Economic Forecasting

Stationarity and AR, MA and ARMA Processes and Autocorrelation Functions

Stationary and Nonstationary Series

A given time series $\{y_t\}$ is stationary when mean and variance are constant or independent of time.

$$E\left(y_t\right) = \mu \qquad \text{constant mean}$$

$$\operatorname{var}\left(y_t\right) = \sigma^2 \qquad \text{constant variance}$$

$$\operatorname{cov}\left(y_t y_{t-s}\right) = \operatorname{cov}\left(y_t y_{t+s}\right) = \gamma_s \quad \text{time independent covariance}$$

Time series y_t is non-stationary if the mean and variance is not constant or is changing over time.

Many economic variables such as GDP, GDP components, inflation, exchange rates, labour force evolve over time. It is important to check whether these series are stationary or non stationary before any econometric estimation because estimation using non-stationary variables may generate a spurious relationship: reported to have relationship when there is no relationship.

Spurious Regression

Consider a stochastic autoregressive AR(1) series:

$$y_t = \rho y_{t-1} + v_t$$

Application of OLS in this may generate a spurious regression, with a high R^2 and very low Durbin-Watson statistics (R^2>d). OLS generates spurious regression if variables involved are non stationary.

Four ways of checking stationarity

- Partial autocorrelation function and Ljung and Box statistics
- Unit Root Test: Dicky-Fuller Test
- Cointegration Test
- Test for Error Correction

Partial autocorrelation function and Ljung and Box statistics

$$\rho^S = \frac{\mathrm{cov}\left(y_t y_{t+1}\right)}{\mathrm{var}\left(y_t\right)} = \frac{\gamma_s}{\gamma_0}$$

where $\hat{\gamma}_S = \dfrac{\sum\left(y_t - \overline{y}\right)\left(y_{t+s} - \overline{y}\right)}{T}$ $\quad \hat{\gamma}_0 = \dfrac{\sum\left(y_t - \overline{y}\right)^2}{T}$

Use Ljung and Box statistics to test whether all ρ are equal to zero

$$Q = T(T+2)\sum_{s=1}^{m} \frac{\hat{\rho}_s^2}{T-s}$$

This statistic has a χ_m^2 distribution.

Unit Root and Non-stationarity

Like autocorrelated errors consider a stochastic autoregressive AR(1) series:

$$y_t = \rho y_{t-1} + v_t \qquad (4)$$

Here $1 \leq \rho \leq 1$. It is a unit-root process if $\rho = 1$. Then (4) becomes a random walk $y_t = y_{t-1} + v_t$.

Application of OLS in this may generate a spurious regression, with a high R^2 and very low Durbin-Watson statistics (R^2>d). To see how (4) is non stationary , let us assume that

$$v_t \sim N\left(0, \sigma_v^2\right).$$

Iterative Substitution in AR(1) Model

$$y_t = \rho y_{t-1} + v_t$$
$$y_{t-1} = \rho y_{t-2} + v_{t-1}$$
$$y_{t-2} = \rho y_{t-3} + v_{t-3}$$
$$y_{t-3} = \rho y_{t-4} + v_{t-4}$$
$$. \quad \cdots \qquad\qquad \cdots$$
$$y_{t-n} = \rho y_{t-n-1} + v_{t-n} \qquad (5)$$

From substitution

$$y_t = \rho y_{t-1} + v_t =$$

$$y_t = \rho\left[\rho y_{t-2} + v_{t-1}\right] + v_t = \rho^2 y_{t-2} + \rho v_{t-1} + v_t$$

$$y_t = \rho\left[\rho y_{t-3} + v_{t-2}\right] + v_t$$
$$y_t = \rho^2\left[\rho y_{t-3} + v_{t-2}\right] + \rho v_{t-1} + v_t = \rho^3 y_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t$$

AR(1) Time Series as a Function of Past Innovations (Impulses or Shocks)

$$y_t = \rho^n y_{t-n} + \rho^{n-1} v_{t-n-1} + ... + \rho^3 y_{t-3} + \rho^2 v_{t-2} + \rho v_{t}$$

(6)

In the limit the term $\rho^n y_{t-n}$ becomes close to zero as $n \to \infty$.

Rearranging (6) we can write $\{y_t\}$ in terms of current and past values of error terms

$$y_t = v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \rho^3 v_{t-3} + \rho^4 v_{t-4} + ... + \rho^{n-1} v_{t-}$$

(7)

Time Dependent Variance

What is the mean of $\{y_t\}$ in (7)?

$$E(y_t) = E(v_t) + \rho E(v_{t-1}) + \rho^2 E(v_{t-2}) + \rho^3 E(v_{t-3}) + \rho^4 E(v_{t-4}) + \ldots + \rho^{n-1} E(v_{t-n})$$

Because of assumption $v_t \sim N\left(0, \sigma_v^2\right)$; $E(y_t) = 0$

What is the variance of $\{y_t\}$?

$$Var(y_t) = Var\left[(v_t) + \rho(v_{t-1}) + \rho^2(v_{t-2}) + \rho^3(v_{t-3}) + \rho^4(v_{t-4}) + \ldots + \rho^{n-1}(v_{t-n})\right] \text{ if } \rho = 1$$

and

if there is no autocorrelation among the random terms $E(v_t v_{t-1}) = 0$

$$Var(y_t) = \sigma_v^2 + \sigma_v^2 + \sigma_v^2 + \ldots.. + \sigma_v^2 = t.\sigma_v^2$$

Thus the variance of Y term increases with time. This makes this series non stationary.

Rule of thumb : A series is non-stationary if $|\rho| \geq 1$.

A series is stationary if $|\rho| \leq 1$.

Dicky-Fuller and Augmented Dicky-Fuller Tests

$$y_t = \rho y_{t-1} + v_t$$

$$\Delta y_t = (1-\rho) y_{t-1} + v_t \,;$$

Random Walk:

$$\Delta y_t = \gamma y_{t-1} + v_t$$

Random Walk with a drift (intercept):

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + v_t$$

Trend stationary process

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + v_t$$

Augmented Dicky Fuller Test

$$\Delta y_t = \alpha_0 + \alpha_1 t + \gamma y_{t-1} + \sum_{i=1}^{m} a_i \Delta y_{t-i} + v_t$$

Null hypotheses:
There is unit root and time series in non-stationary
K=0 ➔ (1-Ψ)=0

Alternative hypothesis:
There is no unit root and time series is stationary
K<0 ➔ (1-Ψ)<0 ➔ Ψ<1

Unit Root Test of Banana in the first difference in logs

$$Y_t = \rho Y_{t-1} + e_t$$

prodcam_L1: ADF tests (T=39, Constant; 5%=-2.94 1%=-3.61)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
2	-3.181*	0.099137	0.04863	-1.807	0.0794	-5.950	
1	-6.235**	-0.26798	0.05013	2.123	0.0407	-5.912	0.0794
0	-6.060**	0.0077992	0.05246			-5.846	0.0266

prodch_L1: ADF tests (T=39, Constant; 5%=-2.94 1%=-3.61)

D-lag	t-adf	beta Y_1	sigma	t-DY_lag	t-prob	AIC	F-prob
2	-3.564*	0.33520	0.08494	0.2420	0.8102	-4.835	
1	-4.104**	0.35912	0.08382	0.7727	0.4447	-4.884	0.8102
0	-4.426**	0.42529	0.08336			-4.919	0.7282

Prodcam_L1: production in Cambodia in the first difference
Prodch_L1: production in China in the first difference

Co-integration

If two economic variables have long-run equilibrium relationship linear combination of these variables may be stationary even if the individual series may be non stationary. These two variables are said to be co-integrated to each other.

Suppose y_t is consumption and X_t is disposable income.

$$e_t = Y_t - \beta_1 - \beta_2 X_t$$

Even if y_t and X_t are I(1) e_t is I(0).

$$e_t = \alpha_0 + \gamma e_{t-1} + v_t$$

If γ is zero then series e_t is stationary and y_t and X_t are I(1).

Error Correction Model

$$\Delta y_t = \alpha_1 + \alpha_2 \left(Y_t - \beta_1 - \beta_2 X_t \right) + v_t$$

The term in the parenthesis is the error term and the coefficient α_2 governs the speed of adjustment towards long-run equilibrium.

Autoregressive AR(1) Process

$$Y_t = \delta + \theta_1 y_{t-1} + e_t$$

$$E(y_t) = E(y_{t-1}) = ... = E(y_{t-k}) = \mu$$

$$E(Y_y) = E(\delta + \theta_1 y_{t-1} + e_t)$$

$\mu = \delta + \theta_1 \mu$; $\mu = \dfrac{\delta}{1-\theta_1}$; Assume $\delta = 0$ means $E(y) = 0$

$$\text{var}(y_t) = \text{var}(\theta_1 y_{t-1} + e_t) \Rightarrow \sigma_y^2 = \frac{\sigma_e^2}{1-\theta_1^2}$$

$$\text{cov}(Y_t Y_{t-1}) = E(y_t - E(y_t))(y_{t-1} - E(y_t)) = \theta_1 \sigma_y^2$$

Some examples:

$$Y_t = 0.8 y_{t-1} + e_t$$

$$Y_t = -0.8 y_{t-1} + e_t$$

Convergence occurs if $|\theta_1| < 1$. The series is called stationary.

Moving Average-MA(1) Process

$$Y_t = \mu + e_t + \alpha_1 e_{t-1}$$

$$E(y_t) = \mu$$

$$\mathrm{var}(y_t) = \mathrm{var}(\mu + e_t + \alpha_1 e_{t-1}) = \sigma_e^2 (1 + \alpha_1)$$

$$\mathrm{cov}(Y_t Y_{t-1}) = E(y_t - \mu)(y_{t-1} - \mu) = \sigma_e^2 \alpha_1$$

Autocorrelation function: it tapers off after k lags

$$\rho_1 = \frac{\mathrm{cov}(y_t y_{t-1})}{\mathrm{var}(y_t)} = \frac{\alpha_1 \sigma_e^2}{\sigma_e^2 (1 + \alpha_1^2)}$$

Some examples of MA (1) process:

$$Y_t = \mu + e_t + 0.8 e_{t-1}$$

$$Y_t = \mu + e_t - 0.8 e_{t-1}$$

ARMA(1,1) Process: Mean and Variance

$$Y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1}$$

$$\gamma_0 = \text{var}(y_t) = E\left[(y_t - \mu)^2\right] = E\left[\left(\delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1}\right)^2\right]$$

$$= \theta_1^2 \gamma_0 + \sigma_e^2 + \alpha_1^2 \sigma_e^2 + 2\theta_1 \alpha_1 E\left[y_{t-1} e_{t-1}\right]$$

$$E\left[y_{t-1} e_{t-1}\right] = E\left(\delta + \theta_1 y_{t-2} + e_{t-1} + \alpha_1 e_{t-2}\right) e_{t-1} = E\left(e_{t-1}\right)^2 = \sigma_e^2$$

$$\gamma_0 = \frac{\left(1 + \alpha_1^2 + 2\theta_1 \alpha_1\right)}{1 - \theta_1^2} \sigma_e^2$$

ARMA(1,1) Process: Covariance

$$E\left[y_{t-1}y_t\right]=E\left[\left(\delta+\theta_1 y_{t-1}+e_t+\alpha_1 e_{t-1}\right)y_{t-1}\right]=\theta_1\gamma_0+\alpha\sigma_e^2$$

$$E\left[y_{t-1}y_t\right]=E\left[\theta_1\left(y_{t-1}y_{t-1}\right)\right]+E\alpha_1\left[y_{t-1}e_{t-1}\right]=\theta_1\frac{\left(1+\alpha_1^2+2\theta_1\alpha_1\right)}{1-\theta_1^2}\sigma_e^2+\alpha_1\sigma_e^2$$

$$E\left[y_{t-1}y_t\right]=\theta_1\frac{\left(1+\alpha_1^2+2\theta_1\alpha_1\right)}{1-\theta_1^2}\sigma_e^2+\alpha_1\sigma_e^2=\frac{\theta_1\left(1+\alpha_1^2+2\theta_1\alpha_1\right)+\alpha_1-\alpha_1\theta_1^2}{1-\theta_1^2}\sigma_e^2$$

$$\gamma_1=\frac{\theta_1+\alpha_1^2\theta_1+2\theta_1^2\alpha_1+\alpha_1-\alpha_1\theta_1^2}{1-\theta_1^2}\sigma_e^2=\frac{\left(\theta_1+\alpha_1^2\theta_1+\theta_1^2\alpha_1+\alpha_1\right)}{1-\theta_1^2}\sigma_e^2=\frac{\left(\theta_1+\alpha_1\right)\left(1+\theta_1\alpha_1\right)}{1-\theta_1^2}\sigma_e^2$$

$$\gamma_k=\theta_1\gamma_{k-1}\qquad k\geq 2$$

ARMA(1,1) Process: Autocorrelation Coefficients

$$\rho_1 = \frac{\operatorname{cov}\left(y_t y_{t-1}\right)}{\operatorname{var}\left(y_t\right)} = \frac{\gamma_1}{\gamma_0} = \frac{\left(1+\alpha_1^2+2\theta_1\alpha_1\right)}{\left(\theta_1+\alpha_1\right)\left(1+\theta_1\alpha_1\right)}$$

$$\rho_k = \theta_1 \rho_{k-1}$$

This process converges when $0 \le \rho_i \le 1$

Partial autocorrelation function and Ljung and Box statistics

$$\rho^S = \frac{\operatorname{cov}(y_t y_{t+1})}{\operatorname{var}(y_t)} = \frac{\gamma_s}{\gamma_0}$$

where $\hat{\gamma}_S = \frac{\sum (y_t - \bar{y})(y_{t+s} - \bar{y})}{T}$ $\hat{\gamma}_0 = \frac{\sum (y_t - \bar{y})^2}{T}$

Use Ljung and Box statistics to test whether all ρ are equal to zero

$$Q = T(T+2)\sum_{s=1}^{m} \frac{\hat{\rho}_s^2}{T-s}$$

This statistic has a χ_m^2 distribution.

AR(2) Process: Mean

$$Y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-1} + e_t$$

$$E(y_t) = E\left(y_{t-1}\right) = \ldots.. = E\left(y_{t-k}\right) = \mu$$

$$E(Y_y) = E\left(\delta + \theta_1 y_{t-1} + \theta_2 y_{t-1} + e_t\right)$$

$$\mu = \delta + \theta_1 \mu + \theta_2 \mu \qquad \mu = \frac{\delta}{1 - \theta_1 - \theta_2}$$

AR(2) Process: Variance and Covariance

$$\text{var}(y_t) = E\left[\left(\theta_1 y_{t-1} + \theta_2 y_{t-2} + e_t\right) y_t\right]$$

$$\gamma_0 = \theta_1 \gamma_1 + \theta_2 \gamma_2 + \sigma_e^2$$

$$\text{cov}(Y_t Y_{t-k}) = \gamma_k = E\left[\left(\theta_1 y_{t-1} + \theta_2 y_{t-2} + e_t\right) y_{t-k}\right] = \theta_1 \gamma_{k-1} + \theta_2 \gamma_{k-2}$$

$$\gamma_k = \theta_1 \gamma_{k-1} + \theta_2 \gamma_{k-2}$$

Dividing both sides by γ_0

$$\rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}$$

AR(2) Process: Partial autocorrelation

$$\rho_k = \theta_1 \rho_{k-1} + \theta_2 \rho_{k-2}$$

$$\rho_k = \rho_{-k}$$

$$\rho_0 = 1$$

$$\rho_1 = \frac{\theta_2}{(1-\theta_1)}$$

$$\rho_2 = \theta_2 + \frac{\theta_1^2}{(1-\theta_2)}$$

MA(2) Process

$$Y_t = \mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}$$

$$E(y_t) = \mu$$

$$\operatorname{var}(y_t) = \operatorname{var}(\mu + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}) = \sigma_e^2 \left(1 + \alpha_1^2 + \alpha_2^2\right)$$

$$\operatorname{cov}(Y_t Y_{t-1}) = E(y_t - \mu)(y_{t-1} - \mu) = \sigma_e^2 \left(\alpha_1 + \alpha_1 \alpha_2\right)$$

$$\operatorname{cov}(Y_t Y_{t-2}) = E(y_t - \mu)(y_{t-2} - \mu) = \alpha_2 \sigma_e^2$$

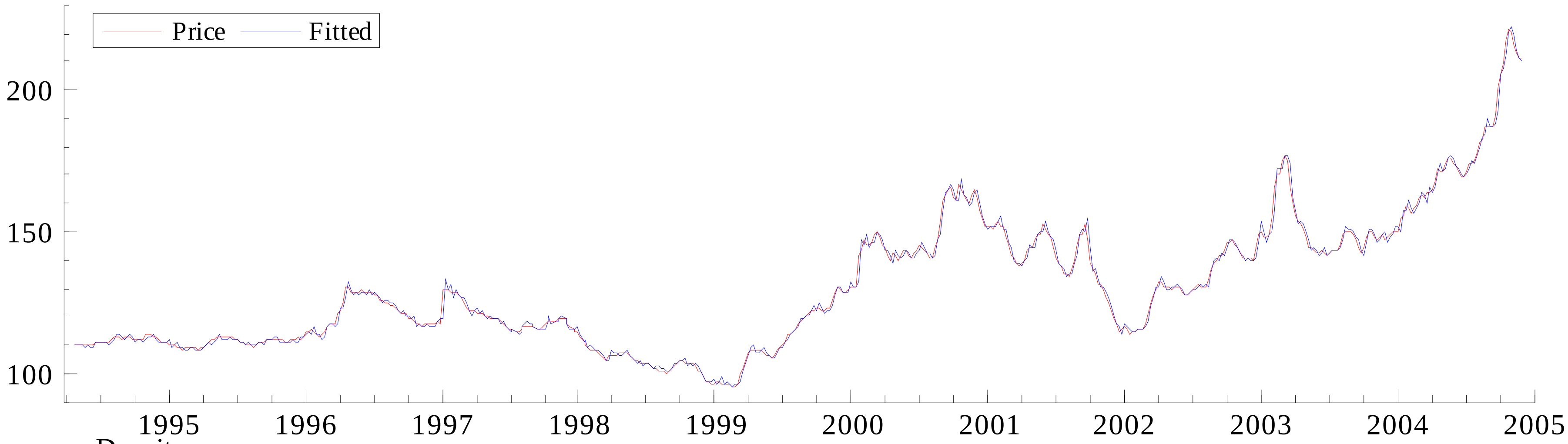
$$\operatorname{cov}(Y_t Y_{t-3}) = 0$$

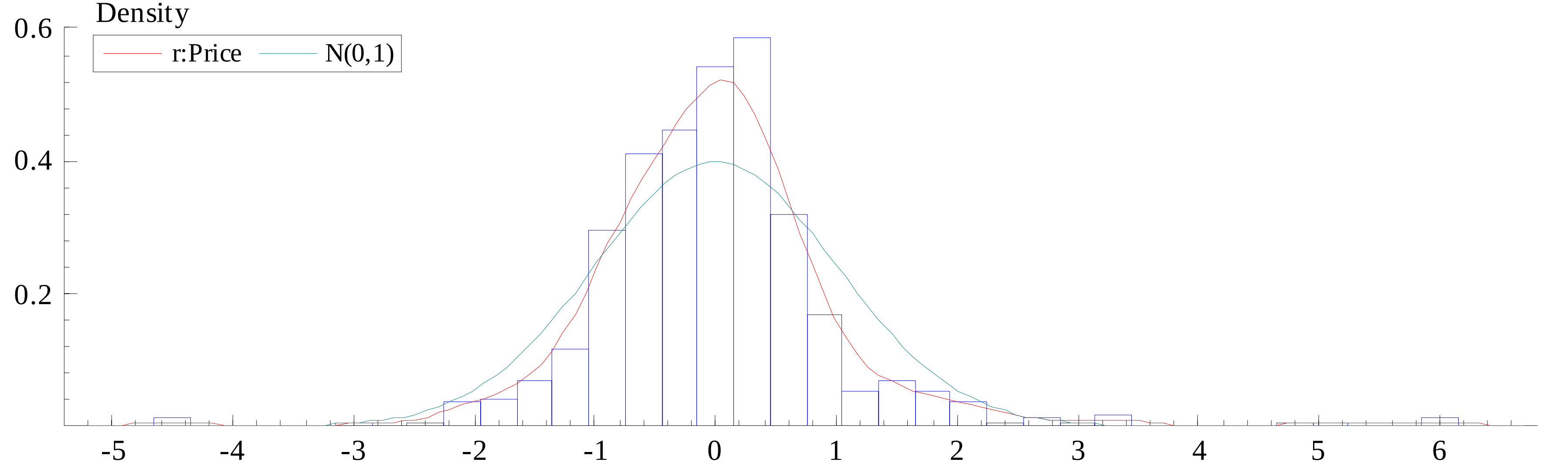
$$\rho_1 = \frac{\operatorname{cov}\left(y_t y_{t-1}\right)}{\operatorname{var}(y_t)} = \frac{\alpha_1\left(1 + \alpha_2\right)}{\left(1 + \alpha_1^2 + \alpha_2^2\right)}$$

$$\rho_2 = \frac{\operatorname{cov}\left(y_t y_{t-2}\right)}{\operatorname{var}(y_t)} = \frac{\alpha_1}{\left(1 + \alpha_1^2 + \alpha_2^2\right)} \quad ; \quad \rho_k = 0$$

MA(2) process has tow period long memory.

Weakly Prices of Diesel in the US AR(6) Model: PcGive Graphics





Weakly Prices of Diesel in the US

Data source: Economagic.com

PcGive AR(6):Forecast

Commands for ARMA (2,2) forecasting model

Arima C / NAR=2 NMA=2 Predict=predc plotac plotpac acf=cacf

plot predc c year /gnu lineonly

dim alpha 3
gen1 alpha:1=0.5
gen1 alpha:2=-0.2
gen1 alpha:3=100

arima c /NAR=1 NMA=1 coef=beta start=alpha
gen1 S=sqrt($sig2)
arima c/NAR=1 NMA=1 coef=beta fbeg=26 fend=30 sigma=s gnu
arima c/NAR=1 NMA=1 coef=beta fbeg=26 fend=35 sigma=s gnu

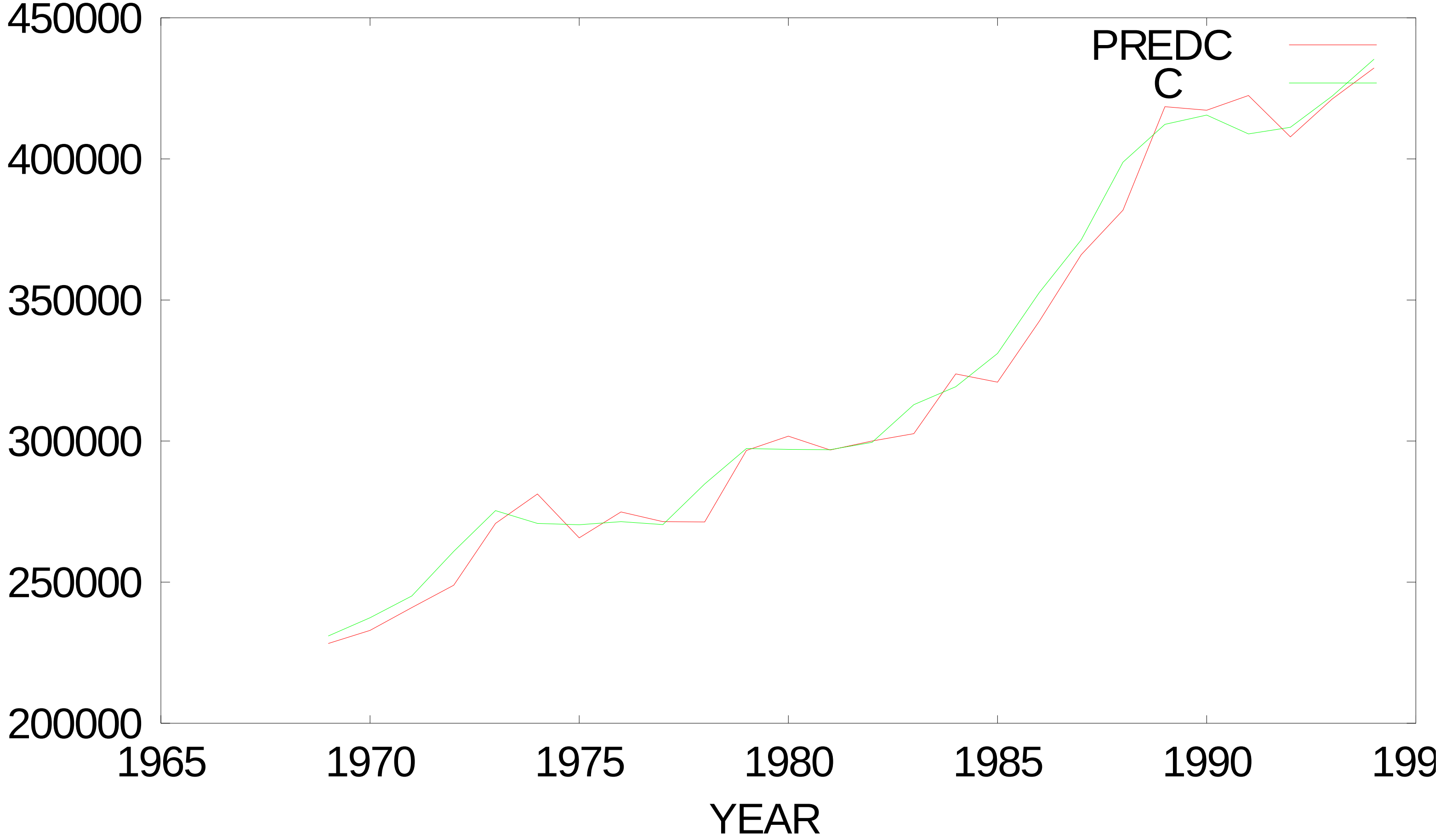
arima c/NAR=1 NMA=1 coef=beta fbeg=26 fend=40 sigma=s gnu

arima c/NAR=1 NMA=1 coef=beta fbeg=26 fend=50 sigma=s gnu

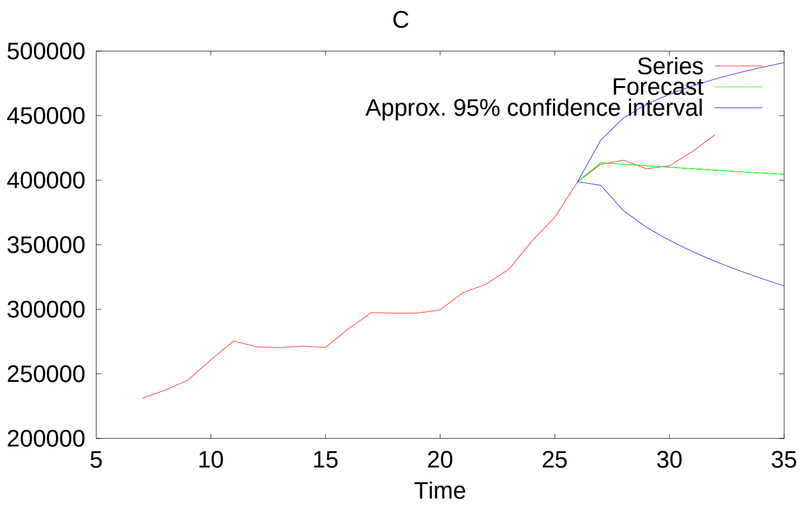
arima c/NAR=1 NMA=1 coef=beta fbeg=26 fend=100 sigma=s gnu

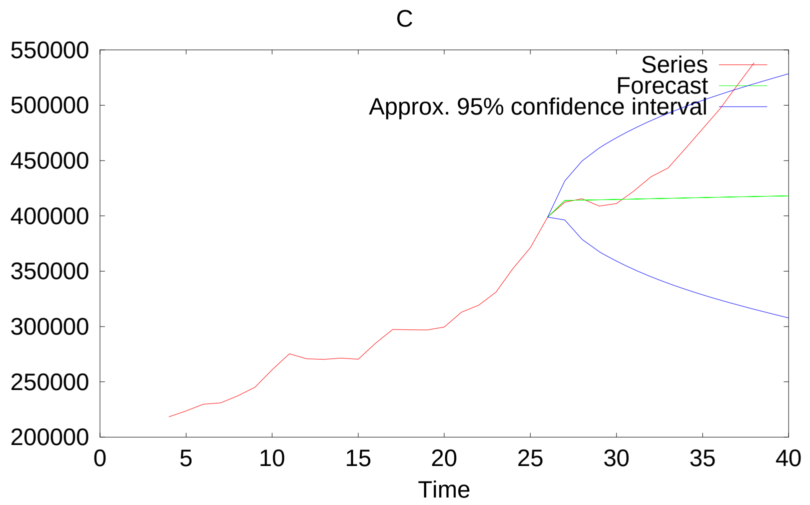
Plot of predected and actual investment series. It is a very good prediction.

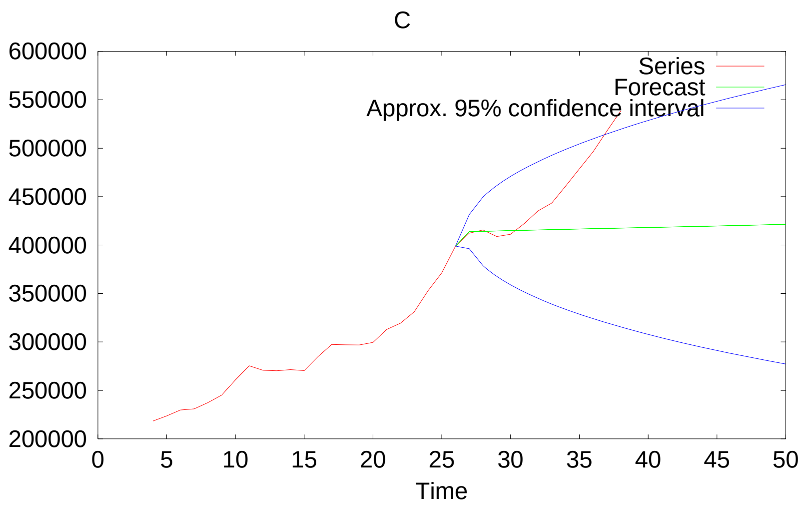
Prediction with an ARIMA Model

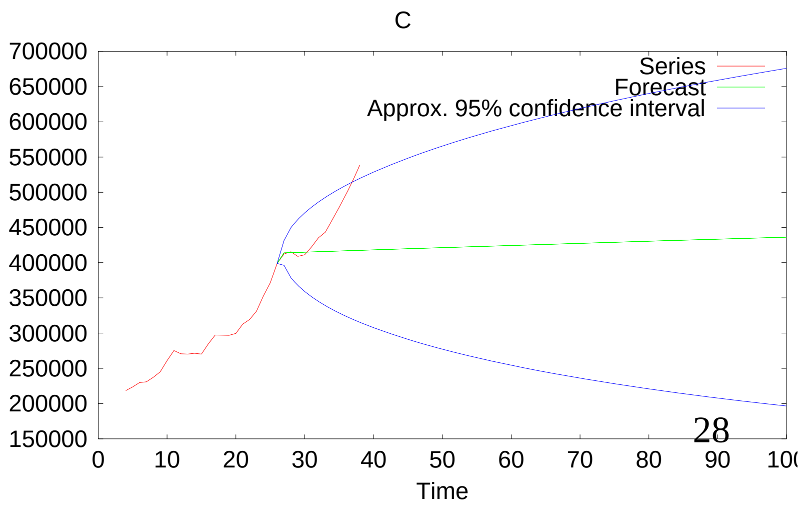


Forecast from an ARIMA Model









Prediction with AR or MA or ARMA Processes