Calculating Interval Forecasts

Chapter 7 (Chatfield)

Monika Turyna & Thomas Hrdina

Department of Economics, University of Vienna

Summer Term 2009

Terminology

- An interval forecast consists of an upper and a lower limit between which a future value is expected to lie with a prescribed probability.
- Example: Inflation in the next quarter will lie in the interval $[1\%, 2.5\%]$ with a 90% probability
- The limits are called *forecast limits* or *prediction bounds* while the interval is referred to as *prediction interval (P.I.)*
- Note: the term confidence interval usually applies to estimates of fixed but unknown parameter values while a P.I. is an estimate of an unknown future value of a random variable

Focus of Attention

- In what follows we concentrate on computing P.I.s for a single variable at a single (future) time point
- We do not cover the more difficult problem of calculating a P.I. for a single variable over a longer time horizon or a P.I. for different variables at the same time point
- Furthermore, we will cover a variety of approaches for calculating P.I.s – the various methods for forecasting time-series require different approaches

Importance of P.I.s

- Predictions in form of point forecasts provide no guidance as to their likely accuracy
- P.I.s, in contrast,
  - allow to assess future uncertainty,
  - enable different strategies to be planned for different outcomes indicated by the P.I. and
  - make a thorough comparison of forecasts from different methods possible

Some Reasons for Disregard

- Rather neglected topic in the statistical literature, i.e. textbooks and journal papers
- No generally accepted method of calculating P.I.s
- Theoretical P.I.s difficult or impossible to evaluate (e.g. for some multivariate or non-linear models)
- Properties of empirically based methods – based on within-sample residuals – have been little studied
- Software packages do not produce P.I.s at all or only on a limited scale

Density Forecasts

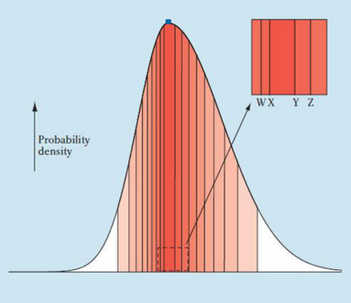
- Finding the entire probability distribution of a future value is called *density forecasting*
- For linear models with normally distributed innovations, the density forecast is usually a normal distribution with the mean equal to the point forecast and the variance equal to that used for computing the prediction interval.
- Of course, given the density forecast one can construct the P.I.s for any desired level of probability
- Problem: when forecast error distribution is not normal

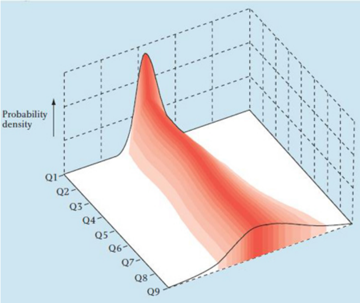
Fan Charts

- Idea: for consecutive future values, construct several P.I.s for different probabilities (e.g. from 10% to 90%) and plot them in the same graph using different levels of shading for different probabilities
- Usually the darkest shade covers the P.I.s with a 10% probability while the lightest shade covers the P.I.s with a 90% probability
- Intervals typically get wider, indicating increasing uncertainty about future values, i.e. they fan out
- "Fan charts *could* become a valuable tool for presenting the uncertainty attached to forecasts [. . .]" (Chatfield, 2000)

Density Forecasts & Fan Charts

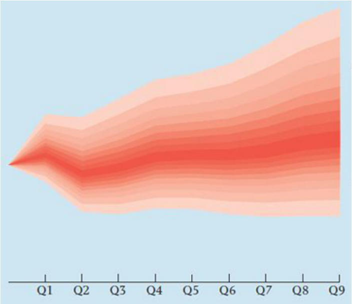
Example

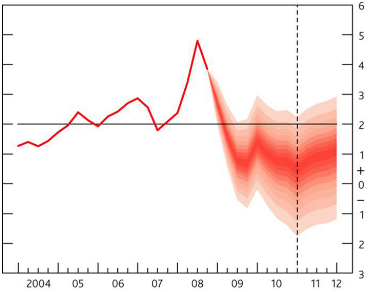




Density Forecasts & Fan Charts

Example (cont.)





Notation

Conditional Forecast Error

- An observed time series $(x_1, x_2, \ldots, x_N)$ is regarded as a finite realisation of a stochastic process $\{X_t\}$
- The point forecast of the random variable X_{N+h}, conditional on data up to time N, is denoted by $\hat{X}_N(h)$ when regarded as a random variable and by $\hat{x}_N(h)$ when regarded as a particular value
- The *conditional forecast error* – conditional on data up to time N and on the particular forecast – is the random variable $e_N(h) = X_{N+h} - \hat{x}_N(h)$
- The observed value of $e_N(h) = x_{N+h} - \hat{x}_N(h)$ only becomes available at time $N + h$

Notation

Forecast Errors & Fitted Residuals

- It is important to differentiate between the (out-of-sample) conditional forecast errors $e_N(h)$, the fitted residuals (within-sample "forecast" errors) and the model innovations
- The out-of-sample observed forecasting errors $e_N(h) = x_{N+h} - \hat{x}_N(h)$ are true forecasting errors
- The within-sample observed "forecasting" errors $[x_t - \hat{x}_{t-1}(1)]$, for $t = 2, \dots, N$, are merely the residuals from the fitted model, i.e. the difference between observed and fitted values

Notation

Forecast Errors & Fitted Residuals (cont.)

- These fitted residuals will not be the same as the true model innovations because they depend on parameter estimates (and perhaps on estimated starting values)
- They are also not true forecasting errors if the parameters have been estimated using data up to N
- However, if one finds the "true" model and the latter does not change then it is reasonable to expect the true forecast errors to have similar properties as the fitted residuals
- Practice: true forecast errors tend to be larger than expected from within-sample fit (change in the model?)

Prediction Mean Square Error (PMSE)

- The uncertainty of an h-step-ahead forecast of a single variable is assessed with its prediction mean square error given by $E[e_N(h)^2]$
- For an unbiased forecast, i.e. where the conditional expectation $E[X_{N+h}] = \hat{x}_N(h)$, it holds that $E[e_N(h)] = 0$ and thus $E[e_N(h)^2] = Var[e_N(h)]$
- Note: we are not interested in the variance of the forecast but in the variance of the forecast error (the particular value of the point forecast $\hat{x}_N(h)$ is determined exactly and has variance zero)

Prediction Mean Square Error (PMSE)

Example

- Consider the zero mean AR(1) process $X_t = \phi_1 X_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ are independent $N(0, \sigma_\epsilon^2)$
- Assuming that we know ϕ_1 and σ_ϵ^2, it can be shown that the "true-model" PMSE – the true forecast error variance – is given by

$$E[e_N(h)^2] = \sigma_\epsilon^2(1 - \phi_1^{2h})/(1 - \phi_1^2) \tag{1}$$

- In practice we would have to replace ϕ_1 and σ_ϵ^2 with sample estimates

Prediction Mean Square Error (PMSE)

Example (cont.)

- Assume that we want to calculate the PMSE of a one-step-ahead forecast, i.e. we want to know $E[e_N(1)^2]$
- Since we do not know the true parameter values, the one-step-ahead forecasting error is given by

$$
\begin{aligned}
e_N(1) &= X_{N+1} - \hat{x}_N(1) \\
&= \phi_1 x_N + \epsilon_{N+1} - \hat{\phi}_1 x_N = (\phi_1 - \hat{\phi}_1) x_N + \epsilon_{N+1}
\end{aligned}
$$

- If the estimate $\hat{\phi}_1$ is unbiased then we get $E[e_N(1)^2] = E[\epsilon_{N+1}^2] = \sigma_\epsilon^2$ which is the same as for $h = 1$ in equation (1)

Prediction Mean Square Error (PMSE)

Bias Correction

- The previous example shows that analysts have to bear in mind the effects of parameter uncertainty on estimates of the PMSE
- Though PMSE estimates can be corrected to allow for parameter uncertainty, the formulas are complex and corrections are merely of order $1/N$
- *"Overall, the effect of parameter uncertainty seems likely to be of a smaller order of magnitude in general than that due to other sources, notably the effects of model uncertainty and the effects of errors and outliers [...]."* (Chatfield, 2000)

- General formula for a $100(1-\alpha)\%$ P.I for X_{N+h}

$$\hat{x}_N(h) \pm z_{\alpha/2}\sqrt{\mathrm{Var}[e_N(h)]}$$

  - Symmetric about $\hat{x}_N(h)$ – assumes the forecast is unbiased $E[e_N(h)^2] = \mathrm{Var}[e_N(h)]$
  - Assumes errors are normally distributed [sometimes for short series $z_{\alpha/2}$ replaced by the respective percentage point of a t-distribution]
  - The latter assumption usually violated even for a linear model with Gaussian innovations, when model parameters estimated from the same data used to compute forecasts
- For any method the main problem is to evaluate $\mathrm{Var}[e_N(h)]$

P.I.s derived from a probability model

- Ignore parameter uncertainty
- Example – ARIMA forecasting:
  - $X_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$
  - then: $e_N(h) = [X_{N+h} - \hat{X}_{N+h}] = \varepsilon_{N+h} + \sum_{j=1}^{h-1} \psi_j \varepsilon_{N+h-j}$
  - thus: $\mathrm{Var}[e_N(h)] = [1 + \psi_1^2 + \cdots + \psi_{h-1}^2]\sigma_\varepsilon^2$
  - In practice replace ψ_i and σ_ε^2 with $\hat{\psi}_i$ and $\hat{\sigma}_\varepsilon^2$

P.I.s derived from a probability model cont'd

- Similar formulas available for: Vector ARMA models, structural state–space models and various regressions (the latter typically allow for parameter uncertainty and are conditional in the sense that they depend on the particular values of the explanatory variables from where a prediction is being made)
- Typically not available for: complicated simultaneous equation models, non–linear models, ARCH and other stochastic volatility models
- P.I.s usually overoptimistic if the true model is not known and must be chosen from a class of models

P.I.s without model identification

What to do when a forecasting method is selected without any formal model identification procedure?

- assume that the method is optimal (in some sense)
- apply some empirical procedure

P.I.s assuming a method is optimal

Example: Exponential smoothing

- Optimal for ARIMA(0,1,1)
- PMSE formula: $\text{Var}[e_N(h)] = [1 + (h-1)\alpha^2]\sigma_e^2$ where $\sigma_e^2 = \text{Var}[e_n(1)]$
- Should one use this formula even if the model has not been formally identified?
- Reasonable if:
  - Observed one-step-ahead forecast errors show no obvious autocorrelation
  - No other obvious features of the data (e.g. trend) which need to be modelled

Forecasting not based on a probability model

- Assume that the method is optimal in the sense that the one-step ahead errors are uncorrelated.
- Easy to check by looking at the correlogram of the one-step-ahead errors:
  - if there is correlation we have more structure in the data which should improve the forecast.
- If we assume that one–step–ahead errors have also equal variance it should be possible to evaluate $\text{Var}[e_N(h)]$ in terms of $\text{Var}[e_N(1)]$
- Example: Applied to Holt–Winters method with additive and multiplicative seasonality; in the additive case found to be equivalent to results from SARIMA for which Holt–Winters is optimal

"Approximate" formulae

- Used if no theoretical formula is available
- Sometimes, due to simplicity, used even if there exists a theoretical formula
- Usually very inaccurate
- Examples:
  1. $\text{Var}[e_N(h)] = h\sigma_e^2$
     where $\sigma_e^2 = \text{Var}[e_N(1)]$
     only true in a random walk model
  2. $\text{Var}[e_N(h)] = (0.659 + 0.341h)^2\sigma_e^2$
  3. Some formulas for the Holt–Winters method [Bowermann and O'Connel 1987]

Methods based on the observed distribution

1
- Apply the forecasting method to all the past data
- Find the within–sample forecast errors at 1,2,3. . . steps ahead from all available time origins
- Find the variance of these errors
- Assuming normality an approximate $100(1-\alpha)\%$P.I. for X_{N+h} is

$$\hat{x}_N(h) \pm z_\alpha s_{e,h}$$

  where $s_{e,h}$ is the standard deviation of the h–steps–ahead errors
- Values of $s_{e,h}$ can be unreliable for small N and large h

Methods based on the observed distribution cont'd

2

- Split the past data into two parts, fit the method to the first part
- Make predictions of the second part
- Resulting "errors" are much more like true forecast errors than those of the first method
- Refit the model with one additional observation in the first part and one less in the second and so on
- Interestingly it has been found that errors follow a gamma rather than normal distribution

Simulation – Monte Carlo approach

- Given the probability time–series model, simulate past and future behavior by generating a series of random variables
- Repeat many times to obtain a large set of outcomes, called pseudo–data
- Evaluate P.I. by finding the interval within which the required percentage of future values lie
- Assumes the model has been correctly identified

Resampling – Bootstrapping

- Sample from the empirical distribution of past fitted "errors"
- The procedure approximates the theoretical distribution of innovations by the empirical distribution of the observed residuals – a distribution–free approach
- Since in the time series context successive observations are correlated over time, thus bootstrapping over fitted errors [which are hopefully approximately independent] rather then independent observations
- However: procedure highly dependent on the fitted model choice

The Bayesian approach

- Given a suitable model, allows computation of a complete probability distribution for a future value
- Once probability distribution known, compute P.I.s by decision–theoretic approach or Bayesian version of the P.I. general formula
- Alternatively: simulate the predictive distribution
- Bayesian model averaging
  - Natural if the analyst relies not on a single model but on a mixture
  - Use Bayesian methods to find a sensible set of models and to average over these in a appropriate way

P.I.s for transformed variables

- Non–linear transformation of variable X_t: $Y_t = g(X_t)$ [e.g. logarithmic or Box–Cox transformation]
- P.I.s for Y_{N+h} can be calculated in an appropriate way
- But: How to get back P.I.s for the original variable?
- Usually point forecast for X_{N+h} namely $g^{-1}[\hat{y}_N(h)]$is biased since $E[g^{-1}(Y)] \neq g^{-1}[E(Y)]$
- E.g.: If the predictive distribution of Y_{N+h} is symmetric with mean $\hat{y}_N(h)$ then $g^{-1}[\hat{y}_N(h)$ is the median of X_{N+h}
- Fortunately: If the P.I. for Y_{N+h} has a probability $(1-\alpha)$ then the retransformed P.I. for X_{N+h} will have the same probability
- Often P.I. for X_{N+h} will be asymmetric

Why are P.I.s too narrow?

- Empirical evidence shows that out-of-sample forecast errors tend to be larger than model-fitted residuals, implying that more than 5% of future observations will fall outside a 95% P.I. on average
- The various reasons why P.I.s are too narrow in general include
  1. parameter uncertainty
  2. non-normally distributed innovations
  3. identification of the wrong model
  4. changes in the structure underlying the model

Why are P.I.s too narrow? (cont.)

- ad 1) This problem can be mitigated by using theoretical PSME formulae which account for parameter uncertainty by incorporating correction terms; corrections of this form seem to be negligible in comparison with those accounting for other sources of uncertainty
- ad 2) A common problem which is often associated with asymmetry or heavy tails; in particular the latter, which is due to outliers and errors in the data, can have severe effects on model identification and on the resulting point forecasts and P.I.s

Why are P.I.s too narrow? (cont.)

- ad 3) This problem is related to model uncertainty and can be mitigated by applying appropriate diagnostic checks (e.g. checking whether the one-step-ahead fitted residuals are uncorrelated and have constant variance; if this is not the case than there is more structure which should be exploited)
- ad 4) The underlying structure changes if it slowly evolves over time or there are sudden shifts; in both cases the model parameters will change (Chatfield argues that an observation that falls outside a P.I. could indicate a change in the underlying model)

Summary

- The basic message from this lecture is that P.I.s are a reasonable extension to point forecasts
- Wherever applicable, P.I.s should be calculated by formulating a model that approximates the DGP well and from which the PMSE can be used to obtain the interval
- To allow for parameter uncertainty correction terms can be used to calculate the PMSE; since the correction is rather small these terms are often omitted
- It is vital to note that the approach just described rests on the assumptions that the correct model has been fitted, the errors are normally distributed and that the future is like the past

Summary (cont.)

- The various "approximate" formulae for calculating P.I.s can be very inaccurate and should therefore not be used
- If theoretical formulae are not available or there is doubt about the assumptions stated above, empirically based resampling methods are a good alternative
- One should always bear in mind that P.I.s tend to be too narrow in practice, in particular because the wrong model has been identified or the model has changed
- Alternatively one could also consider different approaches to calculating P.I.s (e.g. Bayesian model averaging)