Prediction Error Method With Momentum

Chtiba Reda

2022.05.19

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \qquad (1,1)$$

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \qquad (1,1)$$

Where y(t) is the output , e(t),u(t) is the input and A,B,C are

Lag-Polynomials

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \qquad (1,1)$$

Where y(t) is the output , e(t),u(t) is the input and A,B,C are Lag-Polynomials

$$A\left(q^{-1}\right) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$$

$$B\left(q^{-1}\right) = 1 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$

$$C\left(q^{-1}\right) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}$$

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \qquad (1,1)$$

Where y(t) is the output , e(t),u(t) is the input and A,B,C are

Lag-Polynomials

$$A\left(q^{-1}\right) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$$

$$B\left(q^{-1}\right) = 1 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$

$$C\left(q^{-1}\right) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}$$

The Parameter vector here is

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \qquad (1,1)$$

Where y(t) is the output , e(t),u(t) is the input and A,B,C are

Lag-Polynomials

$$A\left(q^{-1}\right) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$$

$$B\left(q^{-1}\right) = 1 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$

$$C\left(q^{-1}\right) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}$$

The Parameter vector here is

$\theta = [a_1, ..., a_{n_a}, b_1, ..., b_{n_b}, c_1, ..., c_{n_c}]^T$.

The ARMAX Model , is used to describe certain dynamics that have a linear input-output mechanism. It can be given in this form:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t) \qquad (1,1)$$

Where y(t) is the output , e(t),u(t) is the input and A,B,C are

Lag-Polynomials
$$A\left(q^{-1}\right) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$$
$$B\left(q^{-1}\right) = 1 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$
$$C\left(q^{-1}\right) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}$$

The Parameter vector here is
$\theta = [a_1, ..., a_{n_a}, b_1, ..., b_{n_b}, c_1, ..., c_{n_c}]^T$.
Objective : Estimation of θ

PEM

Components of the PEM

PEM

Components of the PEM

•Model choice: ARMAX, where the shocks e(t) are W.N with constant variance.

PEM

Components of the PEM

•Model choice: ARMAX, where the shocks e(t) are W.N with constant variance.

•Predictor and Prediction Error:

PEM

Components of the PEM

•Model choice: ARMAX, where the shocks e(t) are W.N with constant variance.

•Predictor and Prediction Error:

$$\hat{y}(t|t-1,\theta) = \left(1 - \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)}\right)y(t) + \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)}u(t)$$

PEM

Components of the PEM

•Model choice: ARMAX, where the shocks e(t) are W.N with constant variance.

•Predictor and Prediction Error:

$$\hat{y}(t|t-1,\theta) = \left(1 - \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)}\right)y(t) + \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)}u(t)$$

$$\epsilon(t,\theta) = \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)}y(t) - \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)}u(t)$$

PEM

Components of the PEM

•Model choice: ARMAX, where the shocks e(t) are W.N with constant variance.

•Predictor and Prediction Error:

$$\hat{y}(t|t-1,\theta) = \left(1 - \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)}\right) y(t) + \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)} u(t)$$

$$\epsilon(t,\theta) = \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)} y(t) - \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)} u(t)$$

•Criterion:

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \epsilon(t,\theta)^2$$

PEM

Components of the PEM

•Model choice: ARMAX, where the shocks e(t) are W.N with constant variance.

•Predictor and Prediction Error:

$$\hat{y}(t|t-1,\theta) = \left(1 - \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)}\right)y(t) + \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)}u(t)$$

$$\epsilon(t,\theta) = \frac{A(q^{-1},\theta)}{C(q^{-1},\theta)}y(t) - \frac{B(q^{-1},\theta)}{C(q^{-1},\theta)}u(t)$$

•Criterion:

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\epsilon(t,\theta)^2$$

•Constraint set on θ :

$$\mathcal{D} = \{\theta | B(0) = 0, C(0) = 1, |c_i| < 1 \quad \forall i = 2, ..., n_c\}$$

Statistical Properties

•Consistency :

Statistical Properties

•Consistency : If u(t) and y(t) are stationary, that is $|a_i| < 1 \quad \forall i = 2, ..., n_a$ and $|b_i| < 1 \quad \forall i = 2, ..., n_b$.

Statistical Properties

•Consistency : If u(t) and y(t) are stationary, that is $|a_i| \ < 1 \quad \forall i = 2, ..., n_a$ and $|b_i| \ < 1 \quad \forall i = 2, ..., n_b$.Then

Statistical Properties

•Consistency : If u(t) and y(t) are stationary, that is $|a_i| < 1 \quad \forall i = 2, ..., n_a$ and $|b_i| < 1 \quad \forall i = 2, ..., n_b$.Then The PEM Estimate $\hat{\theta}$ is consistent.

Statistical Properties

•Consistency : If u(t) and y(t) are stationary, that is $|a_i| < 1 \quad \forall i = 2, ..., n_a$ and $|b_i| < 1 \quad \forall i = 2, ..., n_b$.Then The PEM Estimate $\hat{\theta}$ is consistent.

•Asymptotic Distribution: If the Hessian of the Criterion is non-singular, Then:

Statistical Properties

•Consistency : If u(t) and y(t) are stationary, that is $|a_i| < 1 \quad \forall i = 2, ..., n_a$ and $|b_i| < 1 \quad \forall i = 2, ..., n_b$.Then The PEM Estimate $\hat{\theta}$ is consistent.

•Asymptotic Distribution: If the Hessian of the Criterion is non-singular, Then:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{dist} \mathcal{N}(0, P)$$

$$\text{Where} \quad P = \Lambda \left[\mathrm{E}[\psi(t, \theta_0)\psi(t, \theta_0)^T] \right]$$

Such that , Λ is the Covariance Matrix associated with e(t), θ_0 is the true parameter and ψ is the operator acting on $\epsilon(t, \theta)$ given by $\psi(t, \theta) = -\left[\frac{\partial \epsilon(t,\theta)}{\partial \theta}\right]^T$

Algorithm

$\epsilon(t, \theta)$ does not depend on θ linearly. $\Rightarrow$

Algorithm

$\epsilon(t,\theta)$ does not depend on θ linearly. $\Rightarrow$ $\hat{\theta} = \arg\min_{\theta \in \mathcal{D}} V_N(\theta)$ does not have an analytical form.$\Rightarrow$

Algorithm

$\epsilon(t,\theta)$ does not depend on θ linearly. $\Rightarrow$ $\hat{\theta} = \arg\min_{\theta \in \mathcal{D}} V_N(\theta)$ does not have an analytical form. $\Rightarrow$ Approximate $\hat{\theta}$ using Newton-Raphson Method $\Rightarrow$

Algorithm

$\epsilon(t,\theta)$ does not depend on θ linearly. $\Rightarrow \hat{\theta} = \arg\min_{\theta \in \mathcal{D}} V_N(\theta)$ does not have an analytical form.$\Rightarrow$ Approximate $\hat{\theta}$ using Newton-Raphson Method $\Rightarrow$ Not efficient : Computing the Hessian is cumbersome.

Algorithm

$\epsilon(t,\theta)$ does not depend on θ linearly. $\Rightarrow \hat{\theta} = \arg\min_{\theta \in \mathcal{D}} V_N(\theta)$ does not have an analytical form.$\Rightarrow$ Approximate $\hat{\theta}$ using Newton-Raphson Method $\Rightarrow$ Not efficient : Computing the Hessian is cumbersome.
However, if N is large $\Rightarrow$

Algorithm

$\epsilon(t, \theta)$ does not depend on θ linearly. $\Rightarrow \hat{\theta} = \arg\min_{\theta \in \mathcal{D}} V_N(\theta)$ does not have an analytical form.$\Rightarrow$ Approximate $\hat{\theta}$ using Newton-Raphson Method $\Rightarrow$ Not efficient : Computing the Hessian is cumbersome.
However, if N is large $\Rightarrow$ Approximate $\hat{\theta}$ using Gauss-Newton Algorithm.

Algorithm

$\epsilon(t,\theta)$ does not depend on θ linearly. $\Rightarrow$ $\hat{\theta} = \underset{\theta \in \mathcal{D}}{\arg\min}\, V_N(\theta)$ does not have an analytical form.$\Rightarrow$ Approximate $\hat{\theta}$ using Newton-Raphson Method $\Rightarrow$ Not efficient : Computing the Hessian is cumbersome.
However, if N is large $\Rightarrow$ Approximate $\hat{\theta}$ using Gauss-Newton Algorithm.

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \alpha_k \left[\frac{2}{N}\sum_{t=1}^{N} \psi(t,\theta_k)\psi(t,\theta_k)^T\right]^{-1} \left[\frac{2}{N}\sum_{t=1}^{N} \epsilon(t,\theta_k)\psi(t,\theta_k)\right]$$

Where α_k is a gain coefficient.

Experimenting with the PEM on a simulated ARMAX(1,1,1,1)

We simulate N=1000 output from true system:

Experimenting with the PEM on a simulated ARMAX(1,1,1,1)

We simulate N=1000 output from true system:

$$\left(1-0.8q^{-1}\right)y(t)=\left(0.7q^{-1}\right)u(t)+\left(1+0.8q^{-1}\right)e(t)$$

Experimenting with the PEM on a simulated ARMAX(1,1,1,1)

We simulate N=1000 output from true system:

$$\left(1-0.8q^{-1}\right)y(t)=\left(0.7q^{-1}\right)u(t)+\left(1+0.8q^{-1}\right)e(t)$$

Where the input signal u(t) is W-N,taking value ± 1 with one delay, independent from the W-N e(t) with variance $\lambda^2 = 1$

Experimenting with the PEM on a simulated ARMAX(1,1,1,1)

We simulate N=1000 output from true system:

$$\left(1 - 0.8q^{-1}\right) y(t) = \left(0.7q^{-1}\right) u(t) + \left(1 + 0.8q^{-1}\right) e(t)$$

Where the input signal u(t) is W-N,taking value ± 1 with one delay, independent from the W-N e(t) with variance $\lambda^2 = 1$
After Simulation, we store the data.

Experimenting with the PEM on a simulated ARMAX(1,1,1,1)

We simulate N=1000 output from true system:

$$\left(1 - 0.8q^{-1}\right) y(t) = \left(0.7q^{-1}\right) u(t) + \left(1 + 0.8q^{-1}\right) e(t)$$

Where the input signal u(t) is W-N,taking value ± 1 with one delay, independent from the W-N e(t) with variance $\lambda^2 = 1$ After Simulation, we store the data.Then we would like to apply the PEM on the data to see whether the estimated $\hat{\theta}$ is close to the true one or not .

Experimenting with the PEM on a simulated ARMAX(1,1,1,1)

We simulate N=1000 output from true system:

$$\left(1-0.8q^{-1}\right)y(t)=\left(0.7q^{-1}\right)u(t)+\left(1+0.8q^{-1}\right)e(t)$$

Where the input signal u(t) is W-N,taking value ± 1 with one delay, independent from the W-N e(t) with variance $\lambda^2 = 1$
After Simulation, we store the data.Then we would like to apply the PEM on the data to see whether the estimated $\hat{\theta}$ is close to the true one or not .
Below, is a figure that shows the evolution of the estimated parameters in 70 iterations.

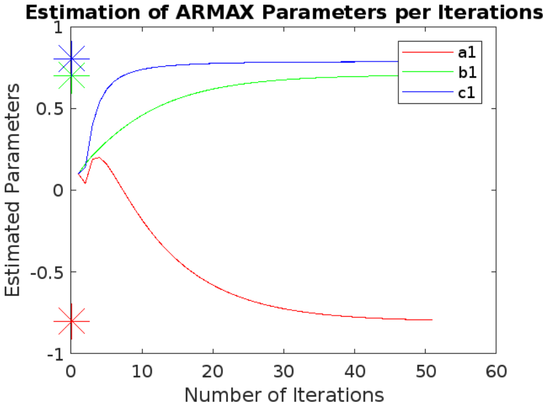

Figure: PEM Estimate

PEMM

We would like to think of a way to improve this algorithm,

PEMM

We would like to think of a way to improve this algorithm,a natural way to do so is to add the Momentum part to the Gauss-Newton Algorithm.

PEMM

We would like to think of a way to improve this algorithm,a natural way to do so is to add the Momentum part to the Gauss-Newton Algorithm.

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \alpha_k \left[\frac{2}{N}\sum_{t=1}^{N}\psi(t,\theta_k)\psi(t,\theta_k)^T\right]^{-1}\left[\frac{2}{N}\sum_{t=1}^{N}\epsilon(t,\theta_k)\psi(t,\theta_k)\right]$$

$$+\beta_k\left[\hat{\theta}_k - \hat{\theta}_{k-1}\right]$$

Where β_k is a gain coefficient.

PEMM

We would like to think of a way to improve this algorithm,a natural way to do so is to add the Momentum part to the Gauss-Newton Algorithm.

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \alpha_k \Big[\frac{2}{N}\sum_{t=1}^{N}\psi(t,\theta_k)\psi(t,\theta_k)^T\Big]^{-1}\Big[\frac{2}{N}\sum_{t=1}^{N}\epsilon(t,\theta_k)\psi(t,\theta_k)\Big]$$

$$+\beta_k\big[\hat{\theta}_k - \hat{\theta}_{k-1}\big]$$

Where β_k is a gain coefficient.The table below, shows some numerical experiments using the PEMM, for various values of β_k and similar conditions to what we had for the PEM experiment.

Table: Comparing PEM with Momentum for different β_k

β_k	$\hat{\theta}_{70}$
$\geq 10^{-3}$	NAN
$\leq 10^{-3}$	$\approx [-0.8, 0.7, 0.8]$