MACQUARIE University

Business Forecasting

Autoregression

Autoregressions

Another useful forecasting model with quasi-explanatory variables is an **autoregression**

Autoregression is a regression where the **independent variables** are **lagged values of the dependent variable**

$$Y_t = f\,(Y_{t-1}, Y_{t-2} \ldots\ldots\ldots\ Y_{t-p})$$

Generally the function is assumed linear $\beta_o + \beta_1$

$$Y_t = \beta_o + \beta_1 * Y_{t-1} + \beta_2 * \ Y_{t-2} + \ldots\ldots + \beta_p\, Y_{t-p}$$

The number of lagged dependent variables used (ie p) is up to the modeller

Rationale for Autoregressions

A typical regression assumption is that **errors** of different observations are **uncorrelated**. This suggests that the Y values of these observations are also uncorrelated

While this is likely to hold for cross-sectional data it is **unlikely to hold for time series**

Y values are likely to be **related over time** eg current interest rates or prices are likely to be highly correlated with interest rates or prices in recent previous periods

Thus **past (Y_{t-i}) values** of these variables may be useful predictors of their **current values (Y_t)**

Further Rationale

The impacts of **explanatory variables** in many cases are **likely to be spread over many time periods**

Current prices, interest rates (X) likely to impact on spending (Y) in the **current period** and also in **future periods**

Hence current spending (Y) is influenced not only by current X but also past values of X

Past values of Y capture some of the influence of the **past values of X**

Including past Y values can **substitute, in part,** for past X in the regression model

Predictive Models

Autoregressions, like trend extrapolations are **quasi-explanatory models**

Time (t) and lagged values of the dependent variable (Y_{t-i}) are likely to be **proxies for other explanatory variables** that may be **numerous, are unobserved or difficult to measure**

Thus autoregressions & trend extrapolations are **more predictive rather than explanatory models**

They can lead to reasonable predictions and, if this is all that is required, then they may suffice as forecasting methods. More for **short to medium term** forecasts

MACQUARIE University

Relationships to Other Methods

$$Y_t = \beta_0 + \beta_1 * Y_{t-1} + \beta_2 * Y_{t-2} + \ldots + \beta_p * Y_{t-p}$$

$$\varepsilon_t$$

The estimated forecast equation for (Y_{t+1}) is

$$Y_{t+1} = b_0 + b_1 * Y_t + b_2 * Y_{t-1} + \ldots + b_p * Y_{t-p+1}$$

If we assume $\mathbf{b_0} = 0$, Y_{t+1} is a **weighted average of past Y values**

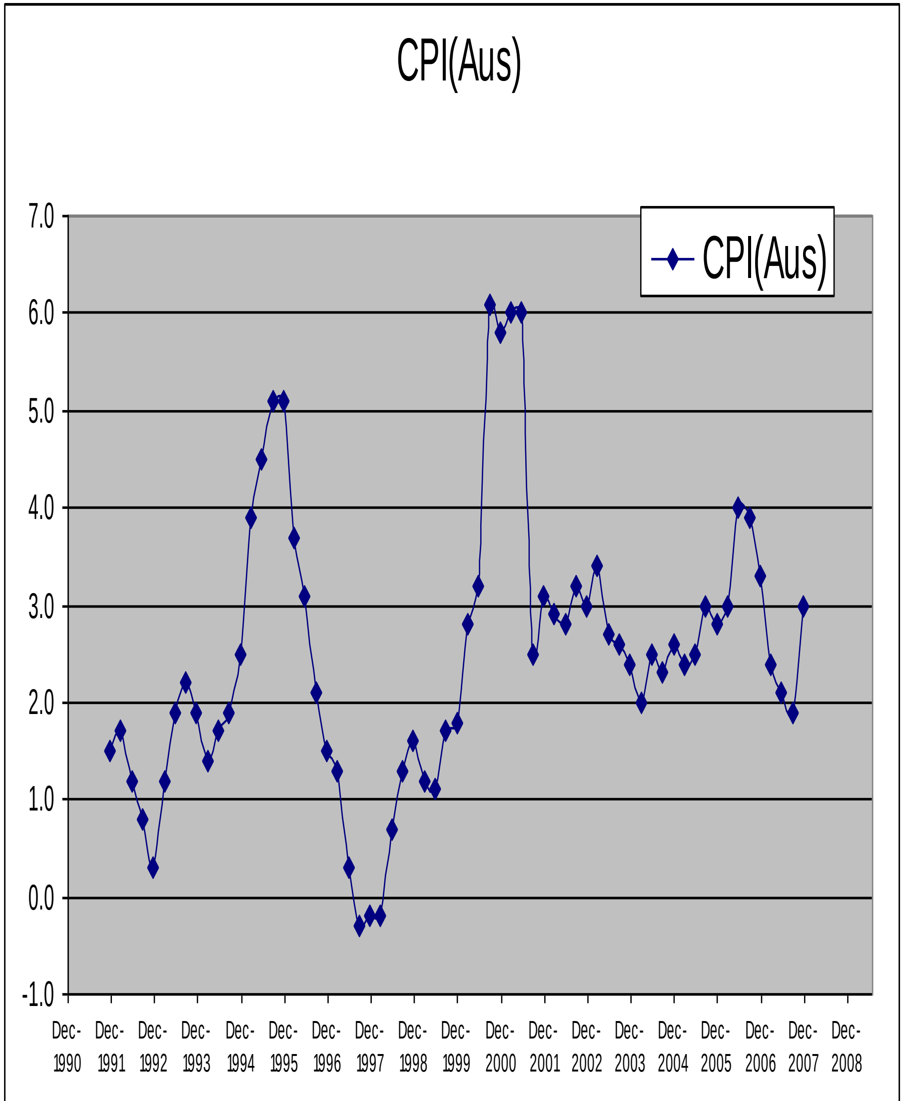
If $b_2, b_3, \ldots, b_p = 0$ then this is a **naïve forecast**

If $b_1, b_2 \ldots b_m = 1/m$ and $b_{m+1}, b_{m+2}, \ldots b_p = 0$ then we have a **moving average of order m**

If $b_1 = \alpha$ ($0 < \alpha < 1$) and $b_{1+i} = (1-\alpha)^i * (\alpha)$ then we have **SES**

Autoregression allows the weights β_i to follow **more flexible patterns than either SES or MA**

Autoregression-Example



Consider CPI (Australia) between 1991(q4) and 2007(q4)

Data looks **generally horizontal** but there are periods of seeming **systematic fluctuation**

Possible Seasonality

(Check ACF and PACF for objective evidence)

CPI values seem **related to near past values of CPI** during these systematic fluctuations

Forecast Model for CPI

Autoregression is a **quick option to forecast CPI** as explanatory variables are likely to be numerous and may not be available

Selection of number of lagged variables to include is not simple

Logically, a model with 6 lagged dependent variables (in this case) should suffice. **Current CPI is unlikely to be affected by observations beyond 6 quarters (1.5 years)**

At least 4 lagged dep vars needed to allow for any **seasonal impacts** (4 quarters - yearly seasonality)

CPI Time Series (Selected Observations)

Quarter	CPI	CPI(-1)	CPI(-2)	CPI(-3)	CPI(-4)	CPI(-5)	CPI(-6)
Dec-1991	1.5	0	0	0	0	0	0
Mar-1992	1.7	1.5	0.0	0.0	0.0	0.0	0.0
Jun-1992	1.2	1.7	1.5	0.0	0.0	0.0	0.0
Sep-1992	0.8	1.2	1.7	1.5	0.0	0.0	0.0
Dec-1992	0.3	0.8	1.2	1.7	1.5	0.0	0.0
Mar-1993	1.2	0.3	0.8	1.2	1.7	1.5	0.0
Jun-1993	1.9	1.2	0.3	0.8	1.2	1.7	1.5
Sep-1993	2.2	1.9	1.2	0.3	0.8	1.2	1.7
Sep-2006	3.9	4.0	3.0	2.8	3.0	2.5	2.4
Dec-2006	3.3	3.9	4.0	3.0	2.8	3.0	2.5
Mar-2007	2.4	3.3	3.9	4.0	3.0	2.8	3.0
Jun-2007	2.1	2.4	3.3	3.9	4.0	3.0	2.8
Sep-2007	1.9	2.1	2.4	3.3	3.9	4.0	3.0
Dec-2007	3.0	1.9	2.1	2.4	3.3	3.9	4.0

Autoregression Estimation Results

Multiple R	0.886051919
R Square	0.785088003
Adjusted R Square	0.762855728
Standard Error	0.706874078
Observations	65

ANOVA

	df	*SS*	*MS*	*F*
Regression	6	105.869238	17.644873	35.31298464
Residual	58	28.9809158	0.49967096	*Significance F*
Total	64	134.8501538		1.26116E-17

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.557030779	0.196179376	2.83939521	0.00622221
CPI(-1)	1.046605234	0.129317959	8.09327059	4.274E-11
CPI(-2)	-0.01102862	0.176147712	-0.0626101	0.950292283
CPI(-3)	-0.18017052	0.168310884	-1.0704627	0.288845623
CPI(-4)	-0.39181654	0.16883985	-2.3206402	0.023843903
CPI(-5)	0.550909566	0.177058857	3.11144879	0.002887435
CPI(-6)	-0.23027878	0.127822106	-1.8015568	0.076813506

Forecasting with Autoregressions

$$E(CPI_t) = 0.55 + 1.04 * CPI_{t-1} - 0.011 * CPI_{t-2} - 0.18 * CPI_{t-3} - 0.39 * CPI_{t-4} + 0.55 * CPI_{t-5} - 0.23 * CPI_{t-6}$$

Thus to Forecast CPI_{t+1}

$$E(CPI_{t+1}) = 0.55 + 1.04 * CPI_{t} - 0.011 * CPI_{t-1} - 0.18 * CPI_{t-2} - 0.39 * CPI_{t-3} + 0.55 * CPI_{t-4} - 0.23 * CPI_{t-5}$$

The values for CPI_t.....CPI_{t-5} are known and can be substituted from the observations

Forecasting Further Ahead

MACQUARIE University

To Forecast CPI_{t+2}

$$E(CPI_{t+2}) = 0.55 + 1.04 * CPI_{t+1} - 0.011 * CPI_t - 0.18 * CPI_{t-1} - 0.39 * CPI_{t-2} + 0.55 * CPI_{t-3} - 0.23 * CPI_{t-4}$$

The values for CPI_t.....CPI_{t-4} are known and can be substituted from the observations **but..... CPI_{t+1} is unknown if we are forecasting at time t**

Substitute $E(CPI_t)$ forecast for unknown CPI_t

The procedure is similar if we wish to forecast CPI_{t+3}. We will need forecasts of CPI_{t+1} and CPI_{t+2}

Thus forecasts are done sequentially

Actual Forecasts

Suppose we wish a forecast for March 08 (t+1)

$$E(CPI_{t+1}) = 0.55 + 1.04 * CPI_t - 0.011 * CPI_{t-1} - 0.18 * CPI_{t-2} - 0.39 * CPI_{t-3} + 0.55 * CPI_{t-4} - 0.23 * CPI_{t-5}$$

$$E(CPI_{Mar08}) = 0.55 + 1.04 * CPI_{Dec07} - 0.011 * CPI_{Sep07} - 0.18 * CPI_{Jun07} - 0.39 * CPI_{Mar07} + 0.55 * CPI_{Dec06} - 0.23 * CPI_{Sep06}$$

$$E(CPI_{Mar08}) = 0.55 + 1.04 * 3.0 - 0.011 * 1.9 - 0.18 * 2.1 - 0.39 * 2.4 + 0.55 * 3.3 - 0.23 * 3.9 = 3.27$$

Forecasting 2 periods ahead

Suppose we wish a forecast for Jun 08 (t+2)

$$E(CPI_{t+2}) = 0.55 + 1.04 * CPI_{t+1} - 0.011 * CPI_{t} - 0.18 * CPI_{t-1} - 0.39 * CPI_{t-2} + 0.55 * CPI_{t-3} - 0.23 * CPI_{t-4}$$

$$E(CPI_{Jun08}) = 0.55 + 1.04 * E(CPI_{Mar08}) - 0.011 * CPI_{Dec07} - 0.18 * CPI_{Sep07} - 0.39 * CPI_{Jun07} + 0.55 * CPI_{Mar07} - 0.23 * CPI_{Dec06}$$

$$E(CPI_{Mar08}) = 0.55 + 1.04 * \underline{3.27} - 0.011 * 3 - 0.18 * 1.9 - 0.39 * 2.1 + 0.55 * 2.4 - 0.23 * 3.3 = 3.34$$