Economic Forecasting
Lecture 5: Forecasting Cycles

Richard G. Pierse

1 Introduction

In last week's lecture we developed the *ARMA* model as a flexible way to model cycles in a stationary series. In this lecture we consider forecasting with the *ARMA* model.

2 Optimal Forecasts

An *optimal forecast* is a forecast that minimises the expected value of the forecaster's loss function

$$L(e_t)$$

given the information set $\mathbf{\Omega}_t$. In the *ARMA* model, the information set will contain current and past values of y_t and the disturbances ε_t so that

$$\mathbf{\Omega}_t = \{y_t, y_{t-1}, y_{t-2}, \cdots, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\}. \quad (1)$$

Under reasonably weak conditions, it can be shown that the optimal forecast of y_{t+h} given the information set at time t, $\mathbf{\Omega}_t$, is the *conditional mean*

$$\mathrm{E}(y_{t+h}|\mathbf{\Omega}_t).$$

3 Forecasting moving average processes

3.1 The first order moving average process

Let us look at the first order moving average model

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1} \quad (2)$$

where

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{var}(\varepsilon_t) = \sigma^2, \quad \mathrm{cov}(\varepsilon_t \varepsilon_{t-s}) = 0, \; s \neq 0.$$

Firstly, consider forecasting one period ahead. Leading the equation by one period, we have

$$y_{t+1} = \varepsilon_{t+1} + \theta\varepsilon_t$$

and, taking expectations conditional on the information set at time t, (1),

$$\mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) = \theta\varepsilon_t$$

since the future innovation ε_{t+1} has expected value zero, or formally,

$$\mathrm{E}(\varepsilon_{t+1}|\mathbf{\Omega}_t) = 0.$$

The one period ahead *forecast error* is

$$e_{t+1,t} = y_{t+1} - \widehat{y}_{t+1,t} = \varepsilon_{t+1}$$

and has variance

$$\mathrm{var}(e_{t+1,t}) = \mathrm{var}(\varepsilon_{t+1}) = \sigma^2 = \sigma_1^2.$$

Now consider forecasting two periods ahead. Leading the equation by two periods we have

$$y_{t+2} = \varepsilon_{t+2} + \theta\varepsilon_{t+1}$$

and, taking expectations conditional on the information set at time t,

$$\mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) = 0$$

since

$$\mathrm{E}(\varepsilon_{t+2}|\mathbf{\Omega}_t) = \mathrm{E}(\varepsilon_{t+1}|\mathbf{\Omega}_t) = 0.$$

The two-step ahead forecast error is

$$e_{t+2,t} = y_{t+2} - \widehat{y}_{t+2,t} = \varepsilon_{t+2} + \theta\varepsilon_{t+1}$$

and has variance

$$\mathrm{var}(e_{t+2,t}) = \mathrm{var}(\varepsilon_{t+2}) + \theta^2\,\mathrm{var}(\varepsilon_{t+1}) = (1+\theta^2)\sigma^2 = \sigma_2^2.$$

More generally, for $h > 1$, we have

$$\mathrm{E}(y_{t+h}|\mathbf{\Omega}_t) = 0$$

and

$$e_{t+h,t} = y_{t+h} - \widehat{y}_{t+h,t} = \varepsilon_{t+h} + \theta\varepsilon_{t+h-1}$$

with

$$\mathrm{var}(e_{t+h,t}) = \mathrm{var}(\varepsilon_{t+h}) + \theta^2\,\mathrm{var}(\varepsilon_{t+h-1}) = (1+\theta^2)\sigma^2 = \sigma_h^2.$$

Thus the optimal h-step ahead forecast for a first order moving average process is zero, for $h > 1$ and the variance of the forecast error is $(1+\theta^2)\sigma^2$. Note also that for $h > 1$, the forecast error is autocorrelated and itself follows an *MA(1)* process.

3.2 The second order moving average process

For the second order moving average model

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \tag{3}$$

we have, for *one-period ahead forecasts*,

$$\mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) = \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}$$

and

$$e_{t+1,t} = \varepsilon_{t+1}$$

with

$$\mathrm{var}(e_{t+1,t}) = \sigma^2 = \sigma_1^2.$$

For *two-step ahead forecasts*,

$$\mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) = \theta_2 \varepsilon_t$$

and

$$e_{t+2,t} = \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1}$$

with

$$\mathrm{var}(e_{t+2,t}) = (1 + \theta_1^2)\sigma^2 = \sigma_2^2$$

and, for forecasts of three or more periods ahead,

$$\mathrm{E}(y_{t+h}|\mathbf{\Omega}_t) = 0$$

and

$$e_{t+h,t} = \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \theta_2 \varepsilon_{t+h-2}$$

with

$$\mathrm{var}(e_{t+h,t}) = (1 + \theta_1^2 + \theta_2^2)\sigma^2 = \sigma_h^2.$$

The forecast errors are autocorrelated and follow an *MA(2)* process.

3.3 The *q-th* order moving average process

Now consider the general *MA(q)* model

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

For $h \leq q$,

$$\mathrm{E}(y_{t+h}|\mathbf{\Omega}_t) = \theta_h \varepsilon_t + \cdots + \theta_q \varepsilon_{t+h-q}$$

and

$$e_{t+h,t} = \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \cdots + \theta_{h-1}\varepsilon_{t+1}$$

with

$$\text{var}(e_{t+h,t}) = (1 + \cdots + \theta_{h-1}^2)\sigma^2 = \sigma_h^2.$$

For $h > q$,

$$\text{E}(y_{t+h}|\mathbf{\Omega}_t) = 0$$

and

$$e_{t+h,t} = \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \cdots + \theta_q \varepsilon_{t+h-q}$$

with

$$\text{var}(e_{t+h,t}) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2 = \sigma_h^2.$$

Consider the properties of the forecast as the forecast horizon h increases. Note that for $h > q$, the optimal forecast is zero so that the *MA(q)* process is not forecastable more than q steps ahead. The forecast error variance increases with h until $h = q+1$, after which it is constant. The forecast error follows an *MA(h-1)* process until $h > q$ when it becomes an *MA(q)* process.

3.4 The infinite order moving average process

Recall from last week that the *Wold Representation Theorem* states that any covariance stationary variable y_t with mean μ has the *infinite order moving average* representation

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$

Forecasting this process h-periods ahead we have

$$y_{t+h} - \mu = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + \cdots$$

and

$$\text{E}(y_{t+h} - \mu|\mathbf{\Omega}_t) = \psi_h \varepsilon_t + \psi_{h+1}\varepsilon_{t+1} + \psi_{h+2}\varepsilon_{t-2} + \cdots .$$

The forecast error $e_{t+h,t}$ is the *MA(h-1)* process

$$e_{t+h,t} = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \cdots + \psi_{h-1}\varepsilon_{t+1}$$

wth variance

$$\text{var}(e_{t+h,t}) = (1 + \psi_1^2 + \cdots + \psi_{h-1}^2)\sigma^2 = \sigma_h^2.$$

As $h \to \infty$, $\text{E}(y_{t+h} - \mu|\mathbf{\Omega}_t) \to 0$ and the optimal forecast becomes simply the unconditional mean μ and the forecast error variance becomes the unconditional variance of y_t

$$\sigma^2 \sum_{i=0}^{\infty} \psi_i^2.$$

As the horizon h increases, the process becomes more difficult to forecast until, in the limit, the best forecast is the unconditional mean μ.

3.5 Interval and density forecasts

So far, we have been looking at point forecasts. Interval and density forecasts can also be defined if we make an assumption about the distribution of the disturbance process ε_t. If we assume that the ε_t are independently normally distributed

$$\varepsilon_t \sim N(0, \sigma^2)$$

then y_{t+h}, conditional on information available in period t, will also be normally distributed with

$$y_{t+h} \sim N(y_{t+h,t}, \sigma_h^2). \tag{4}$$

Equation (4) defines a *density forecast*. It also follows that a 95% h-step ahead *interval forecast* of y_{t+h} is given by

$$y_{t+h,t} \pm 1.96\sigma_h.$$

3.6 Making the forecasts operational

So far we have ignored the fact that the moving average parameters θ are unknown and must be estimated. To make the forecasts operational we must replace the unknown parameters θ by the parameter estimates $\widehat{\theta}$. Similarly, the unknown disturbances ε_t are replaced by the estimated residuals $\widehat{\varepsilon}_t$. For example, in the *MA(2)* case we have

$$\widehat{y}_{t+1,t} = \widehat{\theta}_1 \widehat{\varepsilon}_t + \widehat{\theta}_2 \widehat{\varepsilon}_{t-1}$$

and

$$\widehat{y}_{t+2,t} = \widehat{\theta}_2 \widehat{\varepsilon}_t.$$

The fact that we use estimated parameters in place of true parameter values introduces an additional uncertainty into the forecast. For example the true forecast error is

$$\begin{aligned} \widehat{e}_{t+2,t} &= y_{t+2} - \widehat{y}_{t+2,t} \\ &= \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t - \widehat{\theta}_2 \widehat{\varepsilon}_t. \end{aligned}$$

The variance of this error includes the effect of the estimation error $\widehat{\theta}_2 - \theta_2$ and is difficult to calculate. Because of this difficulty, we ignore parameter estimation uncertainty and use the operational formula

$$\widehat{\sigma}_2^2 = (1 + \widehat{\theta}_1^2)\widehat{\sigma}^2$$

for the estimated variance of our forecast error even though this formula underestimates the true variance because it neglects parameter uncertainty.

4 Forecasting autoregressive processes

Since an autoregressive process can always be inverted to give a moving average process, we do not really need to consider forecasting autoregressive processes specially, having already considered forecasting of moving average processes. Nevertheless, it turns out to be useful to consider forecasting autoregressive processes explicitly since this leads to a useful *chain rule of forecasting.*

Consider the first order autoregressive process defined by

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{5}$$

where

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{var}(\varepsilon_t) = \sigma^2, \quad \mathrm{cov}(\varepsilon_t \varepsilon_{t-s}) = 0,\ s \neq 0.$$

Firstly, consider forecasting this process one period ahead. Leading the equation by one period, we have

$$y_{t+1} = \phi y_t + \varepsilon_{t+1}$$

and, taking expectations conditional on the information set at time t, (1), defines the forecast

$$\mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) = \phi y_t \tag{6}$$

with forecast error variance

$$\mathrm{var}(e_{t+1,t}) = \mathrm{var}(\varepsilon_{t+1}) = \sigma^2.$$

Forecasting two-periods ahead we have

$$y_{t+2} = \phi y_{t+1} + \varepsilon_{t+2}$$

and

$$\mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) = \phi\, \mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) = \phi^2 y_t \tag{7}$$

from (6). The forecast error variance is

$$\mathrm{var}(e_{t+2,t}) = \mathrm{var}(\phi\varepsilon_{t+1} + \varepsilon_{t+2}) = (1+\phi^2)\sigma^2.$$

Similarly,

$$\mathrm{E}(y_{t+3}|\mathbf{\Omega}_t) = \phi\, \mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) = \phi^3 y_t$$

with forecast variance

$$\mathrm{var}(e_{t+3,t}) = \mathrm{var}(\phi^2\varepsilon_{t+1} + \phi\varepsilon_{t+2} + \varepsilon_{t+3}) = (1+\phi^2+\phi^4)\sigma^2$$

and so on. We can build up h-period ahead forecasts recursively, each forecast using previously defined forecasts. This is called the *chain rule of forecasting.*

The variance of the forecasts increases as the horizon h increases until it reaches the unconditional variance of the process

$$(1 + \phi^2 + \phi^4 + \phi^6 + \cdots)\sigma^2 = \frac{\sigma^2}{1 - \phi^2}.$$

For the general *AR(p)* case

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

we have

$$\begin{aligned} \mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) &= \phi_1 y_t + \phi_2 y_{t-1} + \cdots + \phi_p y_{t-p+1} \\ \mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) &= \phi_1 \, \mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) + \phi_2 y_t + \cdots + \phi_p y_{t-p+2} \\ \mathrm{E}(y_{t+3}|\mathbf{\Omega}_t) &= \phi_1 \, \mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) + \phi_2 \, \mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) + \phi_3 y_t + \cdots + \phi_p y_{t-p+3} \end{aligned}$$

and so on. Only the p most recent observations of y_t are needed to generate forecasts indefinitely far into the future. The forecast error variance increases with the forecast horizon h until it reaches the unconditional variance of the process.

5 Forecasting with *ARMA* models

Consider the general *ARMA(p,q)* model

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$

Leading by h-periods, we have

$$y_{t+h} = \phi_1 y_{t+h-1} + \cdots + \phi_p y_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \cdots + \theta_q \varepsilon_{t+h-q}$$

and, taking expectations conditional on the information set at time t,

$$\mathrm{E}(y_{t+h}|\mathbf{\Omega}_t) = \phi_1 y_{t+h-1,t} + \cdots + \phi_p y_{t+h-p,t} + \varepsilon_{t+h,t} + \theta_1 \varepsilon_{t+h-1,t} + \cdots + \theta_q \varepsilon_{t+h-q,t}.$$

All future values of y are replaced by recursively defined optimal forecasts and all future values of ε are replaced by their optimal forecast of zero. All forecast values of y or ε dated at time t or earlier are replaced by their *actual* values.

For example, in the *ARMA(1,1)* model

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{8}$$

we have

$$\mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) = \phi_1 y_t + \theta_1 \varepsilon_t \tag{9}$$

and

$$\mathrm{E}(y_{t+2}|\mathbf{\Omega}_t) = \phi_1 \, \mathrm{E}(y_{t+1}|\mathbf{\Omega}_t) = \phi_1^2 y_t + \phi_1\theta_1\varepsilon_t. \tag{10}$$

To compute the variance of the forecast errors from an *ARMA* model, we note that any *ARMA* model has a moving average representation and that the variance of h-step ahead forecast errors from a moving average model is given by the formula

$$\mathrm{var}(e_{t+h,t}) = \sigma_h^2 = (1 + \psi_1^2 + \cdots + \psi_{h-1}^2)\sigma^2.$$

Substituting backwards for y in the model (8) gives

$$\begin{aligned} y_t &= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1} + \theta_1\varepsilon_{t-2}) + \varepsilon_t + \theta_1\varepsilon_{t-1} \\ &= \varepsilon_t + (\phi_1 + \theta_1)\varepsilon_{t-1} + \phi_1\theta_1\varepsilon_{t-2} + \phi_1^2 y_{t-2}. \end{aligned}$$

We can continue to substitute backwards to get further terms in ε_{t-j} but we need go no further in order to define the variance of the two-step ahead forecast, which only depends on $\psi_1 = \phi_1 + \theta_1$. For the one-step ahead forecast (9) we have

$$\mathrm{var}(e_{t+1,t}) = \sigma^2$$

and, for the two-step ahead forecast (10), we have

$$\mathrm{var}(e_{t+2,t}) = (1 + \psi_1^2)\sigma^2 = (1 + (\phi_1 + \theta)^2)\sigma^2.$$

We can define a 95% 2-step-ahead *interval forecast* of y_{t+2} as

$$\phi_1^2 y_t + \phi_1\theta_1\varepsilon_t \pm 1.96\sigma\sqrt{1 + (\phi_1 + \theta_1)^2}.$$

Replacing the unknown parameters with their estimates, we get the *operational version* of this interval forecast:

$$\widehat{\phi}_1^2 y_t + \widehat{\phi}_1\widehat{\theta}_1\widehat{\varepsilon}_t \pm 1.96\widehat{\sigma}\sqrt{1 + (\widehat{\phi}_1 + \widehat{\theta}_1)^2}.$$

6 Example: Canadian Employment

As an example, we will consider forecasting with the *ARMA* models for the Canadian Employment index which were developed in last week's lecture. We will use two of the better models we fitted: the *MA(8)* model and the *AR(2)* model.

Figure 1 shows the forecasts and 95% confidence bands from the *MA(8)* model. Note that the forecast reverts to the unconditional mean of the series after 8 quarters as expected. The error bands spread out and reach their maximum extent also after 8 quarters.Figure 2 shows the forecast together with historical data for the estimation period: *1963-1994*. It can be seen that the forecast is

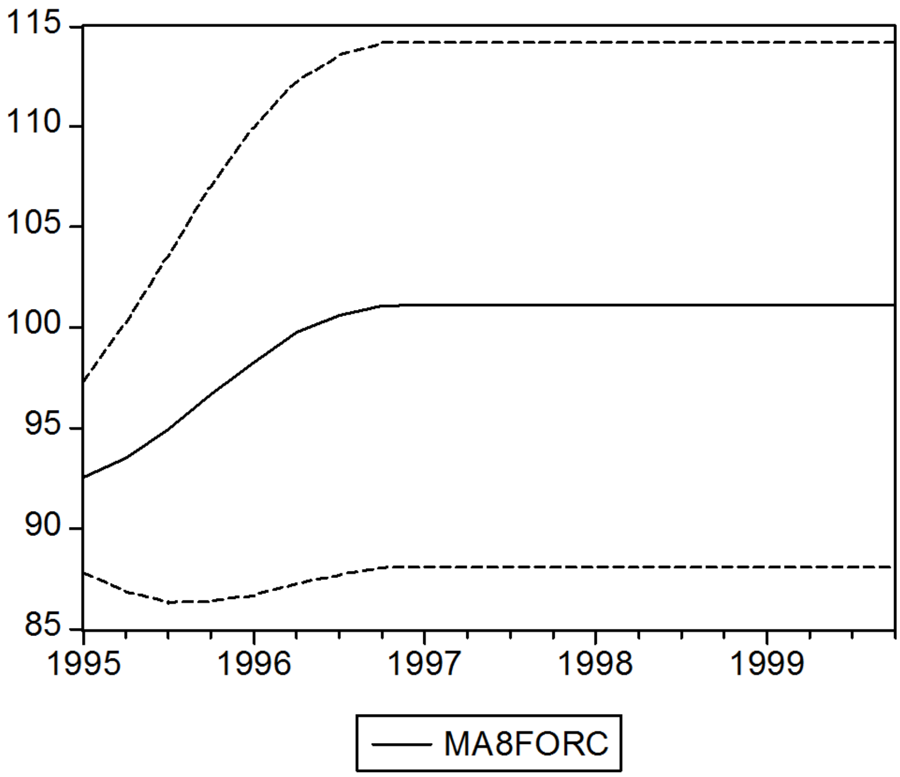

Figure 1: Forecast from *MA(8)* model with 95% confidence bands 1995–1999

notably different from the behaviour of the series in its recent past, in the early *1990s*, and may seem unrealistic.

Figure 3 shows the forecast and 95% confidence bands from the *AR(2)* model. Note that the forecast exhibits much more persistence than that of the *MA(8)* model and has not reverted to the unconditional mean even after 20 quarters. Likewise, the error bands continue to spread out and haven't reached their maximum extent also after 20 quarters.Figure 4 shows the forecast together with historical data for the estimation period: *1963-1994*. It can be seen that the forecast has a lot of persistence and is more closely related to the recent behaviour of the series than the *MA(8)* forecast.

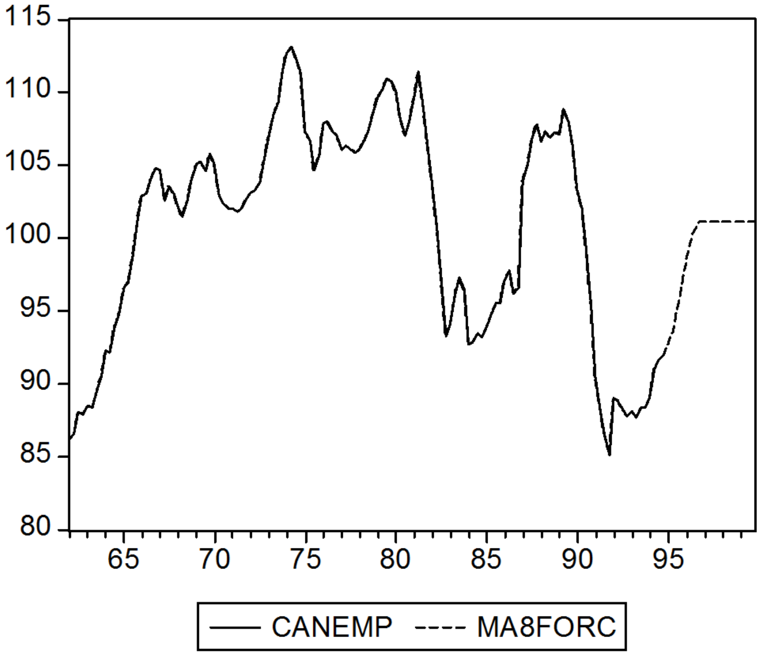

Figure 2: Canadian data: actuals and forecast from *MA(8)* model

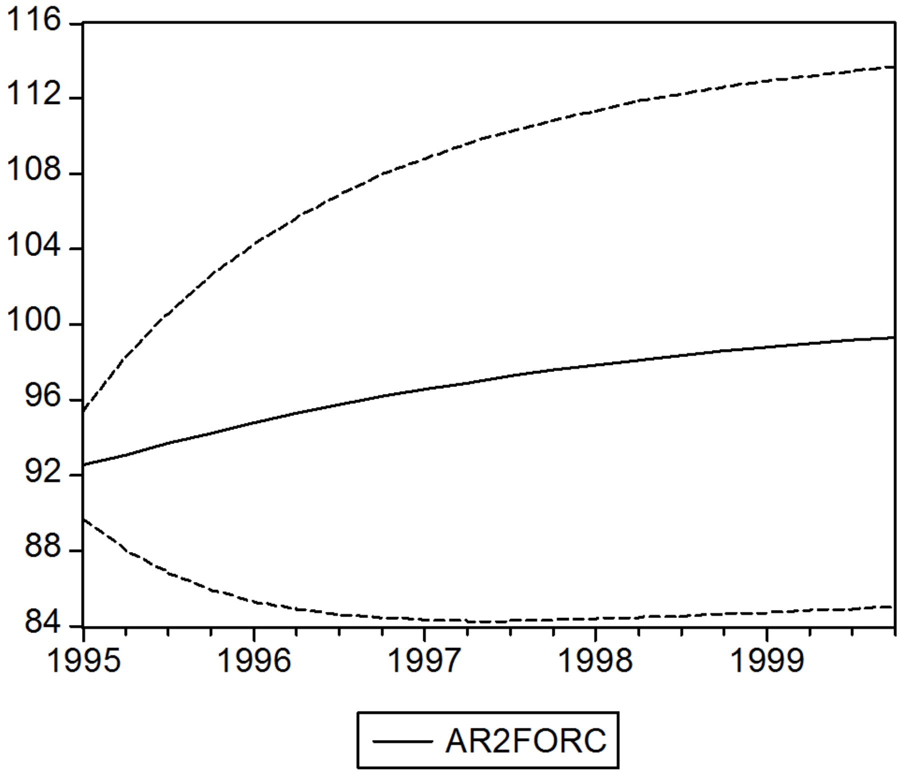

Figure 3: Forecast from *AR(2)* model with 95% confidence bands 1995–1999

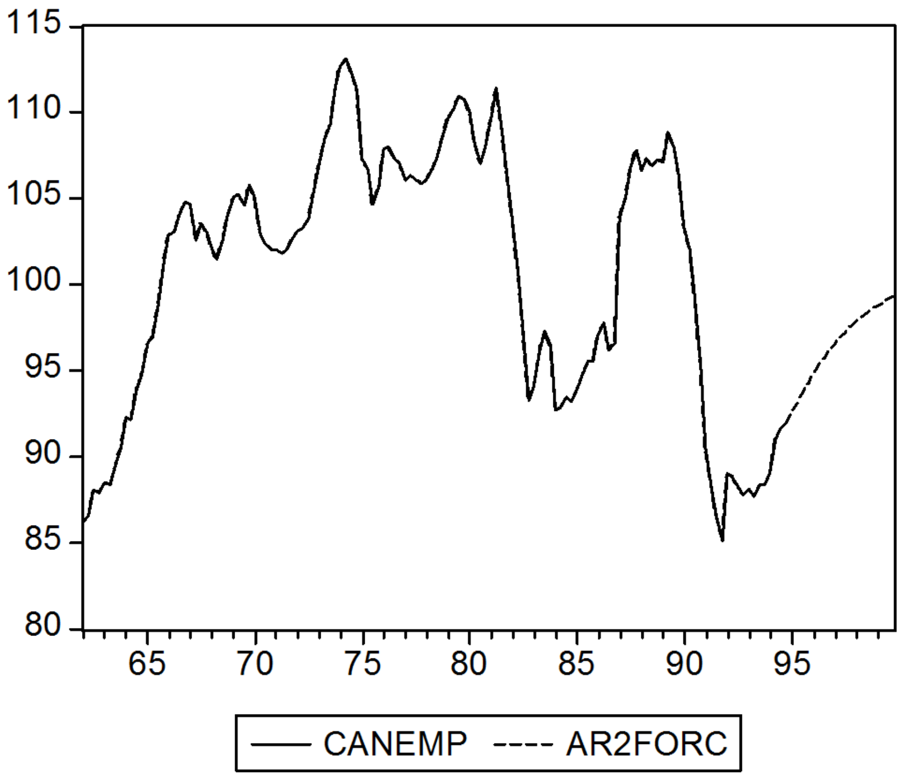

Figure 4: Canadian data: actuals and forecast from the *AR(2)* model