Retail and Marketing Analytics

Session 3

Gokhan Yildirim

Outline

Morning session:

- Omnichannel dynamics
- VAR model

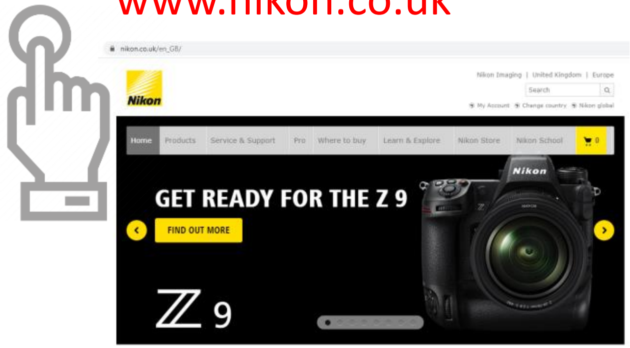
Afternoon session:

- R tutorial
- Simulation Game

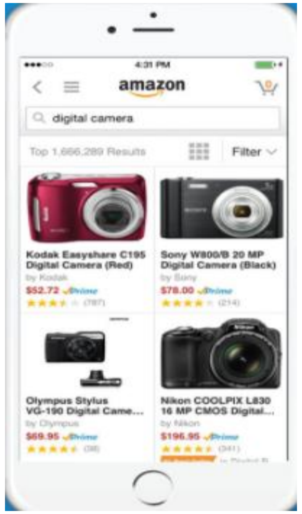
Omnichannel dynamics

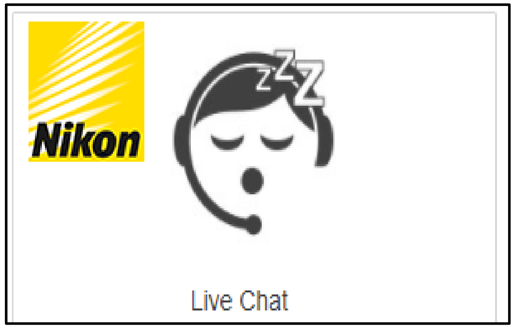
Customer journey is becoming much more complex…

www.nikon.co.uk

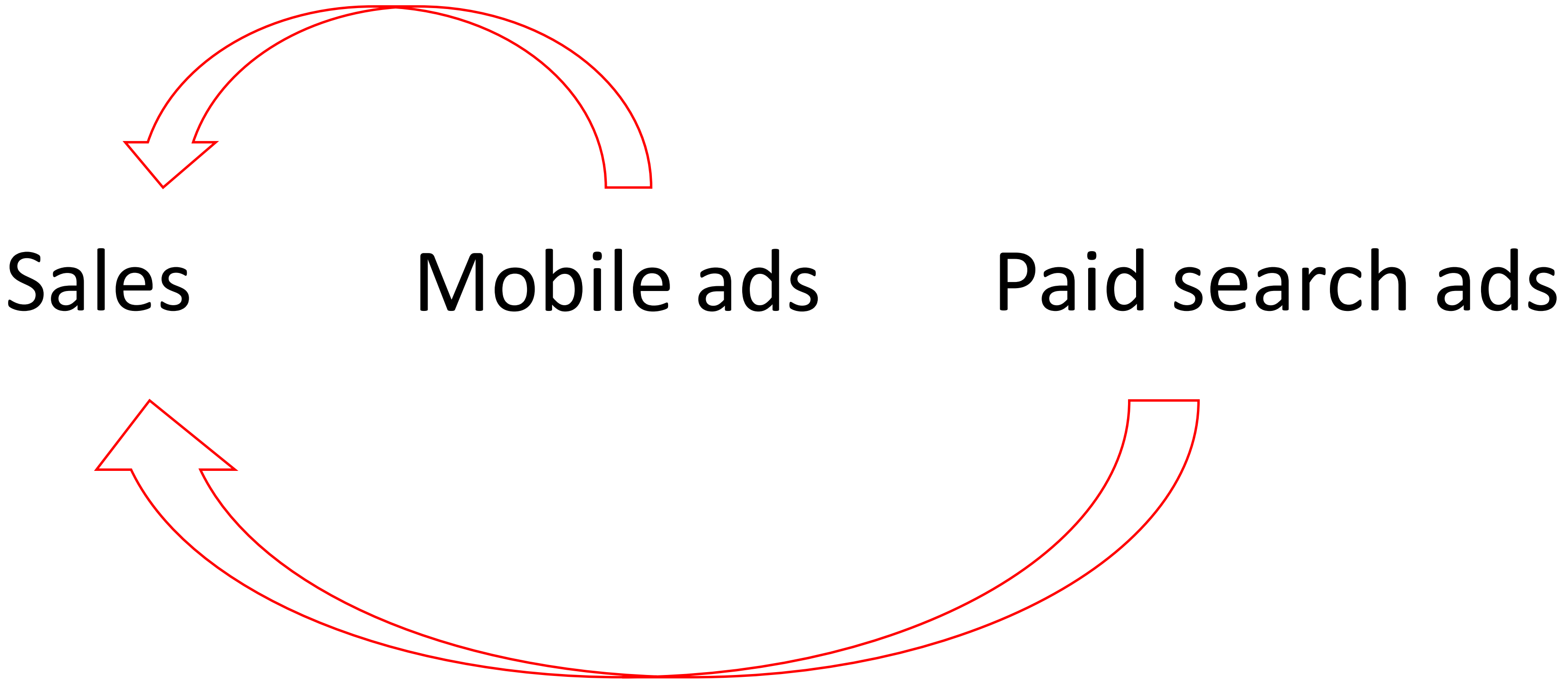


£

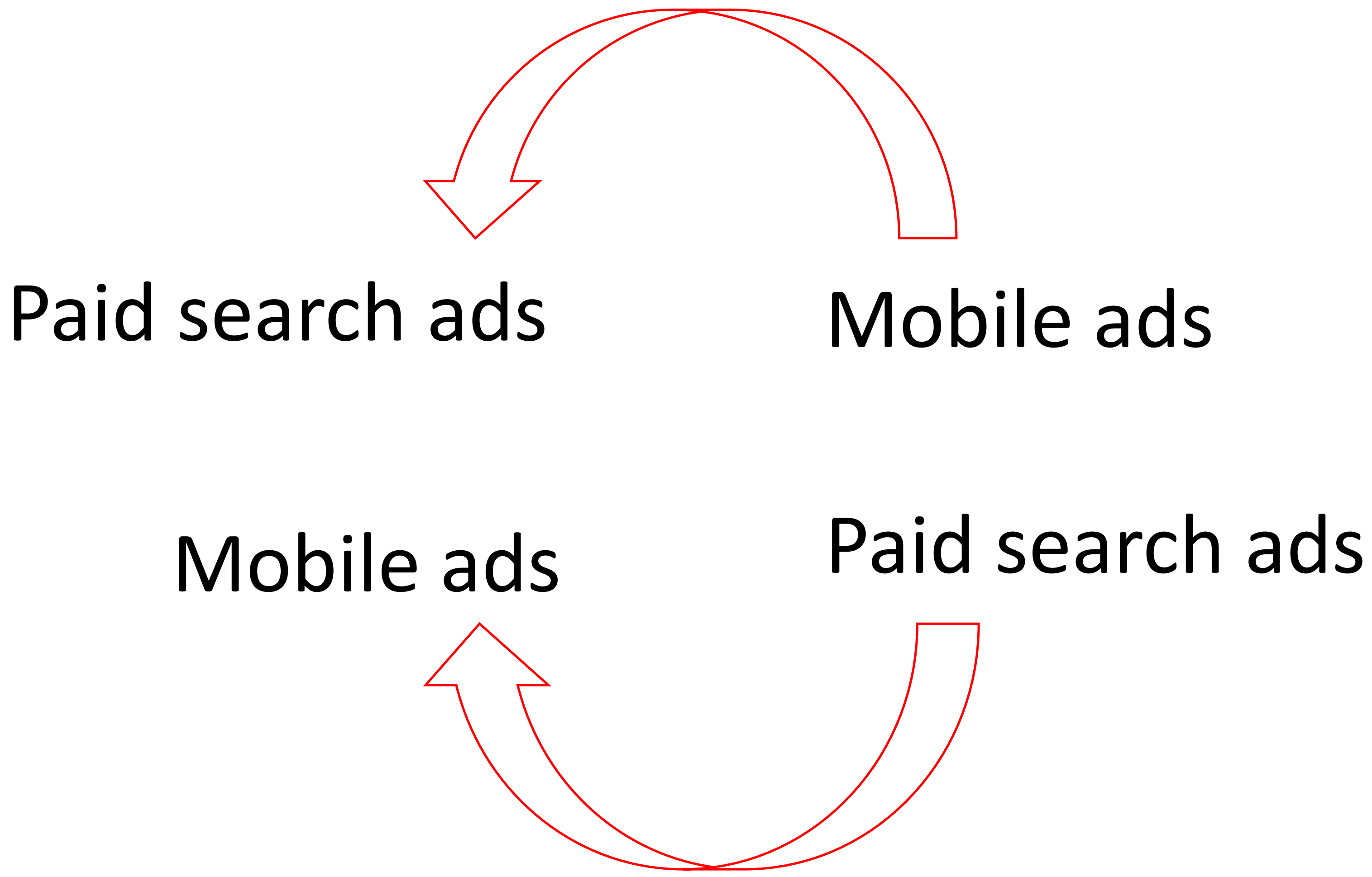




Going beyond the direct effects…



Cross media effects



Carryover effects

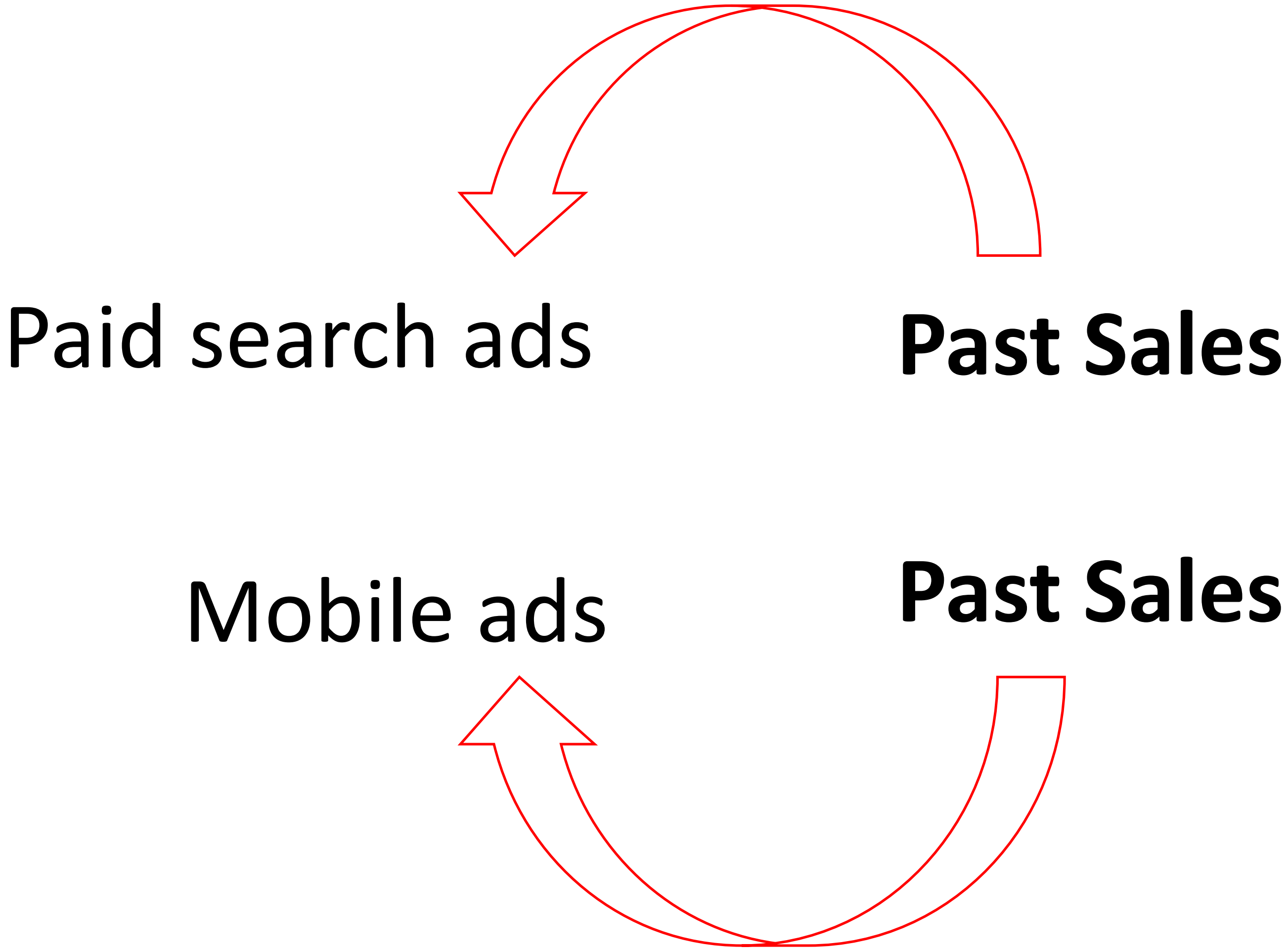
Mobile ads

Past mobile ads

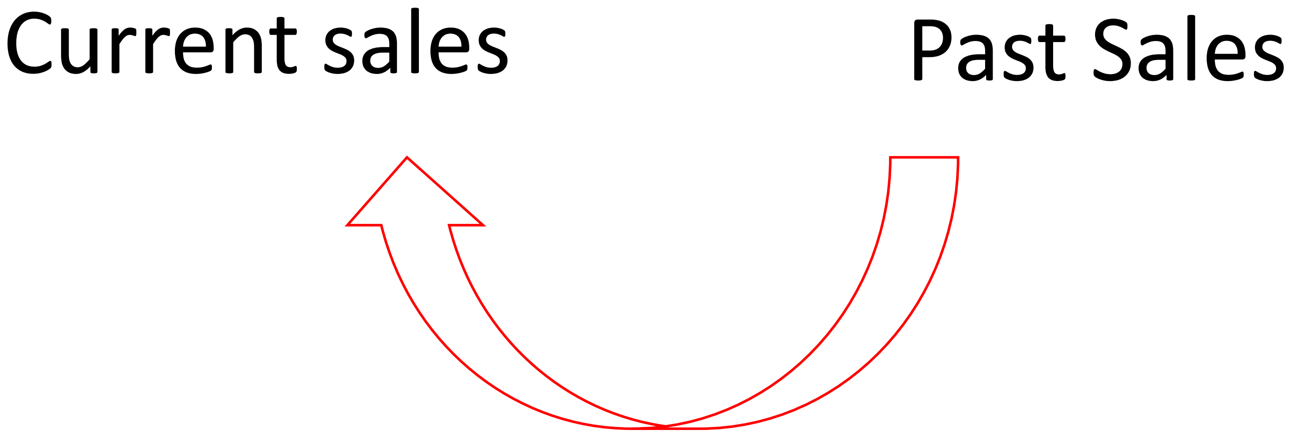
Paid search ads

Past paid search ads

Performance feedback effects



Purchase reinforcement effects



How to model these complex loops?

VAR

What can we do with VAR?

Model features:

- Cross effects
- Dual causality
- Complex feedback loops

In practice:

- Short- vs. long-term effects of marketing programs
- Wear-in and wear-out effects
- Marketing budget allocation using long-term elasticities

VAR Models

A p-th order VAR, VAR(p) is represented as follows:

$$y_t = C + \sum_{i=1}^{p} \Phi_i y_{t-i} + \Gamma X_t + \varepsilon_t, \quad t=1,2,\ldots T$$

VAR(1) with 3 endogenous and 3 exogenous variables:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

VAR Model: example

VAR(1) with **3 endogenous** (**paid search, mobile advertising,** and **sales**) and **3 exogenous variables** (**promotion campaigns 1, 2,** and **3**):

Intercepts

Cross effects

Feedback effects

Promotion effects on
Search, Mobile and Sales

$$\begin{bmatrix} Search_t \\ Mobile_t \\ Sales_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \varphi_{11} & \varphi_{12} & \varphi_{13} \\ \varphi_{21} & \varphi_{22} & \varphi_{23} \\ \varphi_{31} & \varphi_{32} & \varphi_{33} \end{bmatrix} \begin{bmatrix} Search_{t-1} \\ Mobile_{t-1} \\ Sales_{t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} Promo1_t \\ Promo2_t \\ Promo3_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{bmatrix}$$

Direct impact of Search
and Mobile advertising
on Sales

Carryover effects

Purchase reinforcement effects

Endogenous or Exogenous?

Granger Causality

Granger causality

Temporal ordering of events can be used to make an empirical distinction between leading and lagging variables.

This is the basis of a well-known definition of causality due to Granger (1969).

Granger Causality Definition:

- *X* is said to Granger-cause *Y* if *Y* can be better predicted using the histories of both *X* and *Y* than it can by using the history of *Y* alone.

- In other words, X Granger causes Y if the forecast error of the model Y=f(past Y, past X) is lower than that of the model Y=f(past Y).

In more technical terms…

In more technical terms:

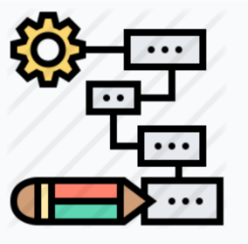
- X is said to Granger cause Y if the mean squared forecast error (MSFE) of Y using the bivariate model Y=f(past Y, past X) is smaller than the MSFE of the univariate model Y=f(past Y).

For the information set containing X and Y, X is said to Granger cause Y if:

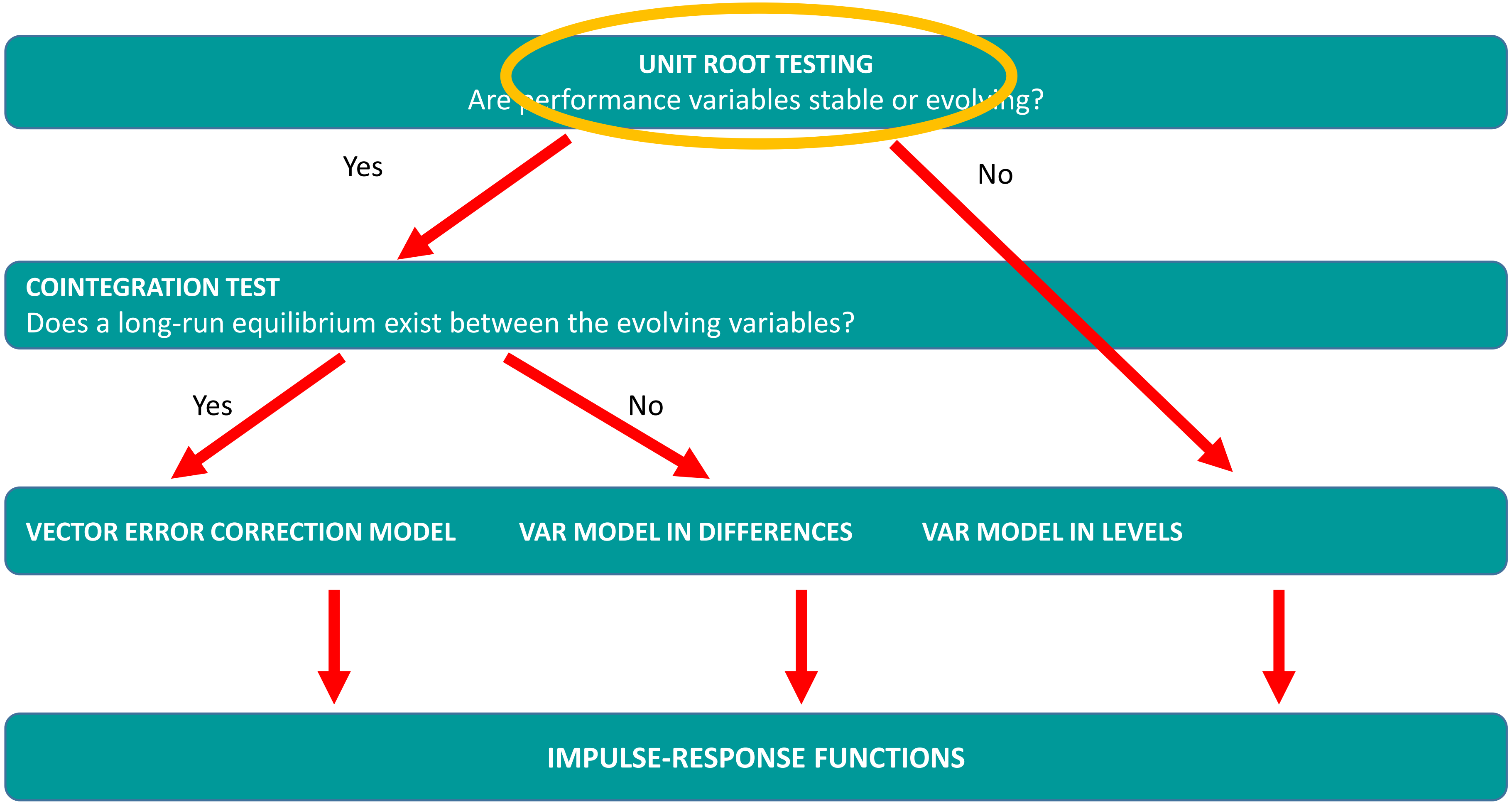
$$\boldsymbol{MSFE(Y_t|Y_{t-1},\dots,Y_{t-k},X_{t-1},\dots,X_{t-m}) < MSFE(Y_t|Y_{t-1},\dots,Y_{t-k})}$$

where k and m are positive integers indicating maximum memory length in Y and X.

Developing a VAR model

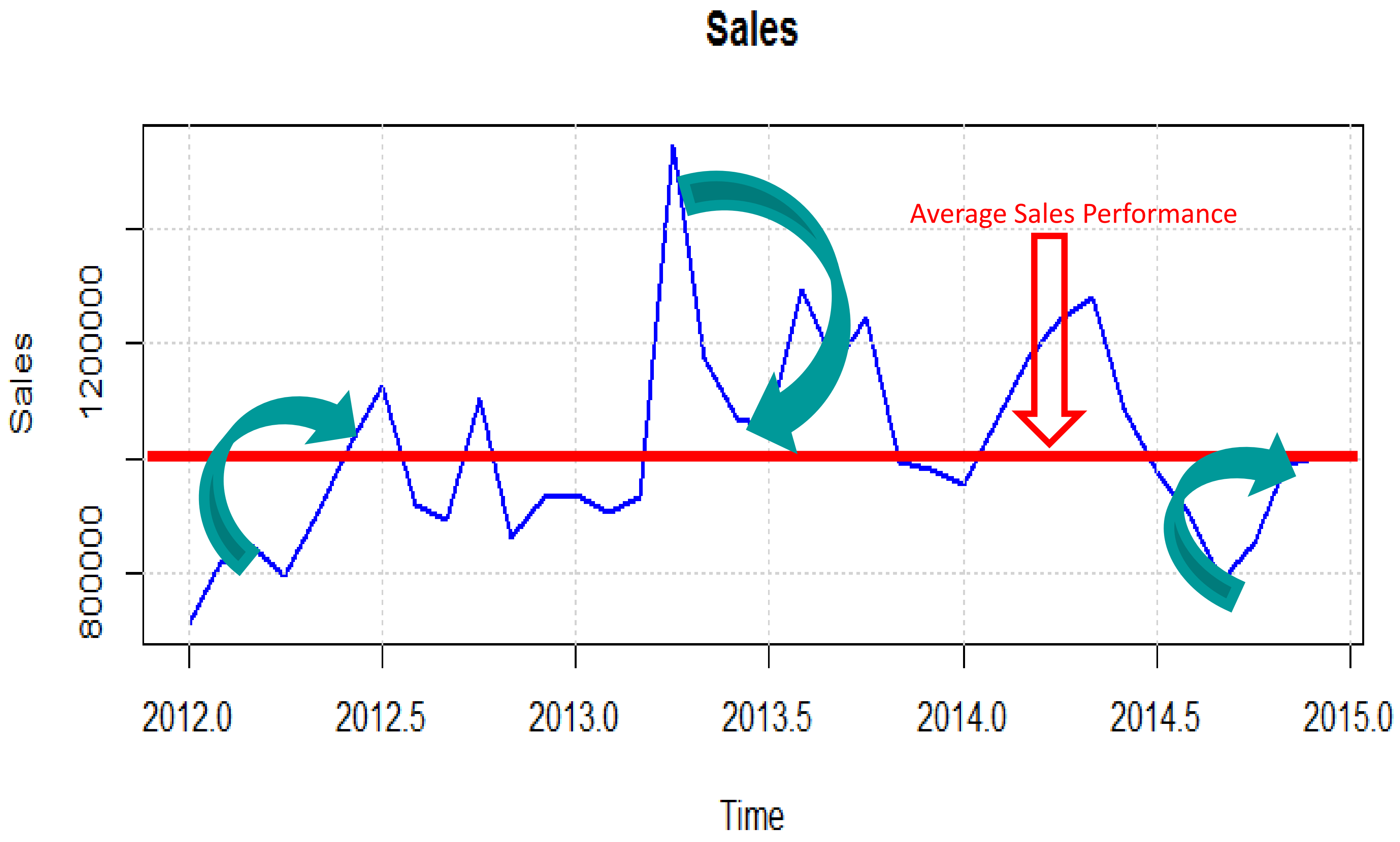


VAR modeling steps

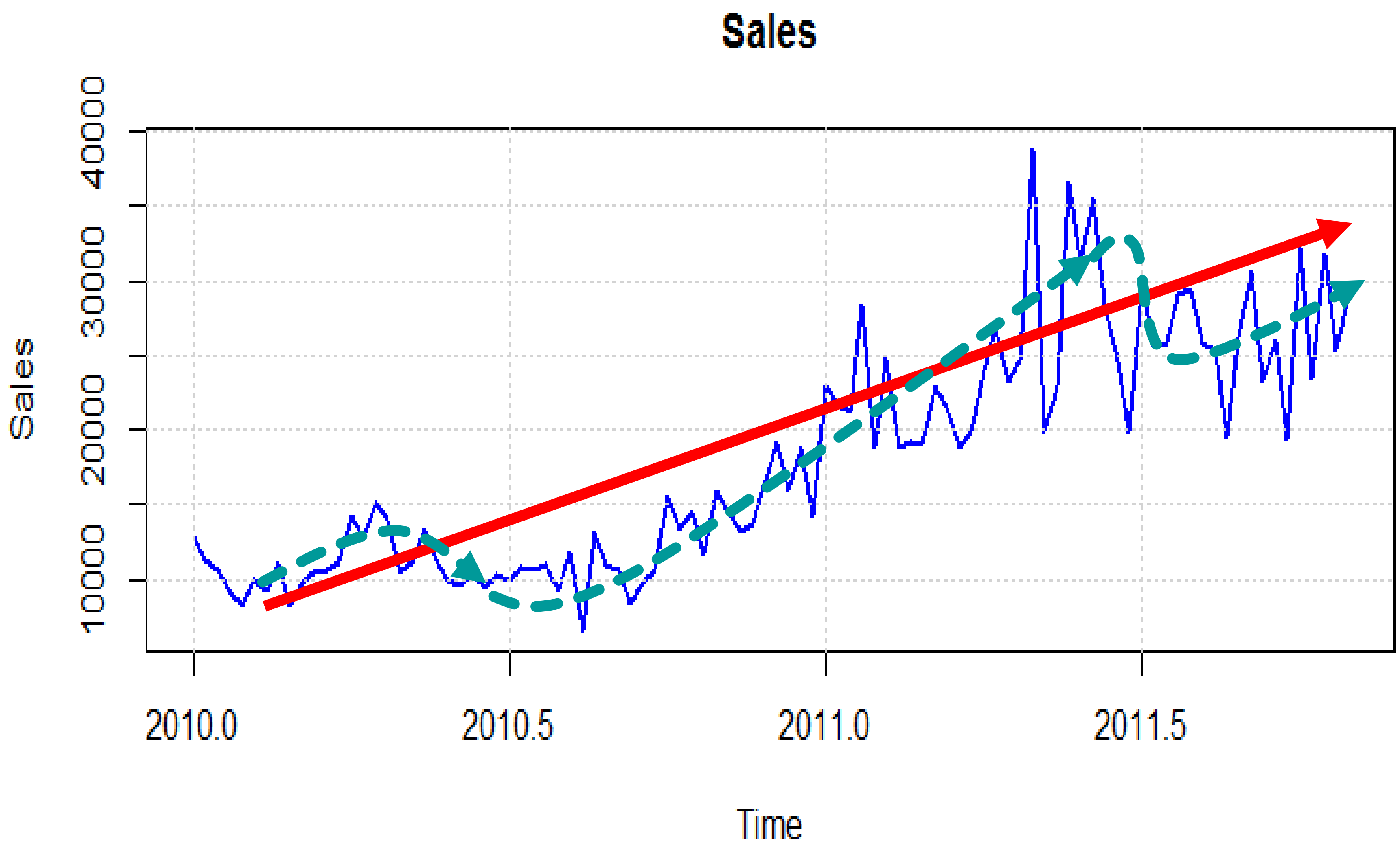


Source: Dekimpe et al. (2006), “Time Series Models in Marketing”, ERIM Report Series, Research in Management.

Mean reversion



Evolving performance



What goes wrong if I ignore unit root(s)?

You find significant relationships when in fact there are no → *Spurious relationships.*

$$y_t = y_{t-1} + v_t \text{ where } v_t \sim NID\left(0, \sigma_y^2\right) \qquad \Delta y_t \sim NID\left(0, \sigma_y^2\right)$$

$$x_t = x_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim NID(0, \sigma_x^2) \qquad \Delta x_t \sim NID(0, \sigma_x^2)$$

INDEPENDENT OF EACH OTHER

$$y_t = \alpha + \beta x_t + \epsilon_t$$

- β must be zero.
- The t-statistic of $\hat{\beta}$ must go to zero as the sample size increases.
- It is not what happens when you have a unit root.

The t-statistic becomes biased towards accepting relations even if variables are independent of each other.

R^2 values become relatively high while the Durbin-Watson statistic becomes low.

Unit Root tests

Unit Root tests identify the presence/absence of a long-run (stochastic-trend) component in the series' data-generating process, and hence distinguish stable from evolving variables.

In other words, the series is tested if it is stationary or not.

A series is said to be integrated of order d, denoted I(d), if it becomes (weakly) stationary after differencing d times.

There are several unit root tests: DF, ADF, KPSS, PP etc.

We will use Augmented Dickey Fuller Test (ADF) Tests.

Unit Roots

ADF test:

$$\Delta y_t = \alpha + \beta t + \boxed{\pi} y_{t-1} + \sum_{k=1}^{K} \gamma_i \Delta y_{t-i} + \varepsilon_t$$

$$H_0\colon \pi = 0$$

$$H_A\colon \pi < 0$$

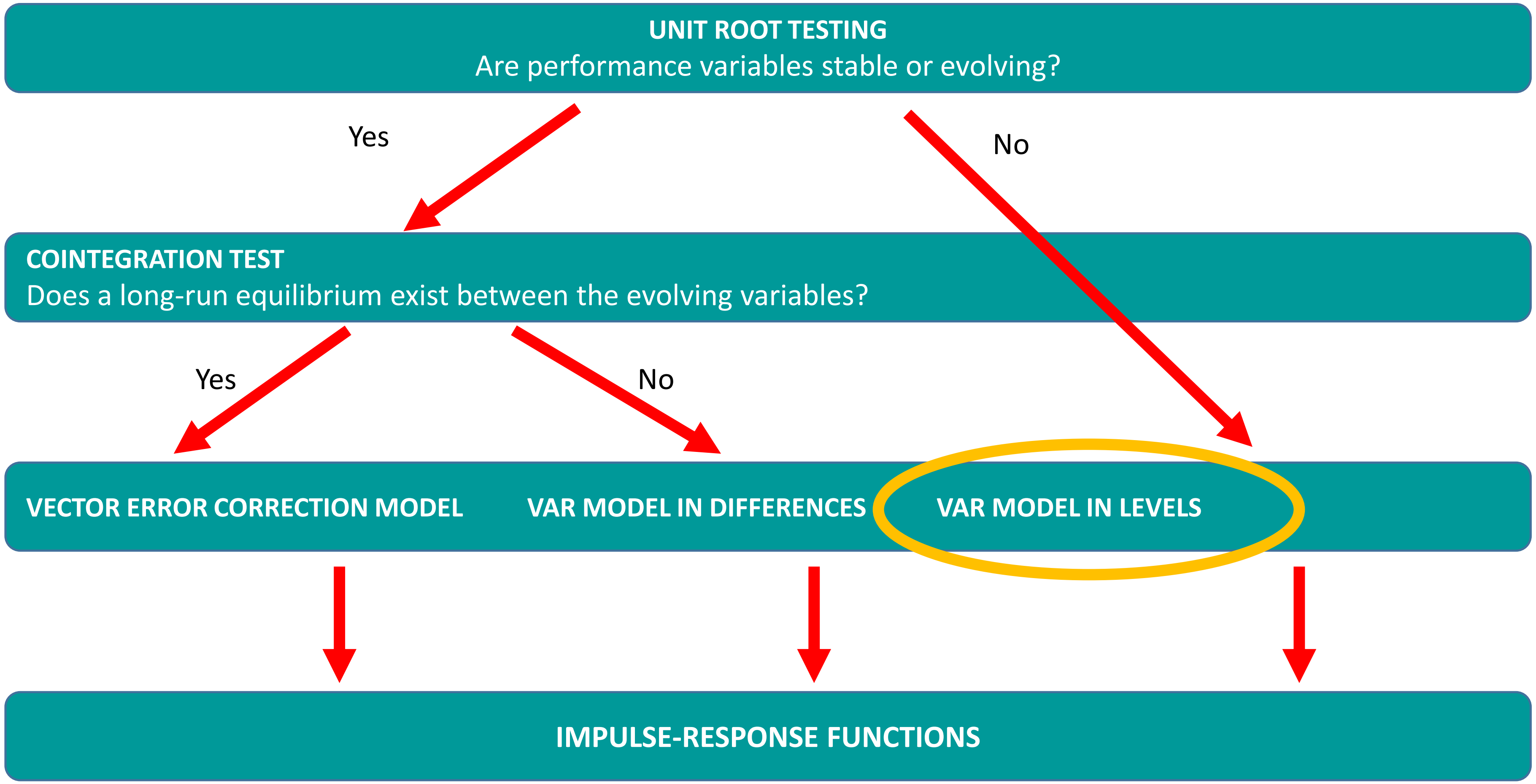
Null hypothesis: the series is non-stationary (evolving)

Alternative hypothesis: the series is stationary (stable)

Compare the test statistic for the parameter π with the Dickey-Fuller critical value.

VAR Model: Estimation

VAR modeling steps



Source: Dekimpe et al. (2006), “Time Series Models in Marketing”, ERIM Report Series, Research in Management.

VAR Model Estimation

Step 1: Determine the optimal lag length such that it minimizes the AIC or BIC criterion.

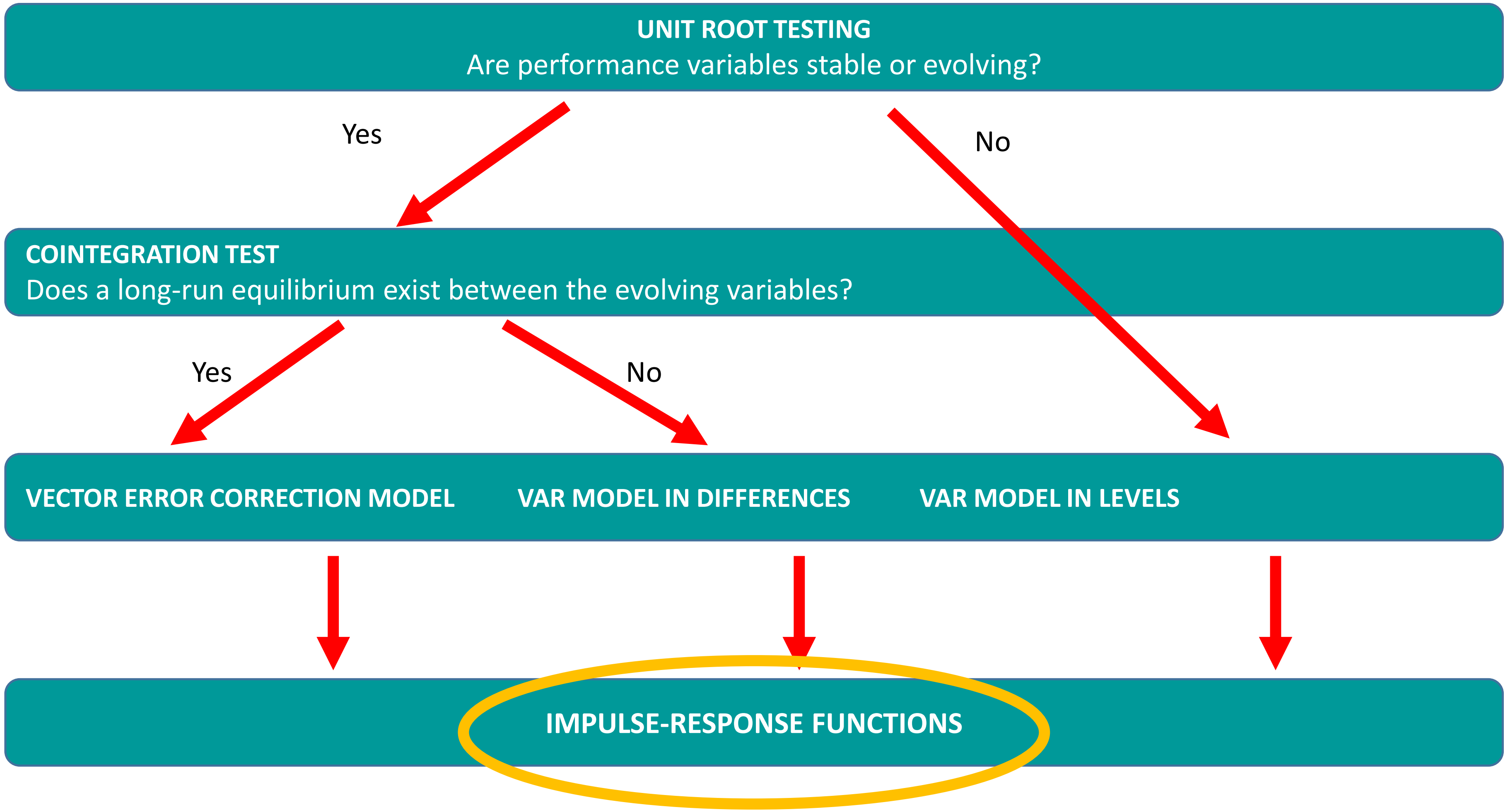
Step 2: Estimate the model using OLS, MLE, GMM (Lutkepohl 2005).

Step 3: Check if the residuals for each equation are free of autocorrelation using simple ACF plots or descriptive statistics (e.g. Breusch-Godfrey test). Add additional lags until the autocorrelation is removed.

Step 4: Drop the variable with the smallest t-statistic (Bruggeman and Lutkepohl 2001).

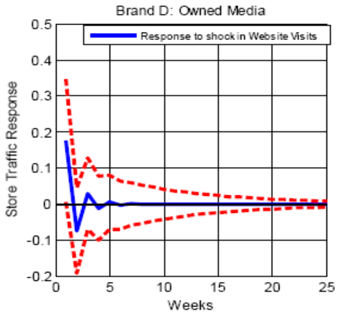
Step 5: Re-estimate the model.

VAR modeling steps

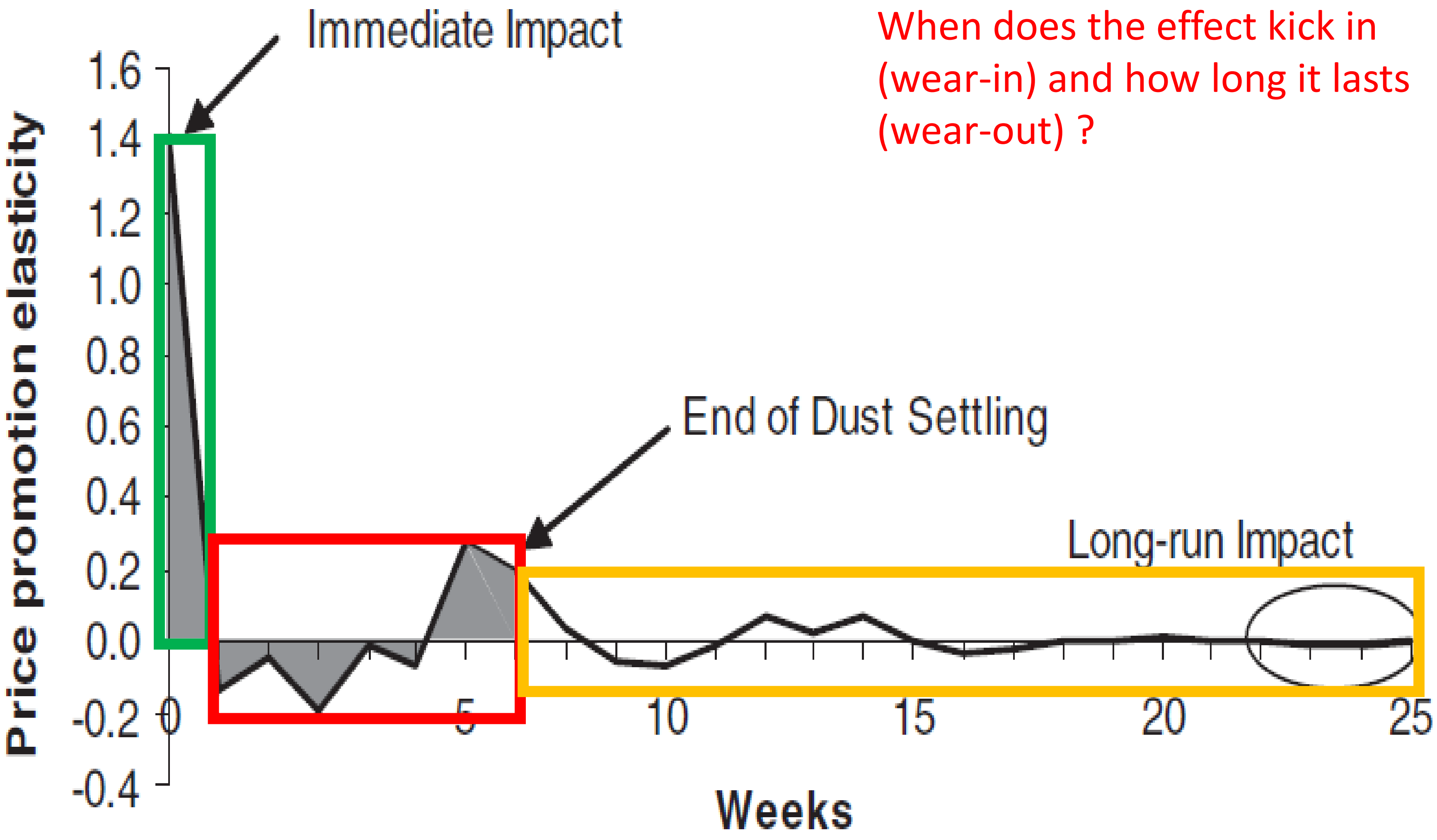


Source: Dekimpe et al. (2006), “Time Series Models in Marketing”, ERIM Report Series, Research in Management.

VAR Models:
Impulse Response Functions



What can I get from IRFs?

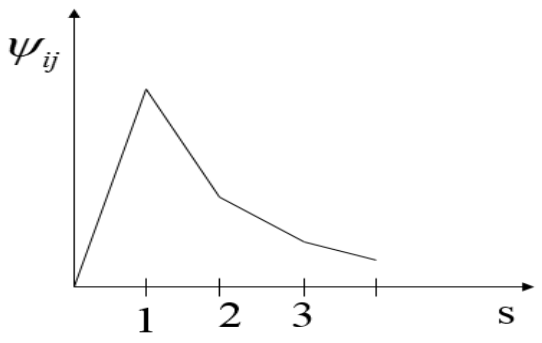


When does the effect kick in (wear-in) and how long it lasts (wear-out) ?

Source: Dekimpe et al. (2005), "Measuring short- and long-run promotional effectiveness on scanner data using persistence modelling," Applied Stochastic Models in Business and Industry, 21, 409-416.

How do I compute the IRFs?

IRF: Response of $y_{i,t+s}$ to one-time impulse in $y_{j,t}$, holding all other variables dated *t* or *earlier* constant.



$$\frac{\partial y_{i,t+s}}{\partial \varepsilon_{jt}} = \psi_{ij}$$

Example: Classical IRFs

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2_{t-1}} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$$

$t < 0 \quad y_{1t} = y_{2t} = 0$

$t = 0 \quad \varepsilon_{10} = 0, \varepsilon_{20} = 1$ (y_{2t} increases by 1 unit)

(no more shocks occur)

$$\begin{bmatrix} y_{10} \\ y_{20} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \textit{(impulse)}$$

$$\begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \phi_{12} \\ \phi_{22} \end{bmatrix}$$

$$\begin{bmatrix} y_{12} \\ y_{22} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\vdots$$

$$\begin{bmatrix} y_{1s} \\ y_{2s} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^s \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \Phi_1{}^s \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Other types of IRFs

Orthogonalized IRFs:

- Since the var-cov is non-diagonal, it is impossible to shock one variable with other variables fixed.
- Use Cholesky decomposition to orthogonalize the shocks.
- The causal order is from the top variables to the bottom variables.

Generalized IRFs:

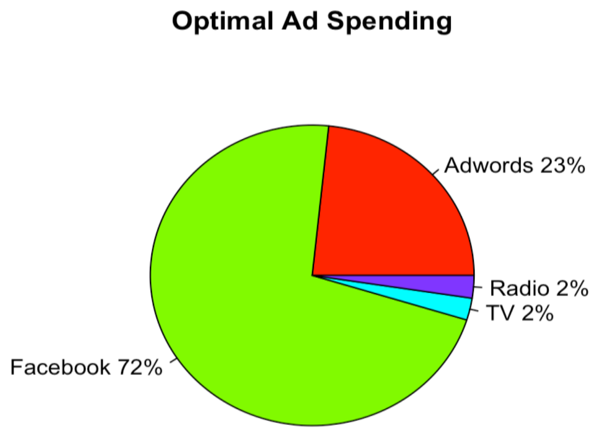
- The shocks are simultaneous.
- GIRF results are insensitive to the order of the model variables.

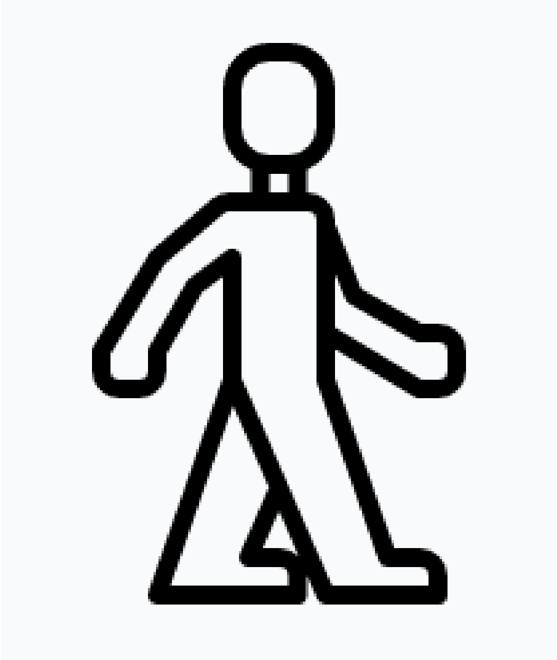
Optimal media allocation

Use the log-term elasticities that come from the IRFs of the log-log model:

$$Optimal\ Allocation_i = \frac{\eta_i}{\sum_{i=1}^{I} \eta_i}$$

η_i is the long-term elasticity of media *i* with respect to performance variable.





Summary

Session summary

Beyond direct effects:

- Feedback and indirect (cross) effects for omnichannel marketing

Dynamic effects:

- **Short- and long-term effects** of marketing actions
- **Wear-in and wear-out effects**

Impulse-response functions

Actionable insights into **media allocation** and marketing strategy

you ready?

1. Do the marketing efforts of ABC kitchen appliances firm pay off?

2. What are the short and long-term effects of flyer and Google adwords campaigns?

3. What is the optimal allocation between offline and online advertising spend?

Thank you!


ANY
QUESTIONS
?