# Remarks and Dedication: George E.P. Box 1919-2013

Kevin Little, Ph.D.

15 October 2013

SWQN Annual Meeting

# BOX AND LUCENO NOTES (1992 EDITION)

# What we all know is true

- Plot process data in time order, “you can see a lot just by looking”
- Apply control chart rules to distinguish signals of special (assignable) causes from patterns of common cause variation.
- Reduce variation by removing special causes and understanding common causes.
- Don’t adjust a process that is in control, else you will increase variability (*Out of the Crisis*, pp. 327-330, “The funnel experiment”).

# But what do we do if…

- Our output process, left to its own devices, will drift away from target?
- Input factors that drive this drift are themselves drifting, not in states of statistical control?
- We can't technically or economically remove the drifts?

# In such a world…

Can we make useful predictions about future performance?

Can we compensate for the drift in a smart way?

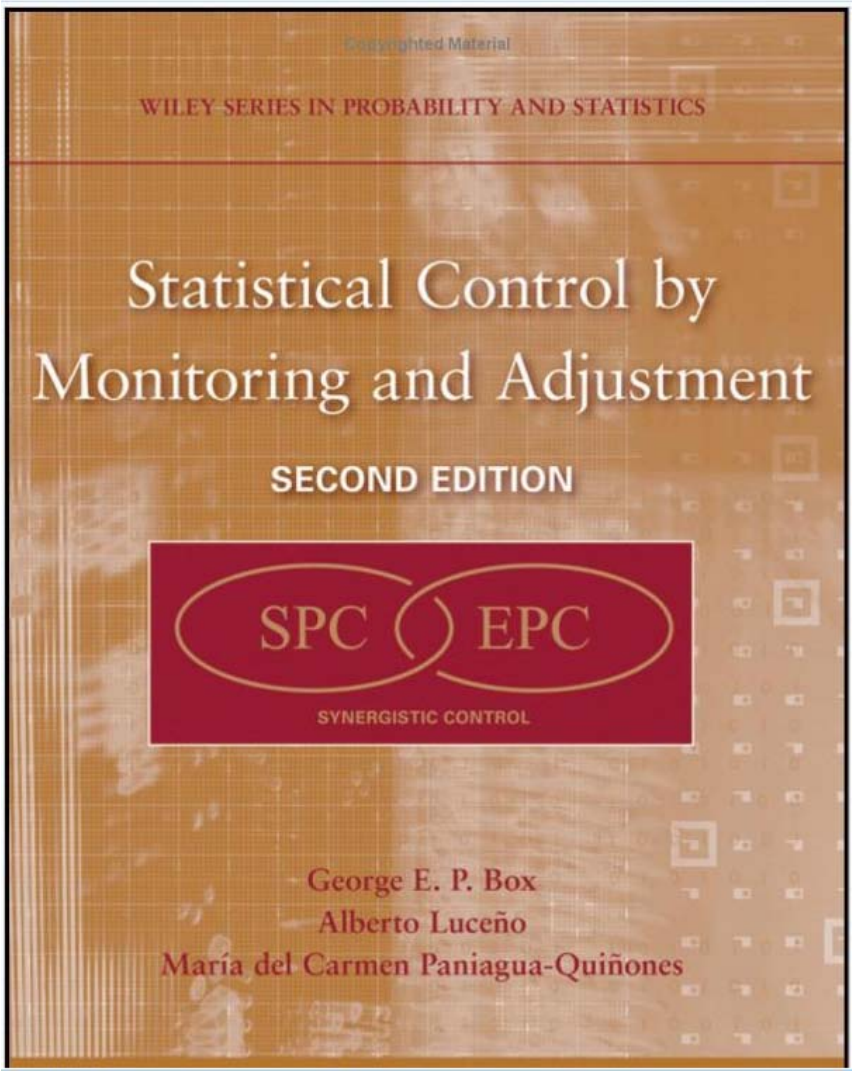
Process World is characterized by flows, flow rates, feedback loops, delays, structural relationships, and (sometimes) factors you can control or adjust to affect the outcomes you want.   This section of the talk is based on chapters 4, 5 and 6 of Statistical Control by Monitoring and Feedback Adjustment, first edition.

Waste water treatment and chemical production of polymers are examples. Biological systems are all in process world.

We'll work with discrete time steps; in many applications, people work with continuous time, too.

# WELCOME TO PROCESS WORLD

A guidebook to process world




# Contents



v

This section of the talk is based on chapters 4-6, first edition (1992).

An overview of chunks of *Statistical Control For Monitoring and Adjustment*

1. The "**Process World**" has features that we can describe with difference equations (Chapter 4)

2. A starting place for process models is a process that wanders "quite a bit" without intervention and adjustment, called the **IMA(1,1)** process (Chapters 1 and 5) (a process "in statistical control" is a special case of IMA(1,1)).

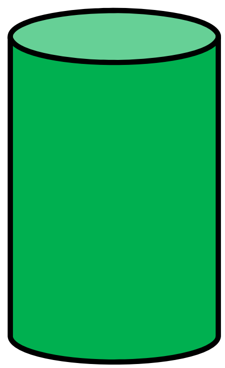
3. An **"Exponentially Weighted Moving Average"** (EWMA) will give you a "best forecast" one step into the future if you are working with the wandering IMA(1,1) process. (Chapter 5)

4. If you have a **knob** that will adjust your IMA(1,1) wandering process, use EWMA to transform the IMA process into a **state of statistical control**. (Chapter 6-9, including issues of cost of adjustment)

# INTRODUCTION TO DIFFERENCE EQUATIONS

# Heat flow:  The can of beans

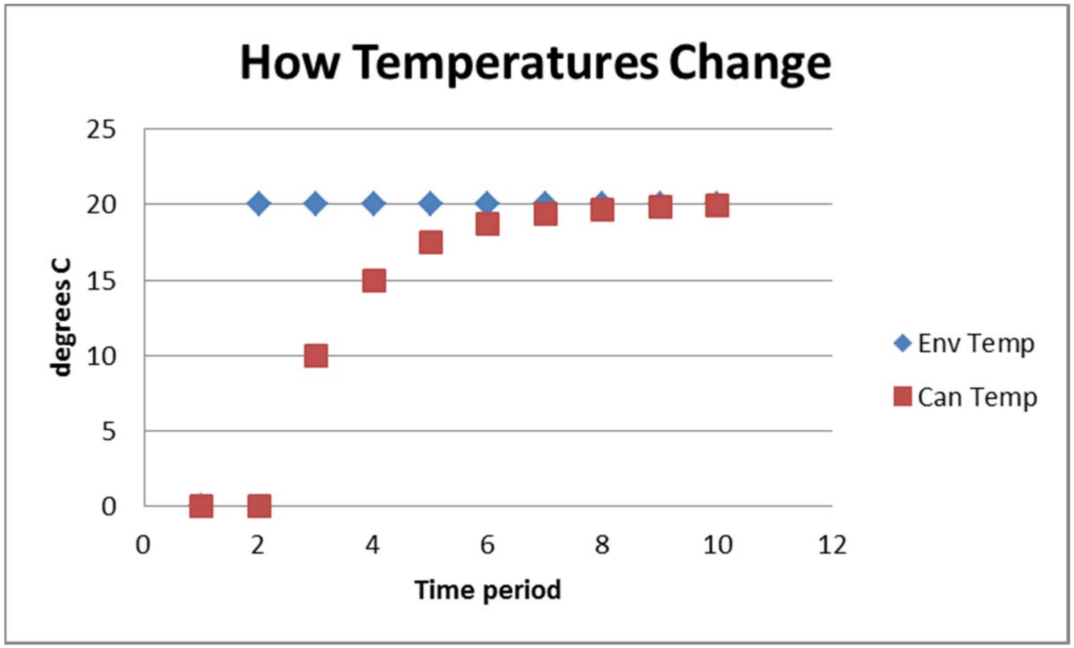
- Can of beans in refrigerator at 0°C.
- Room temperature at 20°C.
- Can will warm to environment temperature at a rate *proportional* to the difference in temperatures at end of time period t.
- Can taken out of refrigerator at end of time period 2.
- At the end of time period 3, can at 10°C.



$X_t$ is the temperature of can's environment
$Y_t$ is the temperature of the can
$1 - \delta$ is the warming rate constant (in the interval (0,1))

p. 89 BL 1st edition

# A model for the problem




The conditions on the previous slide lead to:

$Y_{t+1} - Y_t = (1\text{-}\delta)(X_t - Y_t)$

and $1\text{-}\delta = 0.5$

| t | $X_t$ | $Y_t$ |
|---|---|---|
| 1 | 0 | 0.0 |
| 2 | 20 | 0.0 |
| 3 | 20 | 10.0 |
| 4 | 20 | 15.0 |
| 5 | 20 | 17.5 |
| 6 | 20 | 18.8 |
| 7 | 20 | 19.4 |
| 8 | 20 | 19.7 |
| 9 | 20 | 19.8 |
| 10 | 20 | 19.9 |

# Manipulating the Difference Equation

$Y_{t+1} - Y_t = (1\text{-}\delta)(X_t - Y_t) \Rightarrow Y_{t+1} = (1\text{-}\delta)X_t + \delta\, Y_t$

But $Y_t = (1\text{-}\delta)X_{t-1} + \delta\, Y_{t-1}$

and $Y_{t-1} = (1\text{-}\delta)X_{t-2} + \delta\, Y_{t-2}$

and so on…

Substitute the Y expressions, back in time, to get:

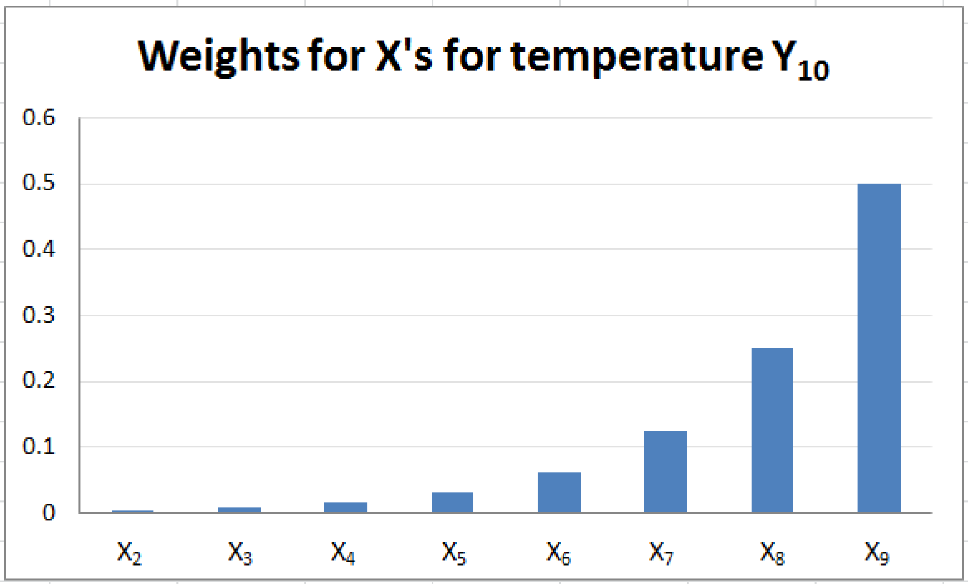
$$Y_{t+1} = (1\text{-}\delta)(X_t + \delta X_{t-1} + \delta^2 X_{t-2} + \delta^3 X_{t-3} + \ldots + \delta^{t-1} X_o) + \delta^t Y_o$$

$$\Rightarrow Y_{t+1} = \tilde{X}_t$$

Info about X's at time t determines $Y_{t+1}$

(for t large, we can ignore $\delta^t Y_o$)

# $\widetilde{X}_t$ is an Exponentially Weighted Moving Average (EWMA)




1. $\widetilde{X}_t$ is a sum of X's
2. A simple average of n items sums each item with weight 1/n.
3. Our sum has exponential weights like (1-δ) $\delta^0$, (1-δ) $\delta^1$, (1-δ) $\delta^2$ etc.
4. As we step in time, the weights move, too
5. It forecasts one step ahead, $\hat{Y}_{t+1} = \widetilde{X}_t$

# For *any set of numbers* $\{z_t\}$ and *any value* $\theta$ in (-1,1)

Let $\hat{z}_{t+1} = \tilde{z}_t = \lambda(z_t + \theta\, z_{t-1} + \theta^2\, z_{t-2} + \ldots)$ with $\lambda = 1 - \theta$

*i.e. the forecast of z at time t+1 is an EWMA($\theta$)*

and define $e_t = z_t - \hat{z}_t$

*i.e. define $e_t$ as the forecast error at time t*

THEN (among other useful relationships*)

$$z_t - z_{t-1} = e_t - \theta\, e_{t-1}$$

*see Table 4.2 BL 1st edition p. 97 reproduced in Appendix 2

# Interim summary

Difference equations capture many useful features of process world.

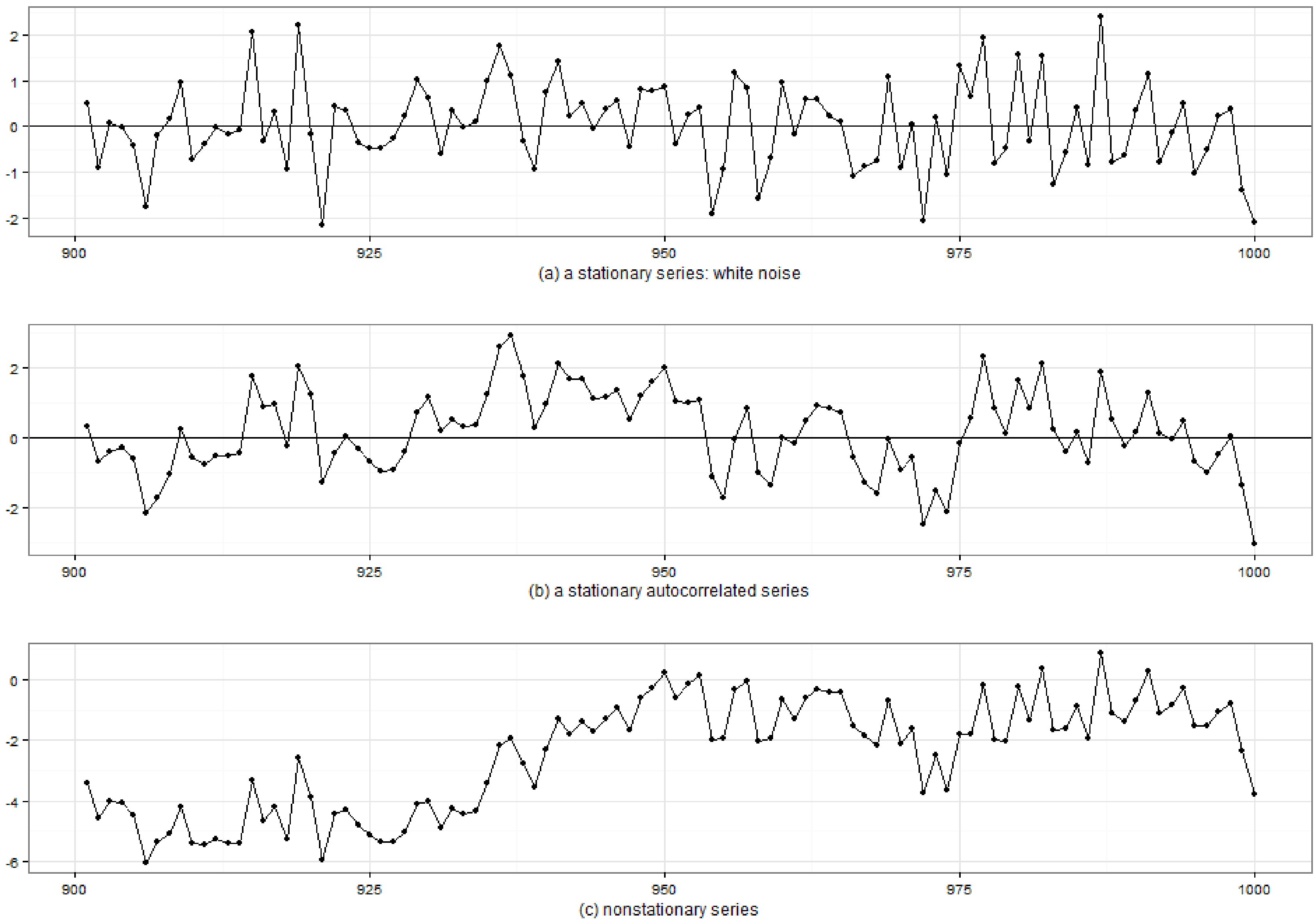
If a difference equation model leads to an EWMA, you have a tool to predict the output, one step ahead.

So far, we have not introduced any statistical variation (common or special causes).

# PROCESS MODELS:  VARIATION

# What are useful models?

Three series like Figure 1.1 Statistical Control by Monitoring and Feedback Adjustment, 1st ed.



# Non-stationary series (c): An Integrated Moving Average

(c) is an example of an integrated moving average, called an "IMA(1,1)" with $\theta = 0.5$

The difference equation form for IMA(1,1):

$$z_{t+1} - z_t = a_{t+1} - \theta a_t$$

where $\{a_t\}$ is a white noise series.

# Two arguments for non-stationary models

1. Theory--2nd Law of Thermodynamics:

"*left to itself*, the entropy (or disorganization) of any system can never decrease." p. 1 1st edition BL

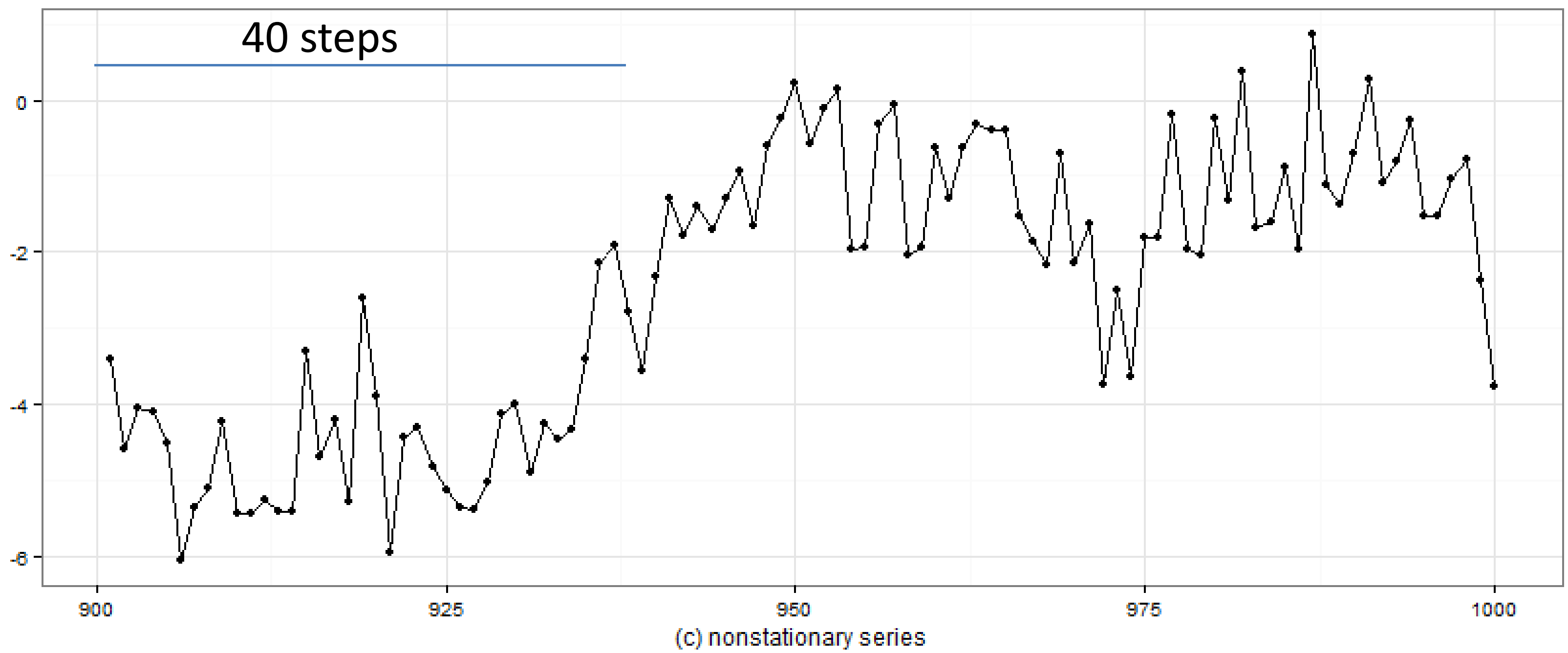
2. Practice--Control methods have been used by engineers for centuries.

Control methods transform non-stationary disturbances to stationary series (like (a) white noise or (b) auto-correlated series.)

p. 5 1st edition BL

# Basic Property of IMA(1,1)

the variability* between observations grows linearly. In the example, with $\theta = 0.5$, the standard deviation 40 steps apart is 3 times the standard deviation one step apart:




*measured by ratio of variance n steps apart to variance 1 step apart

# EWMA GIVES BEST FORECAST FOR IMA(1,1) MODEL

# Four Step Logic

1. If the one-step forecast errors make a “white noise” series, there’s no way to improve forecasts internal to our forecasting system. We’ve got the “best” forecast.
2. For *any set of numbers* $\{z_t\}$ where we set up an EWMA, we get $z_t - z_{t-1} = e_t - \theta\, e_{t-1}$ where $e_t$ is the forecast error $z_t - \hat{z}_t$
3. IMA(1,1) has form: $z_{t+1} - z_t = a_{t+1} - \theta a_t$ , $\{a_t\}$ white noise
4. Given IMA, set up EWMA. Forecast error is now a white noise series and you have the “best.”

# ADJUSTING TO GET BETTER PERFORMANCE

# Adjusting a process--Set up

- Adjustment knob position at time t is $X_t$
- One unit of adjustment produces **g** units change in the output.
- Deviations from target (disturbances) are $\{z_t\}$

*If we could look into the future*, we would adjust the process to exactly compensate the disturbance:

$$gX_t = - z_{t+1}$$

A very good alternative is:

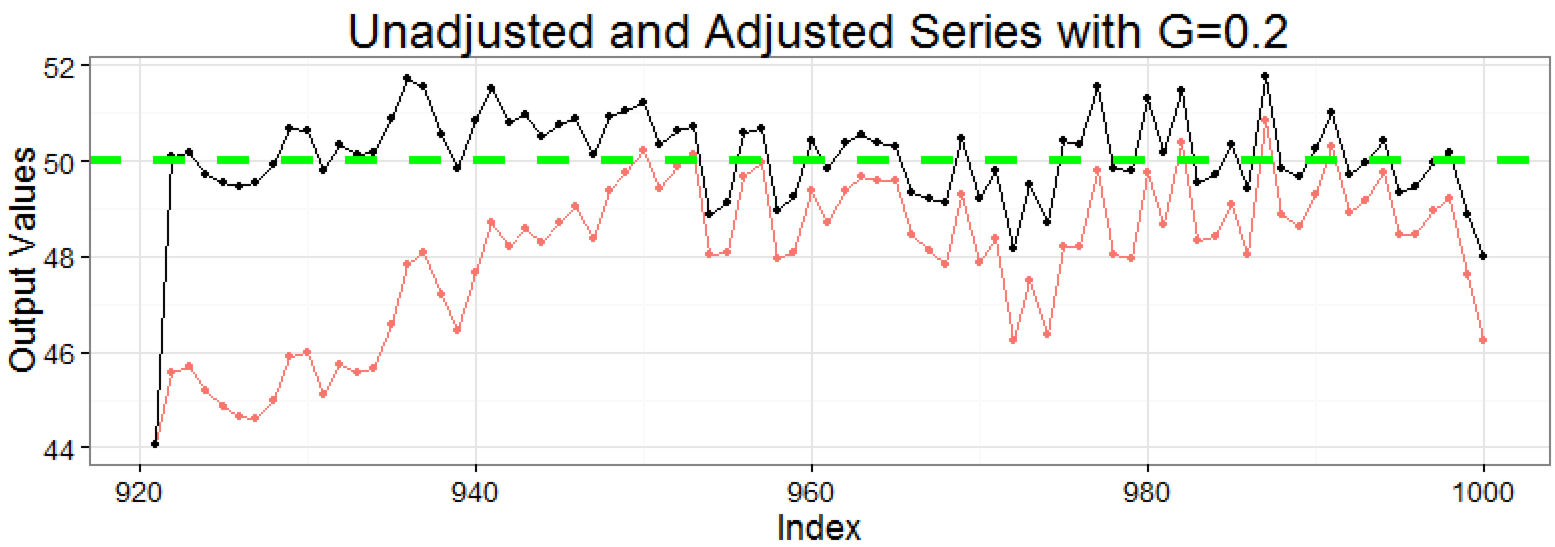
$$gX_t = - \hat{z}_{t+1} = - \tilde{z}_t \quad \text{(our friend, EWMA)}$$

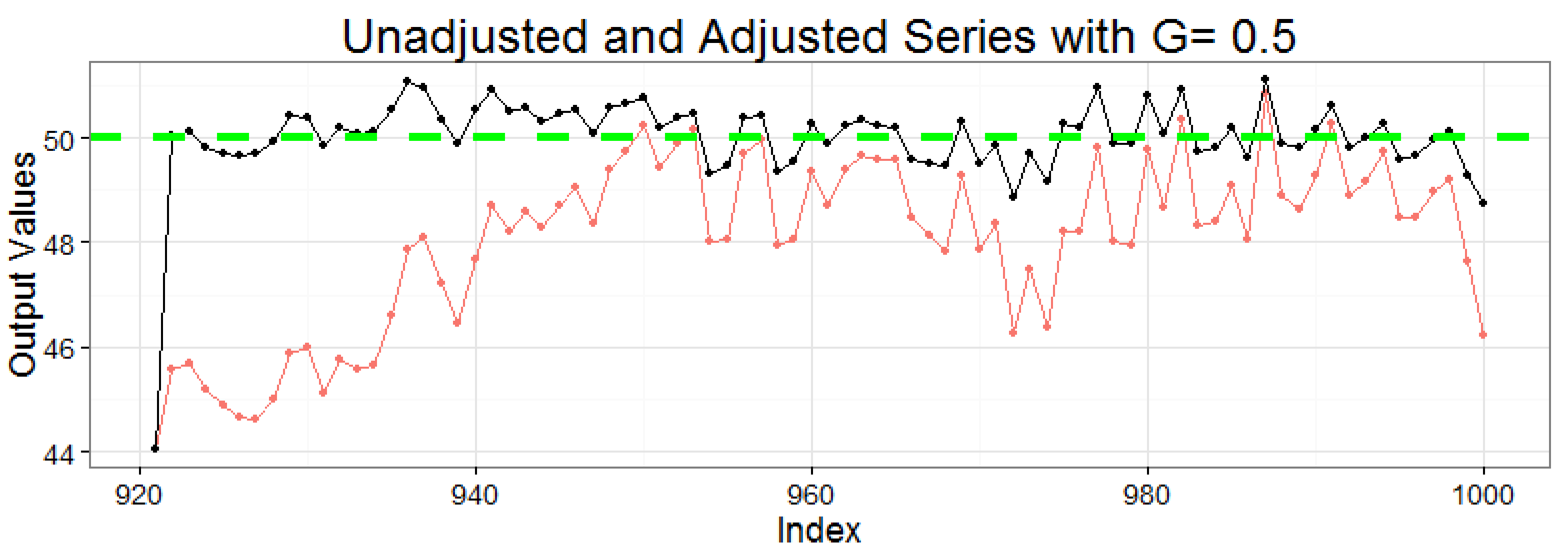
# Adjusting a process--Example

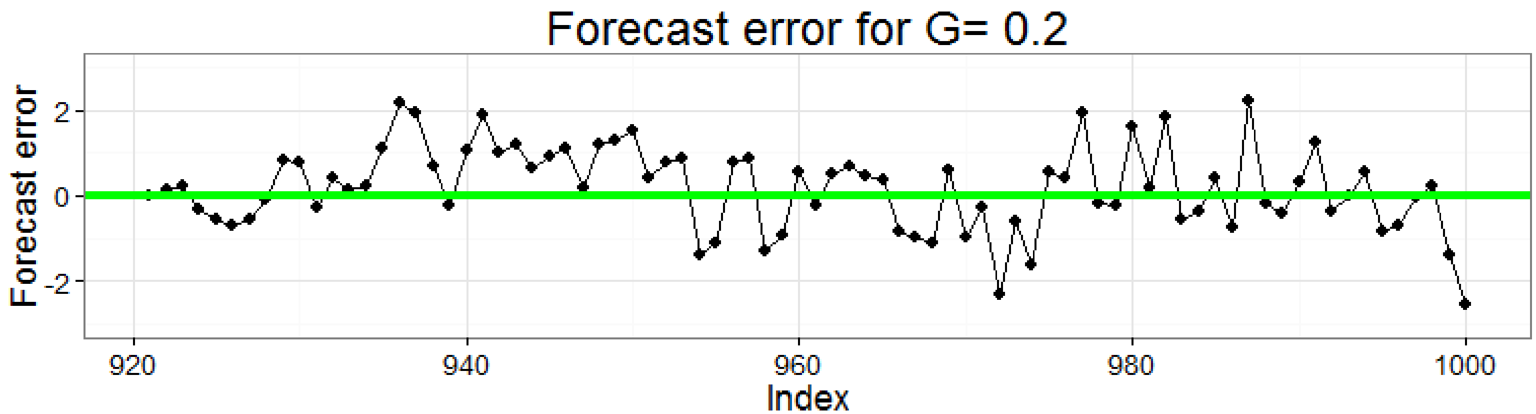
1. The disturbance process is IMA(1,1) with $\theta$ = 0.5
2. We have an adjustment knob, with g = 1.
3. The target is **50**
4. We use the EWMA technology to make adjustments

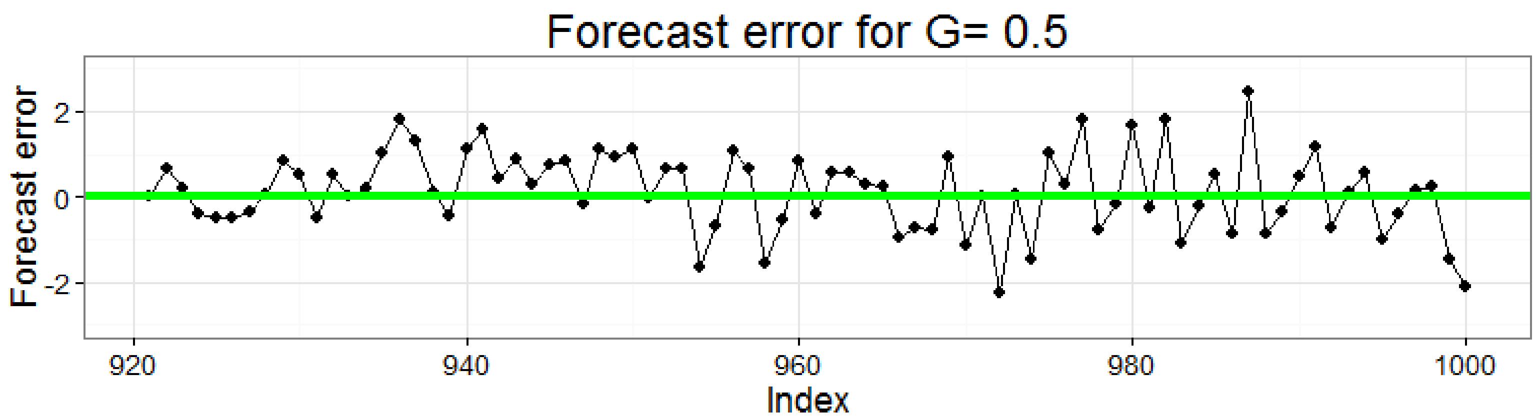
   a. with factor 0.2 (which would be "best" for $\theta$ = 0.8)

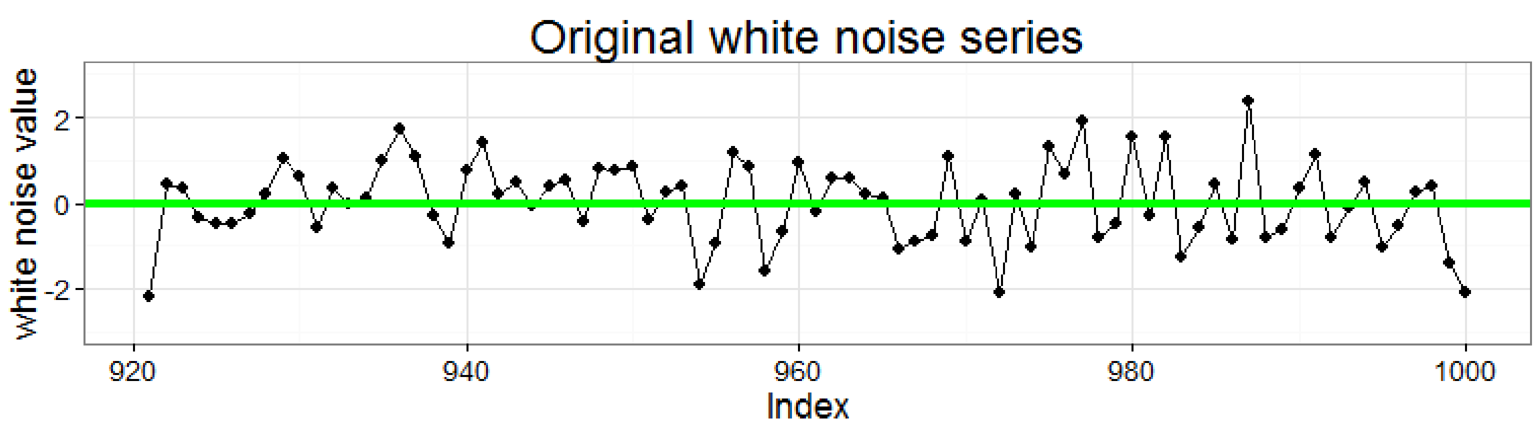
   b. with factor 0.5, the perfect adjustment, because it matches $\theta$ = 0.5.


Unadjusted and Adjusted Series with G=0.2
Output Values
44
46
48
50
52
920
940
960
980
1000
Index

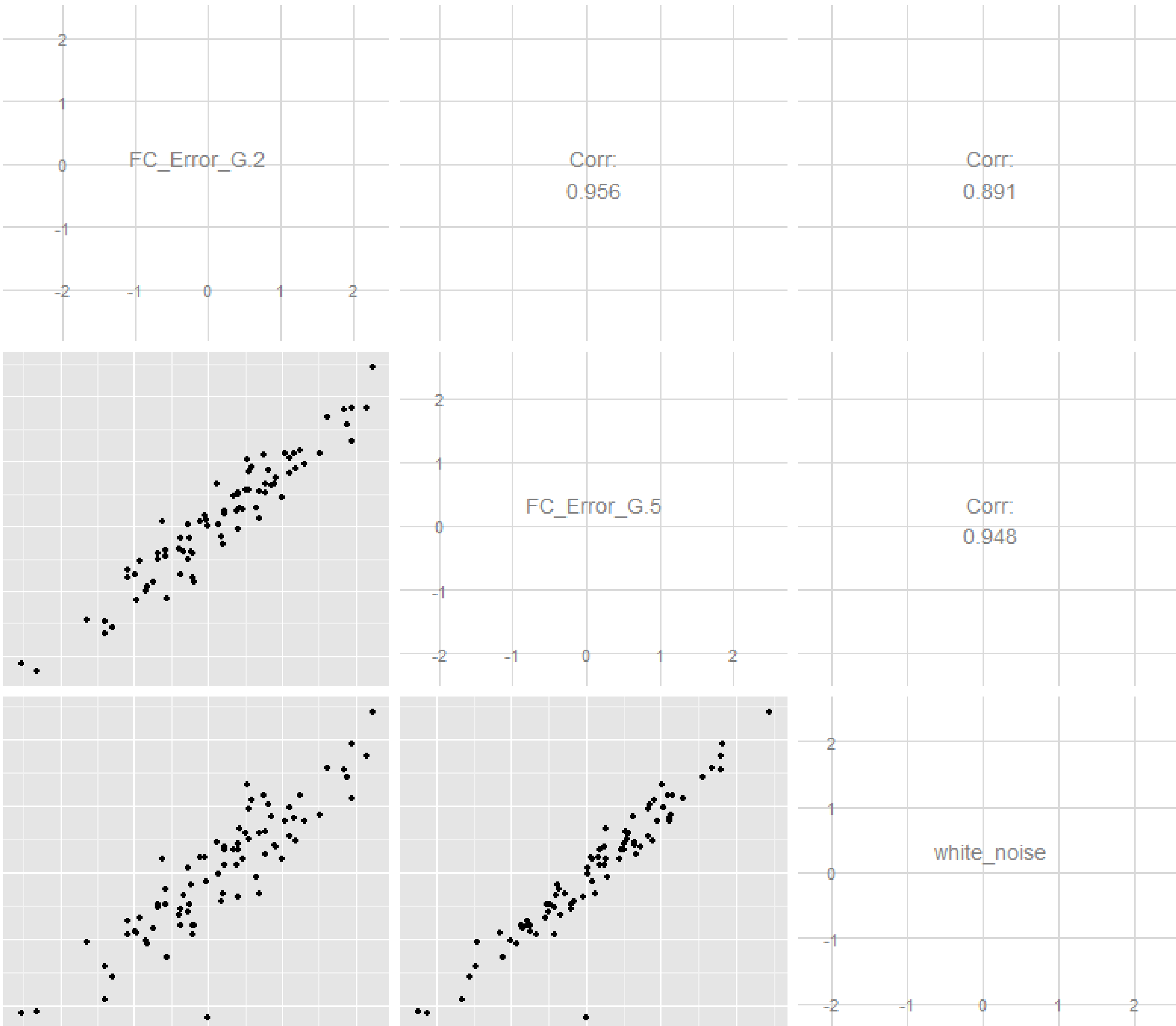

Unadjusted and Adjusted Series with G= 0.5
Output Values
44
46
48
50
920
940
960
980
1000
Index


Forecast error for G= 0.2
Forecast error
2
0
-2
920
940
960
980
1000
Index


Forecast error for G= 0.5
Forecast error
2
0
-2
920
940
960
980
1000
Index


Original white noise series
white noise value
2
0
-2
920
940
960
980
1000
Index

Check of errors



# SUMMARY

# In much of the real world…

- Our output process, if left to its own devices, will drift away from target.
- Input factors that drive this drift are themselves drifting, not in states of statistical control.
- We often can’t technically or economically remove the drifts.

# In such a world…

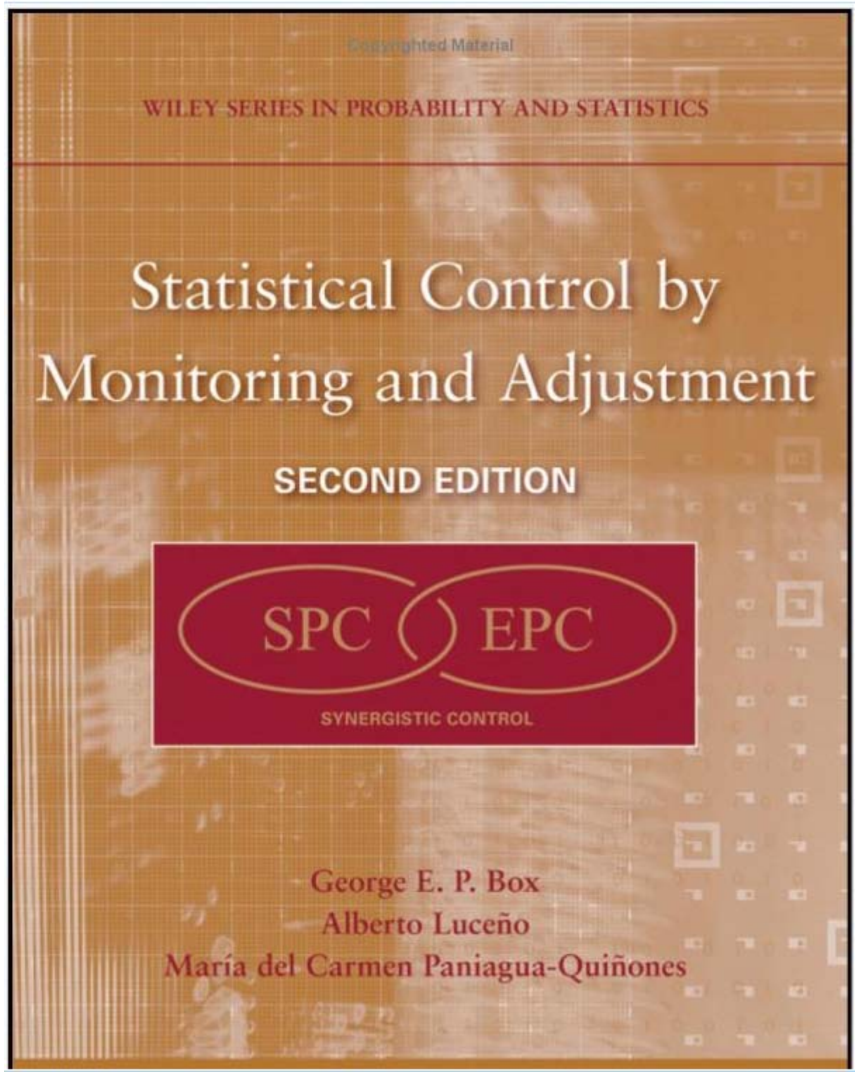
Can we make useful predictions about future performance?

Yes

Can we compensate for the drift in a smart way?

Yes, if we have an adjustment knob

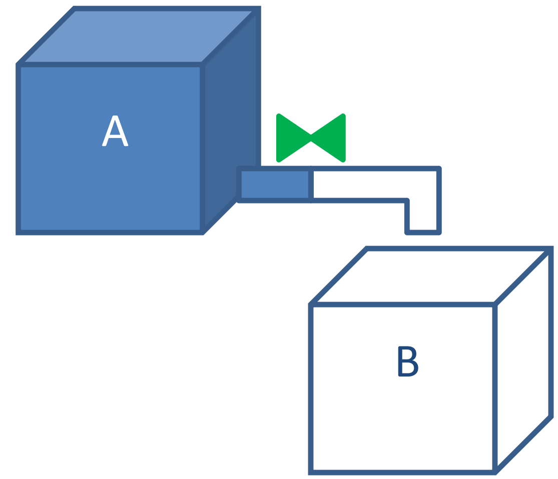
# I recommend this guidebook to process world!



# APPENDIX: CHAPTERS 4 AND 5 BL 1ST EDITION ADDITIONAL NOTES

# Process World: Water flowing from tank A to tank B

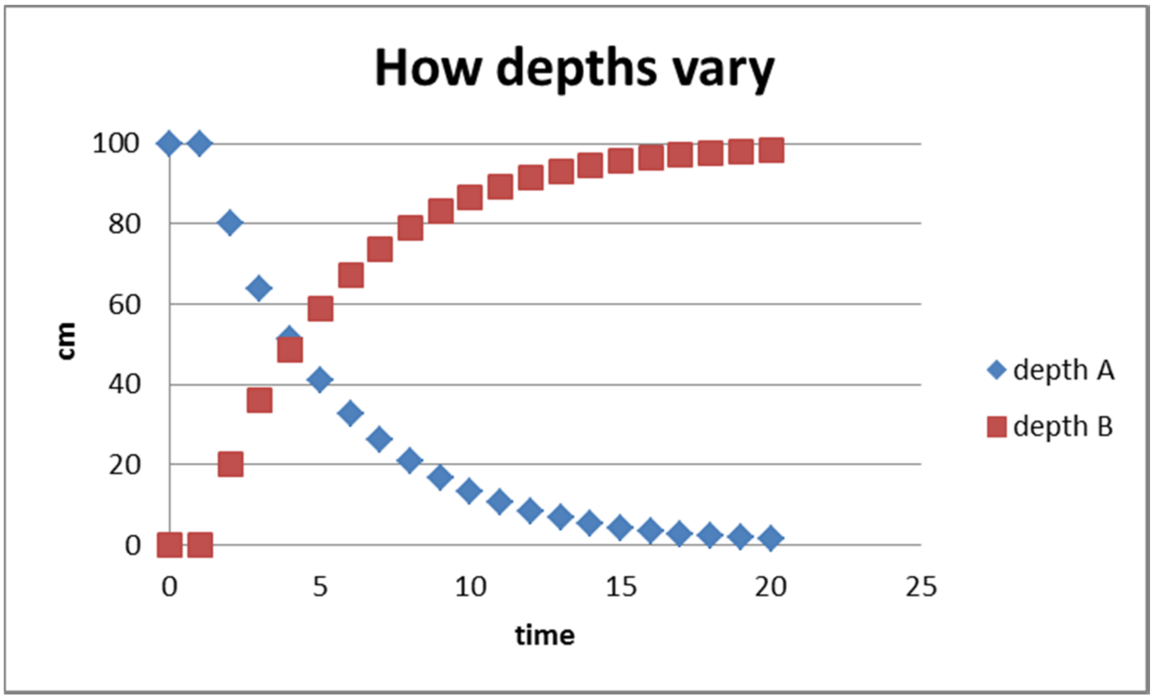
- Two identical tanks A and B, cubes with side 100 cm.
- Rate of flow into B proportional to the head (depth) in tank A
- At time 0, tank A is full, tank B is empty.
- At end of time step 1, the valve is opened.
- At the end of time step 2, tank B is filled to 20 cm.
- How do the depths of the tanks vary over time?




$A_t$ =depth of tank A at end of time period t; $B_t$ = depth of tank B at end of time period t.

Exercise 4.1 BL 1st edition p. 89

# A model for the problem (assuming you don't fiddle with the valve!)




The conditions on the previous slide lead to:

$$B_{t+1} = 0.2A_t + B_t$$

| t | $A_t$ | $B_t$ |
|---|---|---|
| 0 | 100 | 0 |
| 1 | 100 | 0 |
| 2 | 80 | 20 |
| 3 | 64 | 36 |
| 4 | 51.2 | 48.8 |
| 5 | 41.0 | 59.0 |
| 6 | 32.8 | 67.2 |
| 7 | 26.2 | 73.8 |
| 8 | 21.0 | 79.0 |
| 9 | 16.8 | 83.2 |
| 10 | 13.4 | 86.6 |

# A “difference” equation arises…

because we can rewrite our equation

$$B_{t+1} = 0.2A_t + B_t$$

$$\Rightarrow B_{t+1} - B_t = 0.2A_t$$

to describe the **difference** in tank B depth over one time interval

# Table 4.2 BL 1st edition p. 97

<u>Important Relationships for Exponentially Weighted Moving Averages</u> (which hold for any series of numbers $\{z_t\}$ and any value of $\theta$ in (-1,1)—though the practical values of interest are usually in (0,1)):

Let $\hat{z}_{t+1} = \tilde{z}_t = \lambda(z_t + \theta\, z_{t-1} + \theta^2\, z_{t-2} + \ldots)$ with $\lambda = 1 - \theta$ and $e_t = z_t - \tilde{z}_{t-1} = z_t - \hat{z}_t$

| | Algebra | Words |
|---|---|---|
| **A.** | $\tilde{z}_t = \lambda z_t + \theta \tilde{z}_{t-1}$ | The EWMA at time t is an interpolation between the current observation and the previous EWMA |
| **A1.** | $\hat{z}_{t+1} = \lambda z_t + \theta \hat{z}_t$ | The one step ahead forecast at time t is an interpolation between the current observation and the previous forecast |
| **B.** | $\tilde{z}_t - \tilde{z}_{t-1} = \lambda( z_t - \tilde{z}_{t-1})$ | The EWMA difference is a fraction of the difference between the current observation and the previous EWMA. |
| **B1.** | $\hat{z}_{t+1} - \hat{z}_t = \lambda e_t$ | The difference in forecasts is a fraction of the forecast error. |
| **C.** | $z_t - z_{t-1} = e_t - \theta\, e_{t-1}$ | The difference in observations is a simple function of the forecast errors. |

# The Variogram

"If left to themselves, machines continue to go out of adjustment, tools continue to wear out, and people continue to miscommunicate and forget. In all such circumstances, it is reasonable to expect that, if no corrective action is taken, observations spaced further apart will differ more and more."

BL 1st edition, p. 114

Let $V_m$ = variance($z_{t+m} - z_t$)

In particular,

$V_1$ = variance($z_{t+1} - z_t$)

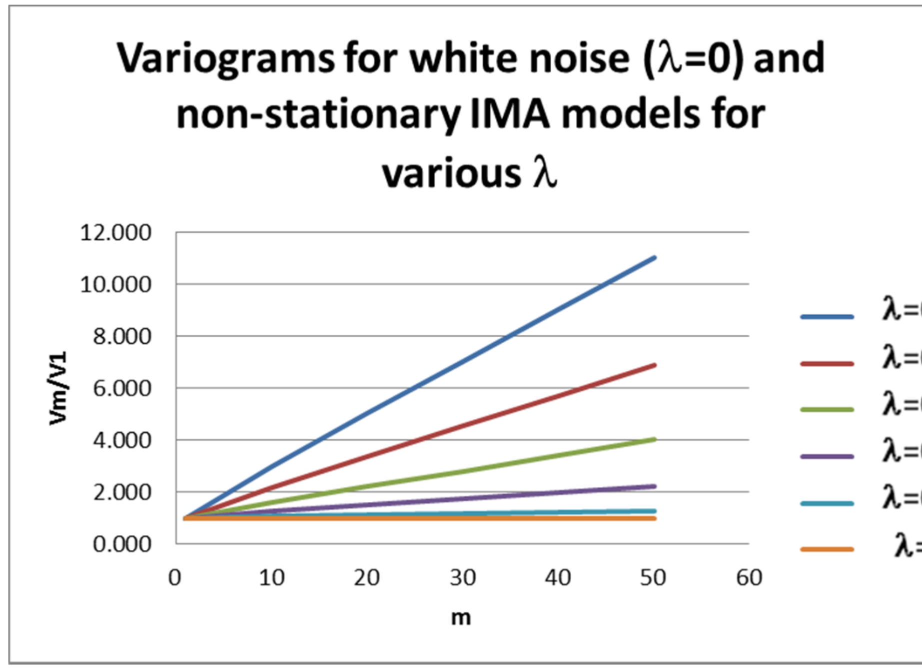
Variogram:

plot of $V_m / V_1$ versus m

# Variogram Insight: IMA(1,1) is the simplest model with increasing $V_M/V_1$

"…a realistic model [for uncontrolled disturbances] would be one in which a steady increase in $V_m/V_1$ occurred as m increased."

$V_m/V_1$ increasing linearly in m *corresponds exactly* to the IMA(1,1) non-stationary model. BL 1st edition, p. 114




See figure 5.3 BL 1st edition, p. 115

# Contact information

Kevin Little, Ph.D.

klittle@iecodesign.com

608-251-4355