ARMAX for System Identification

Dr. Robert Oates

(Based on notes written by Prof. William Harwin)

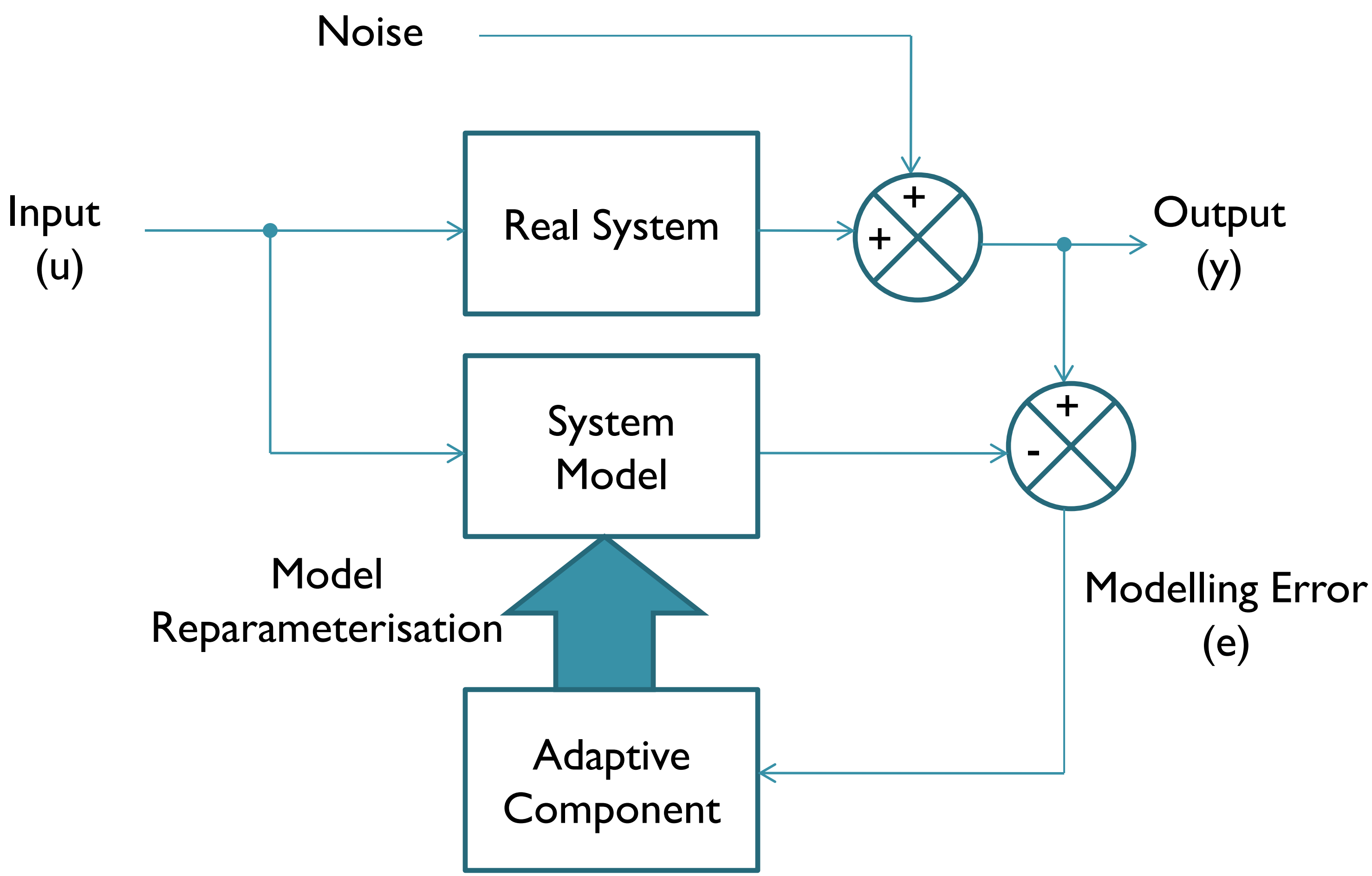
Summary

- Systems Identification
- Least Squares Linear Regression
- Example – How many Harrys are Pottering?
- ARMAX Models and Least Squares Linear Regression
- Common ARMAX Variants

WARNING – MATHS!

- I tend to avoid using equations in presentations
- It is totally unavoidable here

Systems Identification and Modelling



Systems Identification and State Estimation

- There is an obvious connection between
  - Identifying unknown parameters that relate input to output (Systems ID)
    - ARMAX
    - ARMA
    - ARX
  - Identifying unknown parameters that represent a system's state (State Estimation)
    - Particle Filters
    - Kalman Filters

Systems Identification and State Estimation

- In fact both sets of algorithms use Gauss' **Least Squares Linear Regression**

- This is an algorithm designed to minimise the errors for a given function

Least Squares Linear Regression

- Notation
  - $\mathbf{y}$ – The matrix of observed outputs of a system
  - $\mathbf{U}$ – The inputs to the system
  - $\boldsymbol{\theta}$ – The system model parameters
  - $\hat{\mathbf{y}}$ – The estimate of y based on a system model and the inputs to the system

Least Squares Linear Regression

- Calculating the estimate of the output

$$\hat{\mathbf{y}} = \mathbf{U}\boldsymbol{\theta}$$

Least Squares Linear Regression

- Calculating the model error

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$$

- Calculating the sum-of-squares error (SSE)

$$e_{SSE} = \sum_{i=1}^{n} e_i^2$$

$$e_{SSE} = \mathbf{e}^{\mathbf{T}}\mathbf{e}$$

Least Squares Linear Regression

- Substituting in the definition of **e**

$$e_{SSE} = (\mathbf{y} - \hat{\mathbf{y}})^{\mathrm{T}}(\mathbf{y} - \hat{\mathbf{y}})$$

- Substituting in the definition of $\hat{\mathbf{y}}$ and transposing the 1st bracket

$$e_{SSE} = (\mathbf{y}^{\mathrm{T}} - \boldsymbol{\theta}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}})(\mathbf{y} - \boldsymbol{\theta}\mathbf{U})$$

Least Squares Linear Regression

- Expanding and rearranging

$$e_{SSE} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{U}\left(\mathbf{U}^{\mathrm{T}}\mathbf{U}\right)^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}$$

$$+\left(\boldsymbol{\theta} - \left(\mathbf{U}^{\mathrm{T}}\mathbf{U}\right)^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}\right)^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\left(\boldsymbol{\theta} - \left(\mathbf{U}^{\mathrm{T}}\mathbf{U}\right)^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}\right)$$

Least Squares Linear Regression

- Expanding and rearranging

$$e_{SSE} = \mathbf{K} + \mathbf{x}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{x}$$

$$\mathbf{K} = \mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}$$

$$\mathbf{x} = \boldsymbol{\theta} - (\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}$$

Least Squares Linear Regression

- But wait!
- If $\mathbf{x} = \boldsymbol{\theta} - (\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}$

Is the only thing we can change, the smallest possible value of error is when $\mathbf{x} = 0$ – i.e.

$$\boldsymbol{\theta} = (\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}$$

LSE Example
Mr Gauss, meet Mr Potter

- The National Office of Statistics keeps track of all the baby names for certain years (1998, 1999, 2007,2008, 2009)

- In 1997 “*Harry Potter and the Philosopher’s Stone*” was released

- Assuming that this causes an exponential growth in the number of “Harrys” can we use the data to predict 2009’s result?

Building the Model

- Let's assume that the number is growing exponentially and use the following model

$$\mathbf{y} = e^{\mathbf{U}\boldsymbol{\theta}}$$

$$\log(\mathbf{y}) = \mathbf{U}\boldsymbol{\theta}$$

- Let's also use two model parameters, making $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = \begin{bmatrix} b_1 & b_0 \end{bmatrix}$$

Building the Model

- As the dimensionality of **U** and $\boldsymbol{\theta}$ has to match, we pad **U** with 1s

$$\mathbf{U} = \begin{bmatrix} 1998 & 1 \\ 1999 & 1 \\ 2007 & 1 \\ 2008 & 1 \end{bmatrix}$$

Building the Model

- To use the standard formulation of the LSE we'll linearise the data by taking logs

$$\mathbf{y}^{'} = \log(\mathbf{y}) = \log\left(\begin{bmatrix} 4761 \\ 4914 \\ 5851 \\ 6008 \end{bmatrix}\right) = \begin{bmatrix} 8.468213 \\ 8.499844 \\ 8.674368 \\ 8.7008477 \end{bmatrix}$$

Calculating the Parameters

- Using the equation

$$\boldsymbol{\theta} = (\mathbf{U}^{\mathbf{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathbf{T}}\mathbf{y}$$

We can calculate the values of b_0 and b_1 that minimise the error

$$\boldsymbol{\theta} = \begin{bmatrix} 0.02269839 \\ -36.87906196 \end{bmatrix}$$

Example R Code

Predicting the Future

- Using the optimal model parameters we can estimate how many "Harrys" will be born in 2009

$$\hat{\mathbf{y}} = e^{[2009 \;\; 1]\boldsymbol{\theta}} = 6136$$

- Actual number of Harrys from 2009 : 6143
  - (from National Office for Statistics)

ARMAX and Recursive Least Squares

- For modelling generic systems ARMAX is the standard
  - AR – AutoRegressive
    - ***The current output has a relationship to the previous values of the output***
  - MA – Moving average
    - ***The noise model used***
  - X – eXogeneous inputs
    - ***The system relies not only on the current value of the input, but the history of inputs***

ARMAX and Recursive Least Squares

- An ARMAX Model

$$
\begin{aligned}
y_i &= \\
&- a_1 y_{i-1} - a_2 y_{i-2} \ldots - a_{sa} y_{i-sa} \\
&+ b_1 u_{i-1} + b_2 u_{i-2} \ldots + b_{sb} u_{i-sb} \\
&+ c_1 e_{i-1} + c_2 e_{i-2} \ldots + c_{sc} e_{i-sc} + e_i
\end{aligned}
$$

ARMAX and Recursive Least Squares

$$y_i = \boldsymbol{\varphi}_\mathbf{i}\boldsymbol{\theta}$$

$$\boldsymbol{\varphi}_\mathbf{i} = \begin{bmatrix} -y_{i-1} & -y_{i-2} & \dots & u_{i-1} & u_{i-2} & \dots & e_{i-1} & e_{i-2} & \dots \end{bmatrix}^T$$

$$\boldsymbol{\theta} = \begin{bmatrix} a_1 & a_2 & \dots & b_1 & b_2 & \dots & c_1 & c_2 & \dots \end{bmatrix}^T$$

Collecting all of these instances together gives:

$$\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\theta} + e$$

ARMAX and Recursive Least Squares

- Using the normal LSE for an ARMAX model

$$\boldsymbol{\theta} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \mathbf{y}$$

- But there are problems
  - Every time a new observation is added we have to recalculate the entire thing
  - We're performing a matrix inversion on a big matrix- VERY costly

ARMAX and Recursive Least Squares

- Recursion to the rescue!
- Can we rephrase the update into a recursive form?

$$\boldsymbol{\theta} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{y}$$

becomes

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} + \text{Correction}$$

ARMAX and Recursive Least Squares

- Kalman of "Kalman Filters" fame provides us with

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} + K_n \varepsilon_n$$

- Where:

K_n is the Kalman Gain

$$\varepsilon_n = y_n - \hat{y}_n$$

ARMAX and Recursive Least Squares

- But there's still a problem!
- New equations

$$\varepsilon_n = y_n - \boldsymbol{\varphi}_n^T \hat{\boldsymbol{\theta}}_{n-1}$$

$$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_{n-1} + K_n \varepsilon_n$$

$$K_n = \mathbf{P}_n \boldsymbol{\varphi}_n$$

$$\mathbf{P}_n = \left(\mathbf{P}_{n-1}^{-1} + \boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^T\right)^{-1}$$

- We're **still** performing a massive matrix inversion

ARMAX and Recursive Least Squares

- However, there is **The Matrix Inversion Lemma** that allows us to convert this inversion into a more manageable form

$$\mathbf{P}_n = \left(\mathbf{P}_{n-1}^{-1} + \boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^T\right)^{-1}$$

$$= \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1}\boldsymbol{\varphi}_n \boldsymbol{\varphi}_n^T \mathbf{P}_{n-1}}{1 + \boldsymbol{\varphi}_n^T \mathbf{P}_{n-1} \boldsymbol{\varphi}_n}$$

Noise

- The previous model assumes we know what the noise terms are
- This is plainly ridiculous!

- However, we already compute an estimate of the noise ε_n

- We can simply substitute this in for e_n

Variants of ARMAX

- There are two important variants of ARMAX
  - ARMAX with Forgetting
  - Instrumental ARMAX

ARMAX With Forgetting

- For when the system's properties are slowly shifting
- Allows less emphasis to be placed on older observations
- All equations stay the same except for **P**
- Introduce a "forgetting factor" (λ) between 0-1

$$\mathbf{P}_n = \frac{1}{\lambda}\left(\mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1}\boldsymbol{\varphi}_n\boldsymbol{\varphi}_n^T\mathbf{P}_{n-1}}{\lambda + \boldsymbol{\varphi}_n^T\mathbf{P}_{n-1}\boldsymbol{\varphi}_n}\right)$$

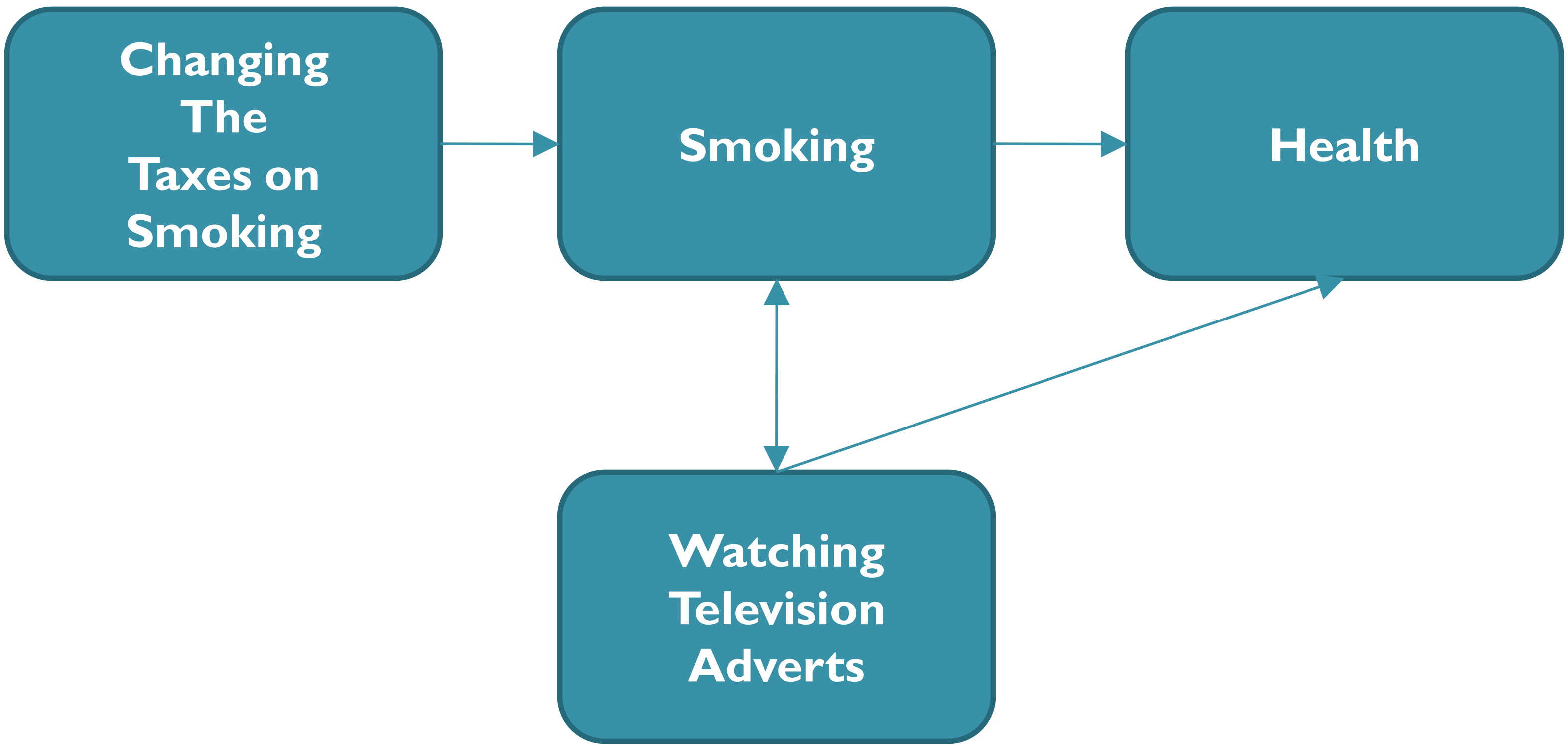
Instrumental ARMAX

- ARMAX assumes that the noise is not changing in proportion to the input
- Not true for many systems
- Introduce an **instrumental variable** that allows us to avoid the results becoming skewed by assuming independence

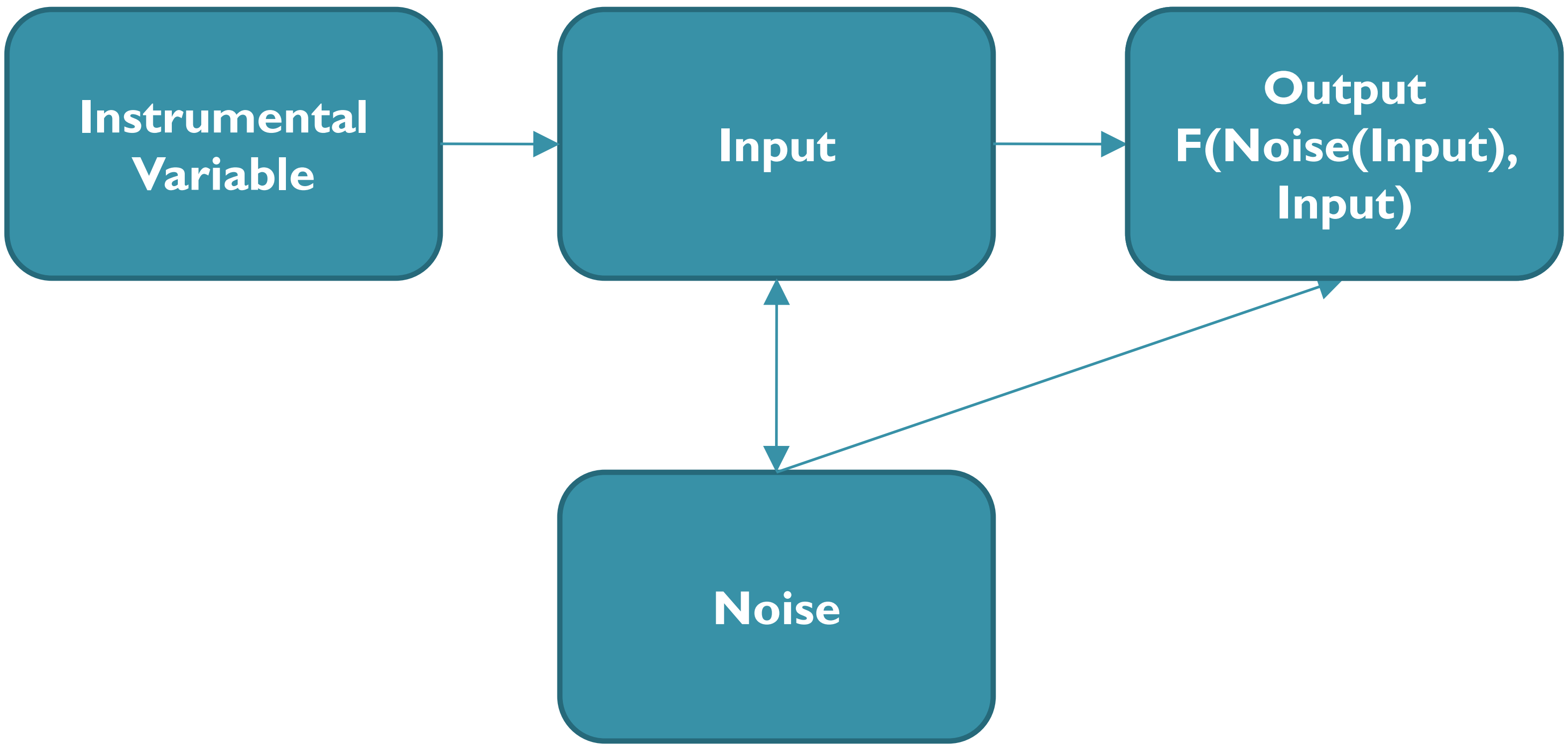
Instrumental Variables - Example

- Smoking is correlated with poor health
- But can we be sure that is a causal relationship?
  - What if there is something which causes both?
  - What if poor health causes smoking!?

Instrument Variables - Example



Instrumental Variables and Estimators



Instrumental ARMAX

- Two of the standard ARMAX Equations change

Instrumental ARMAX

- Two equations change

$$K_n = \mathbf{P}_n \mathbf{z}_n$$

$$\mathbf{P}_n = \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1}\mathbf{z}_n \mathbf{\Phi}_n^T \mathbf{P}_{n-1}}{1 + \mathbf{\Phi}_n^T \mathbf{P}_{n-1}\mathbf{z}_n}$$

Instrumental ARMAX

- All that remains is to find values for **z** that are not correlated with the noise, but are correlated with the input

- A popular choice is an estimate of the current output which uses old model parameters – so unaffected by the latest noise

Summary

- LSE is an effective tool for estimating parameters that you can't directly measure
- ARMAX is a generalised model for discrete, time-varying systems
- Many different variants and techniques available to address a variety of problems