, \sigma^2 \Gamma_p^{-1}\right)$$

the quadratic form

$$(\phi - \hat{\phi})' \Gamma_p (\phi - \hat{\phi}) n / \sigma^2$$

is asymptotically a χ_p^2 distribution and the expectation can be approximated by

$$MSE(z_{T+1}) = \sigma^2\left(1 + \frac{p}{n}\right)$$

An unbiased estimate of σ^2 is ${n\hat{\sigma}^2}/_{(n-p)}$ where $\hat{\sigma}^2$ is the MLE estimate.

Inserting this estimate in the last equation, we have an estimation of the out-of-sample forecast error.

To minimize this value, the order p of the AR model should be chosen by minimizing

$$FPE = \frac{\hat{\sigma}^2(n+p)}{(n-p)}$$

The final criterion, final prediction error (FPE) combines fitting, as given by $\hat{\sigma}^2$, with parsimony, due to the penalty introduced by the term $\frac{(n+p)}{(n-p)}$ for increasing the order p.

MODEL SELECTION CRITERIA.

An equivalent form for this criterion is

$$logFPE = log\hat{\sigma}^2 + log\, n\left(1 + \frac{p}{n}\right) - log\, n\left(1 - \frac{p}{n}\right)$$

Multiplying the equation for n, we obtain the AIC criterion

$$AIC = nlog\hat{\sigma}^2 + 2p$$

The insight of this criterion is

$$AIC = -2(\log maximized\ likelihood) + 2(numer\ of\ parameters)$$

For ARMA models this reduces, dropping constant, to

$$AIC = nlog\hat{\sigma}^2 + 2(p + q)$$

MODEL SELECTION CRITERIA.

- Bayesian Information Criterion (BIC).

$$BIC = n \log \hat{\sigma}^2 + (\log n) p$$

- In this criteria the penalty is grater than in AIC, so BIC tends to select simpler models.

$$AIC = -2(deviance) + 2(p + q)$$

$$BIC = -2(deviance) + (\log n)(p + q)$$