Synchronous bootstrapping of loss reserves

Greg Taylor
Taylor Fry Consulting Actuaries
University of Melbourne
University of New South Wales

Gráinne McGuire
Taylor Fry Consulting Actuaries

ASTIN Colloquium

Zurich

4-7 September 2005

TAYLOR FRY
CONSULTING ACTUARIES

Overview

- We shall be concerned with **forecasting**
  - There are multiple **data sets**
  - There is dependency (correlation) between them (Seemingly Unrelated Regressions)
  - We require estimation of forecast error
    - Separately by **data set**
    - For aggregated **data sets**
  - Forecast error to be estimated by bootstrapping

Overview

- We shall be concerned with **forecasting** (loss reserving)
  - There are multiple **data sets (lines of business (LoBs))**
  - There is dependency (correlation) between them (Seemingly Unrelated Regressions)
  - We require estimation of forecast error
    - Separately by **data set (LoB)**
    - For aggregated **data sets (LoBs)**
  - Forecast error to be estimated by bootstrapping

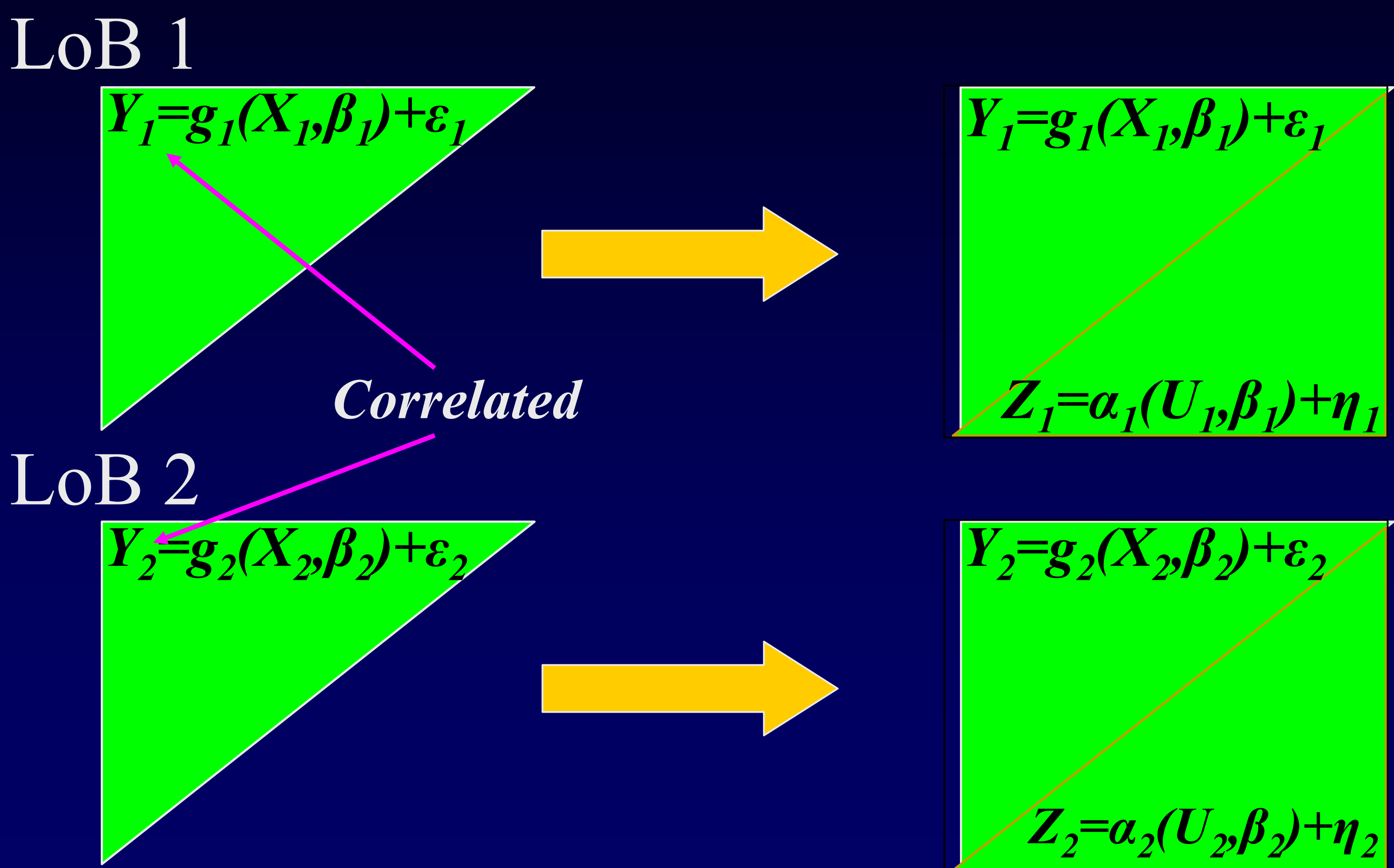
TAYLOR FRY
CONSULTING ACTUARIES

Statement of problem

- Consider an insurance portfolio consisting of LoBs labelled i=1,2,…,I
- Let Z_i denote some technical liability associated with LoB i
  - e.g. loss reserve
- The Z_i are **not** necessarily stochastically independent
- Let $Z=\Sigma_i Z_i$ = **Total Liability** across all LoBs
- Estimate the distribution of Z

TAYLOR FRY
CONSULTING ACTUARIES

Data and model set-up



Sources of correlation

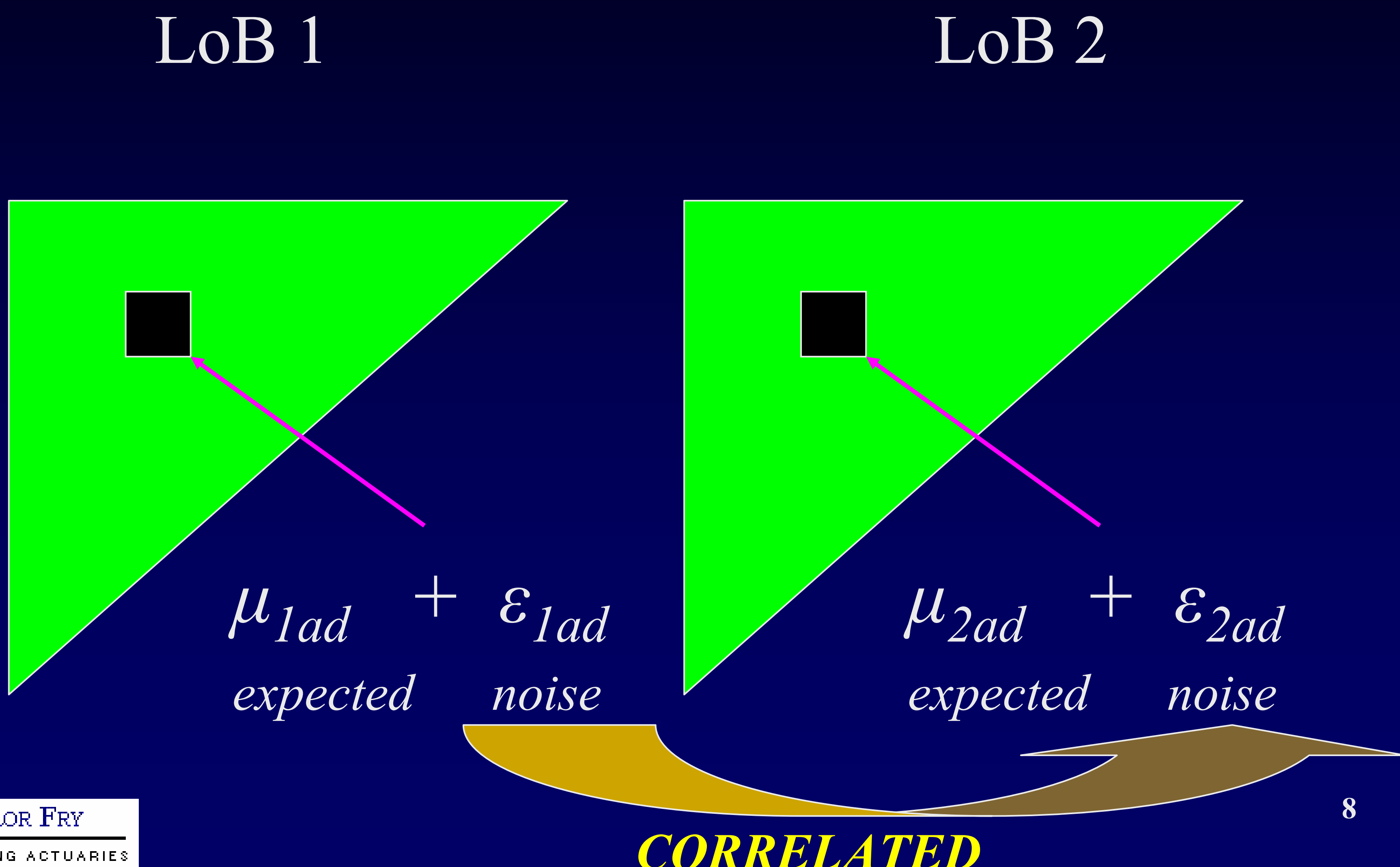
Correlated noise

- $Y_i = \mu_i + \varepsilon_i = g_i(X_i,\beta_i) + \varepsilon_i$
- $Y_j = \mu_j + \varepsilon_j = g_j(X_j,\beta_j) + \varepsilon_j$

- $C_{ij} = \mathrm{Corr}(Y_i,Y_j) = \mathrm{Corr}(\varepsilon_i,\varepsilon_j)$

TAYLOR FRY
CONSULTING ACTUARIES

Correlated noise (cont'd)



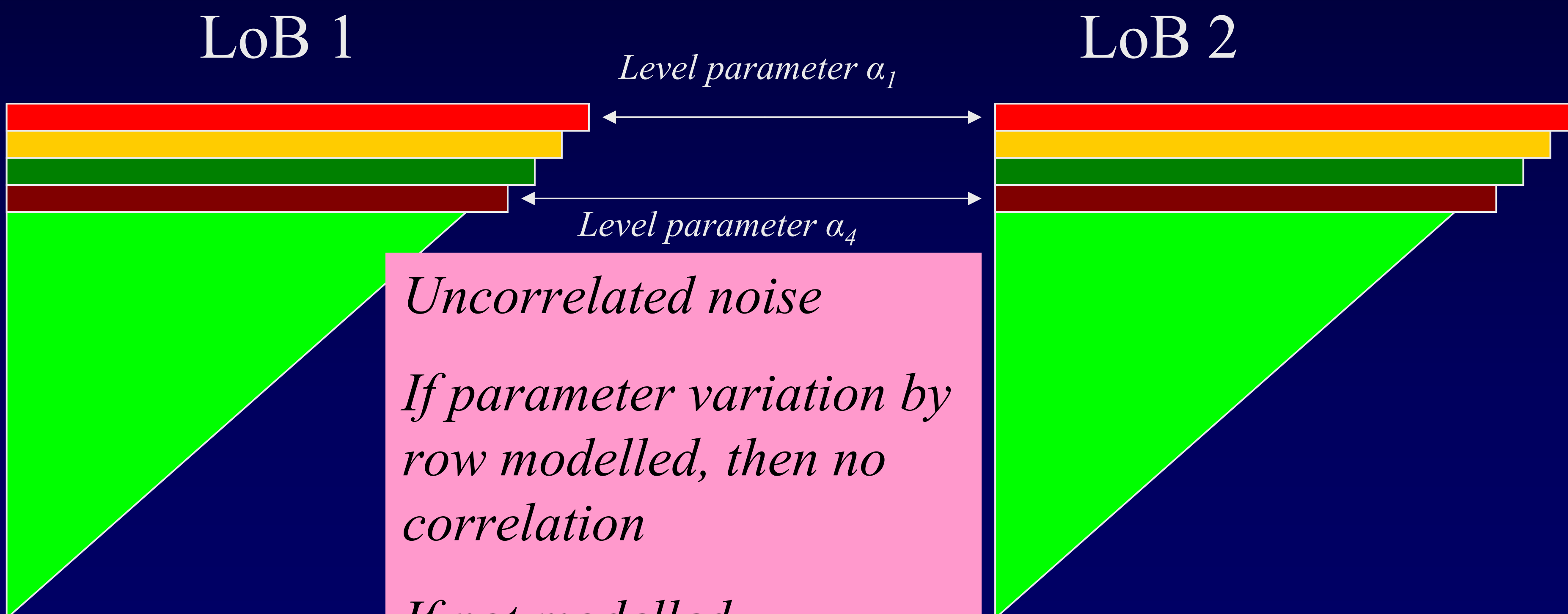
TAYLOR FRY
CONSULTING ACTUARIES

Parameter correlation

- Assumed model
  - $Y_i = \mu_i + \varepsilon_i = g_i(X_i,\beta_i) + \varepsilon_i$
- **BUT** model mis-specified. True model is
  - $Y_i = \mu^+_i + \varepsilon^+_i = g^+_i(X^+_i,\beta_i,\gamma_i) + \varepsilon^+_i$
- where
  - γ_i is an additional vector of unrecognised parameters
  - g^+_i, X^+_i are a function and design matrix that accommodate the unrecognised parameters
  - ε^+_i and ε^+_j are independent
- Note that
  - $\varepsilon_i = \varepsilon^+_i + b_i$ where $b_i = g^+_i(X^+_i,\beta_i,\gamma_i) - g_i(X_i,\beta_i)$ [bias]
  - "Covariance" $E[\varepsilon_i \varepsilon_j^T] = b_i b_j^T$
    - Mis-specification creates bias and correlation

Parameter correlation (cont'd)

- Shared row parameters
  - With **unrecognised** variation between rows

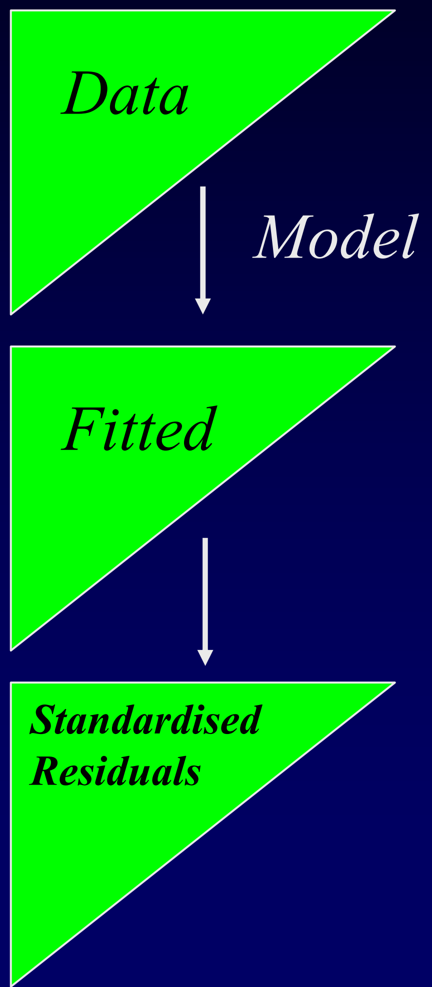


Uncorrelated noise

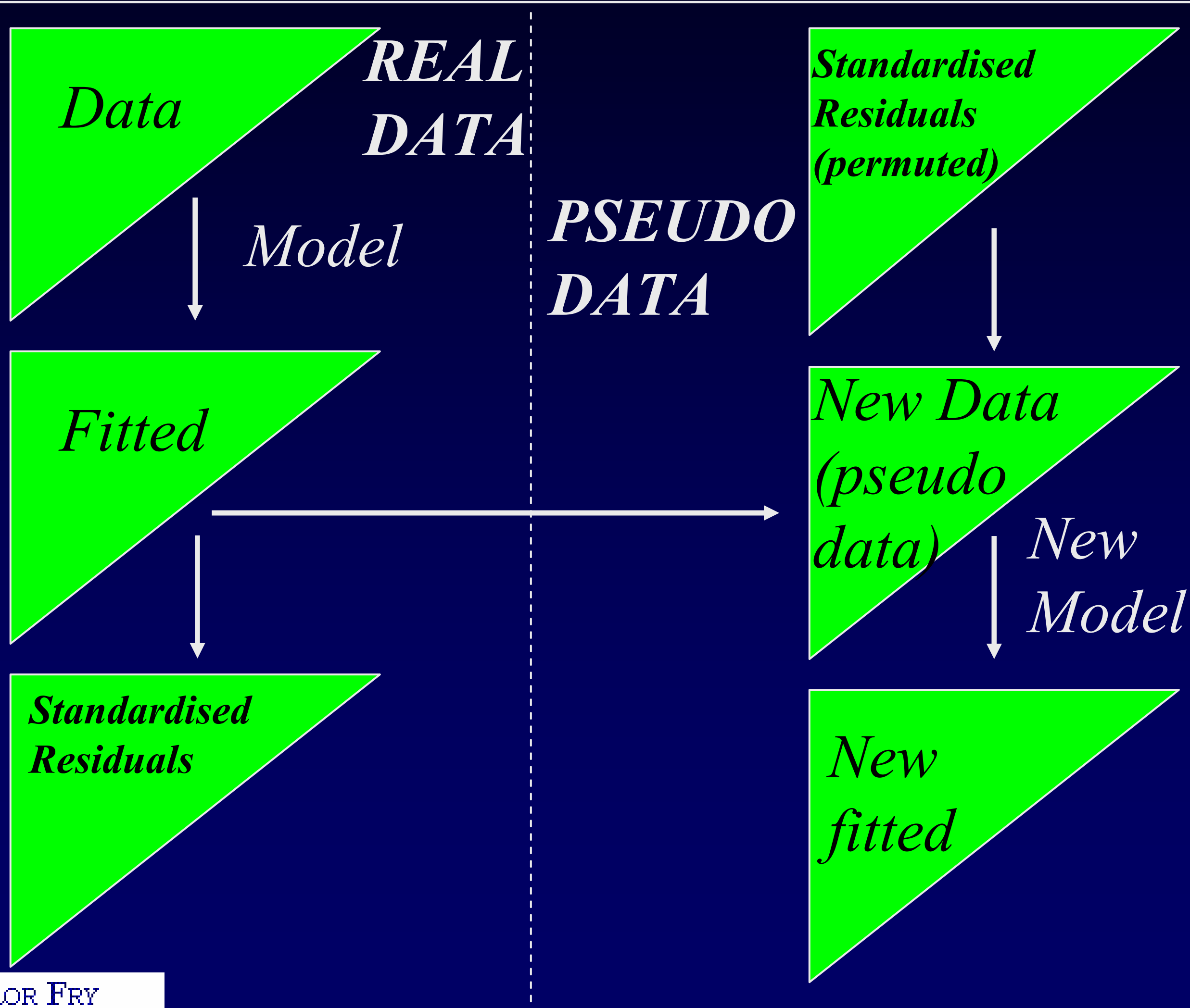
If parameter variation by row modelled, then no correlation

If not modelled, correlation created

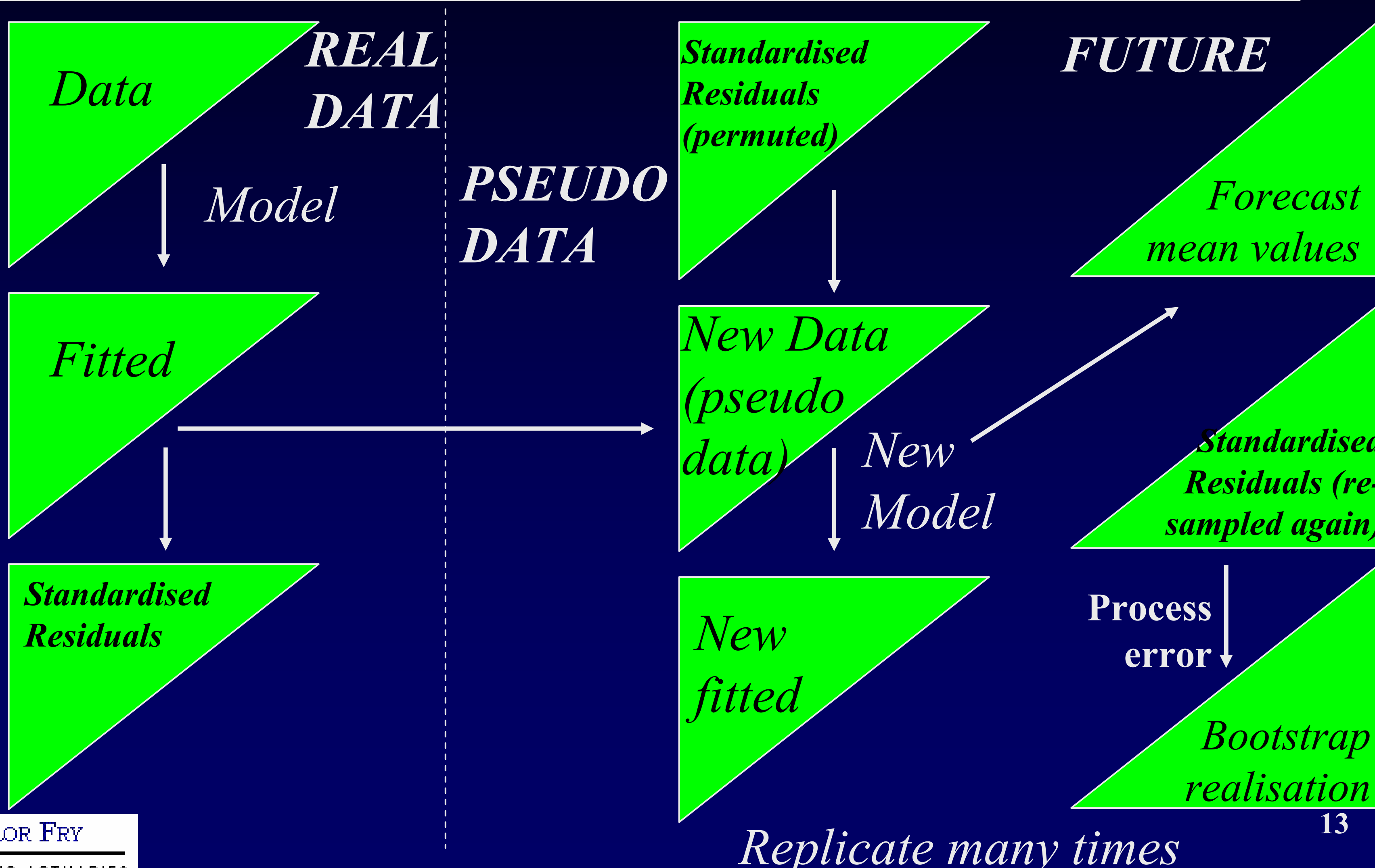
Conventional bootstrap re-visited



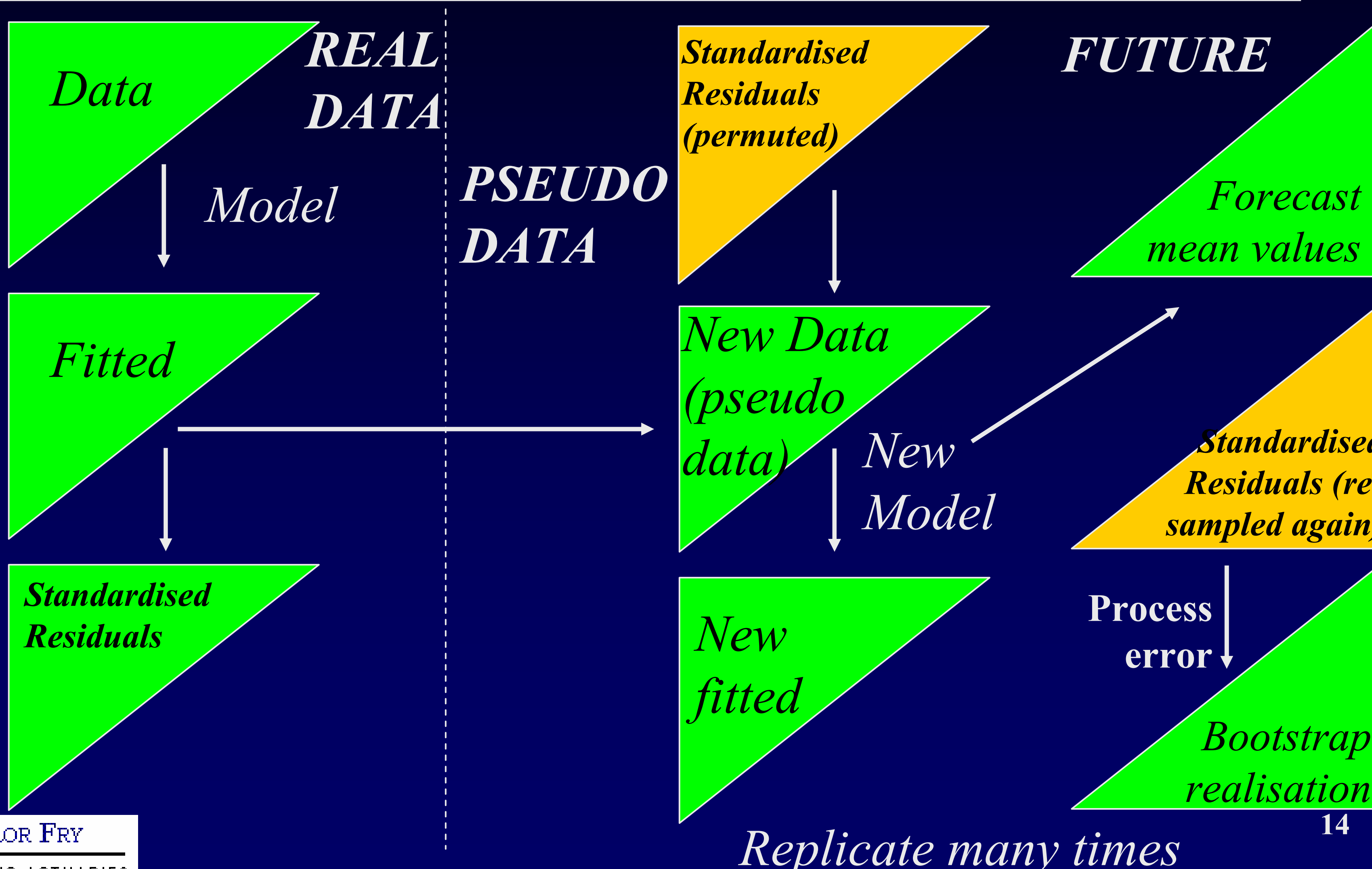
Conventional bootstrap re-visited



Conventional bootstrap re-visited



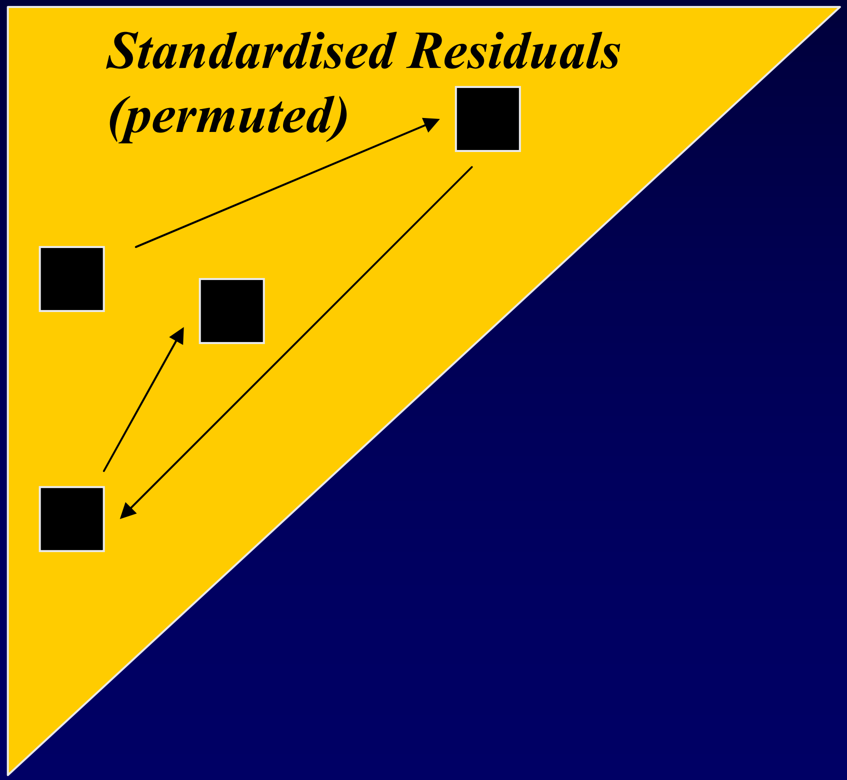
Conventional bootstrap re-visited



TAYLOR FRY
CONSULTING ACTUARIES

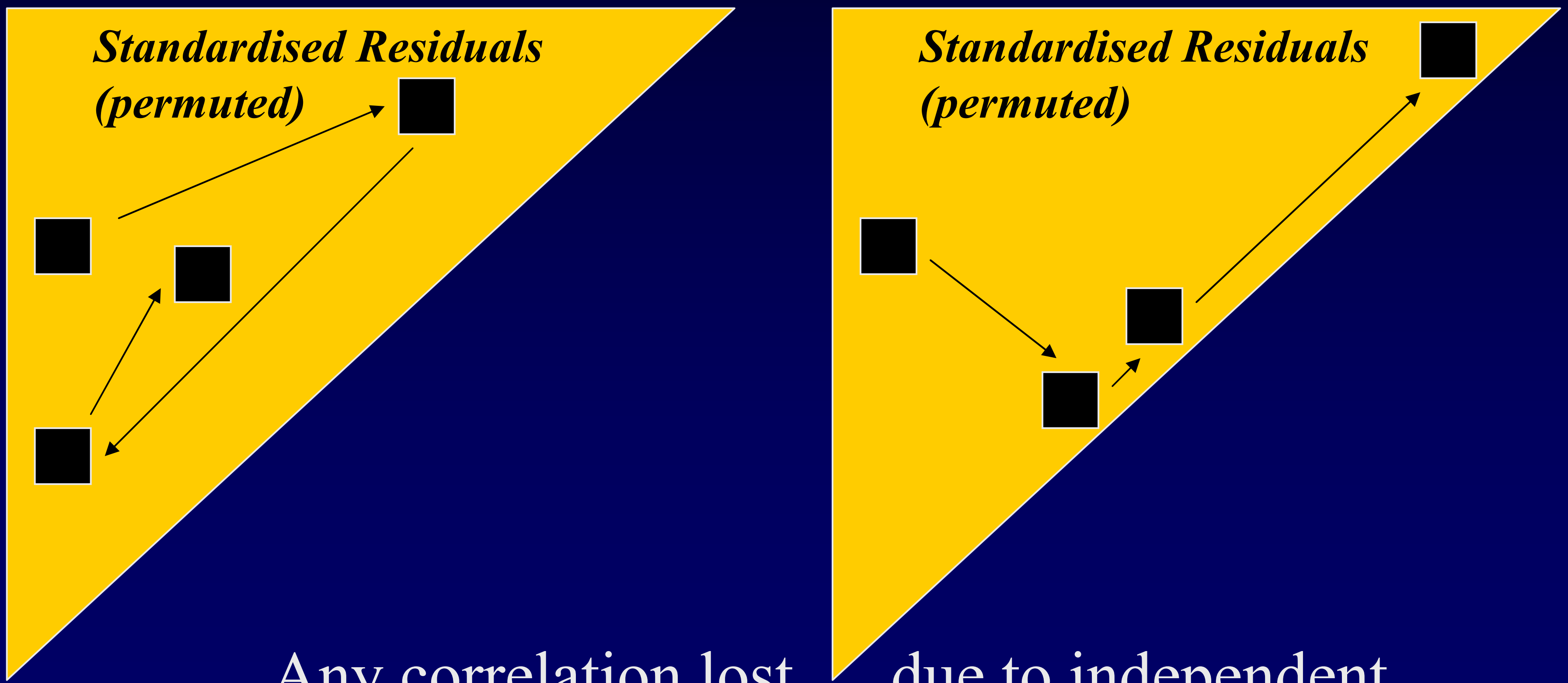
Conventional bootstrap re-visited

Example



Conventional bootstrap of two data sets

DATA SET 1 DATA SET 2



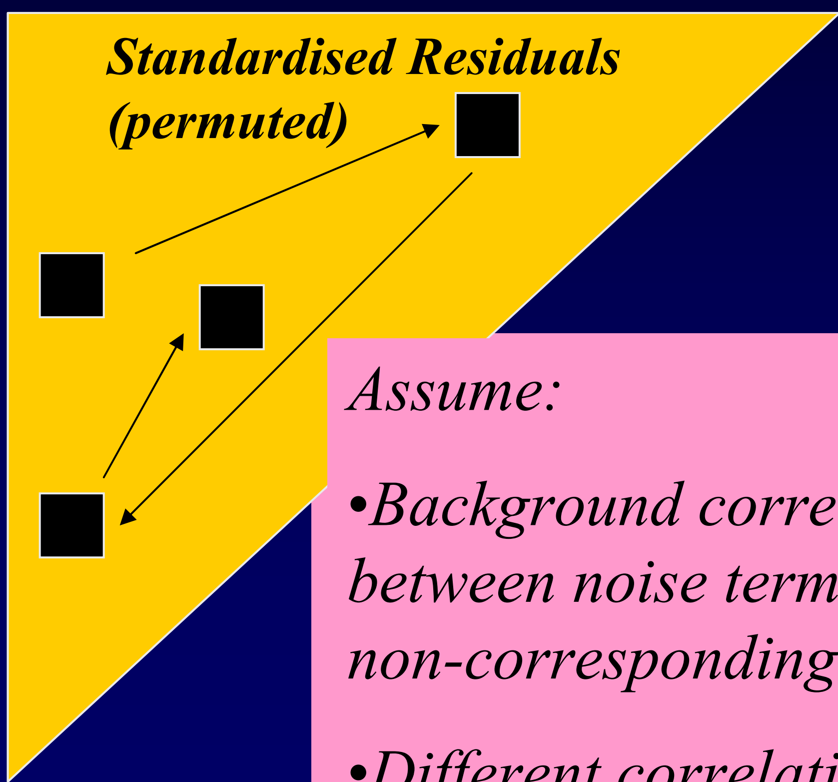
Any correlation lost due to independent permutation of residuals

Synchronous bootstrapping

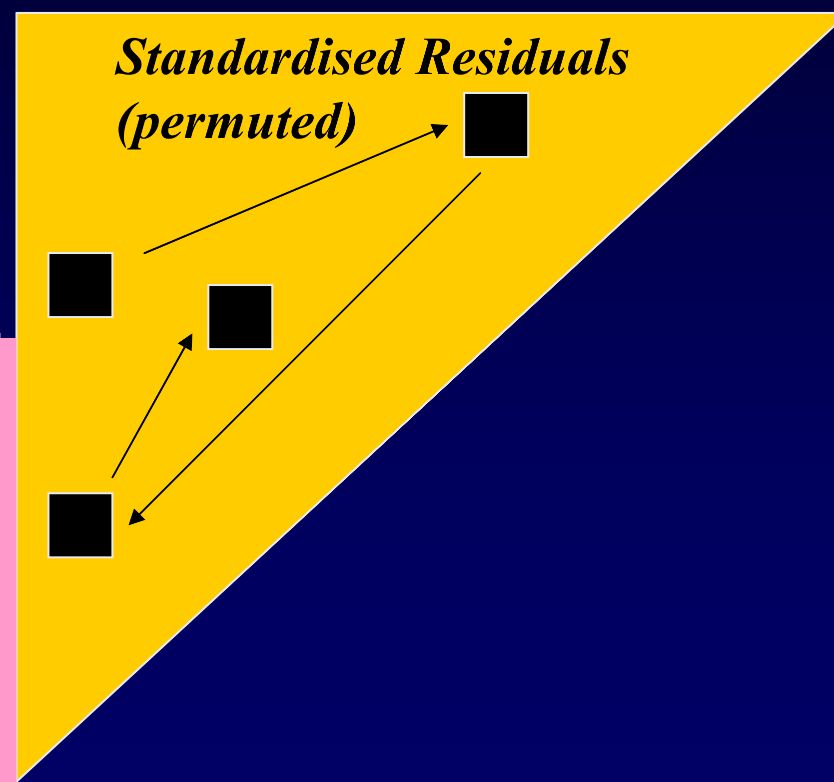
- Conventional (independent) bootstrapping of correlated data sets destroys the correlations
- The solution is to synchronise the permutations applied to the residuals of the different data sets
  - Kirschner, Kerley & Isaacs (CAS, 2002)
- The form of synchronisation depends on the assumed form of correlation

Synchronous bootstrap – correlated noise

DATA SET 1



DATA SET 2

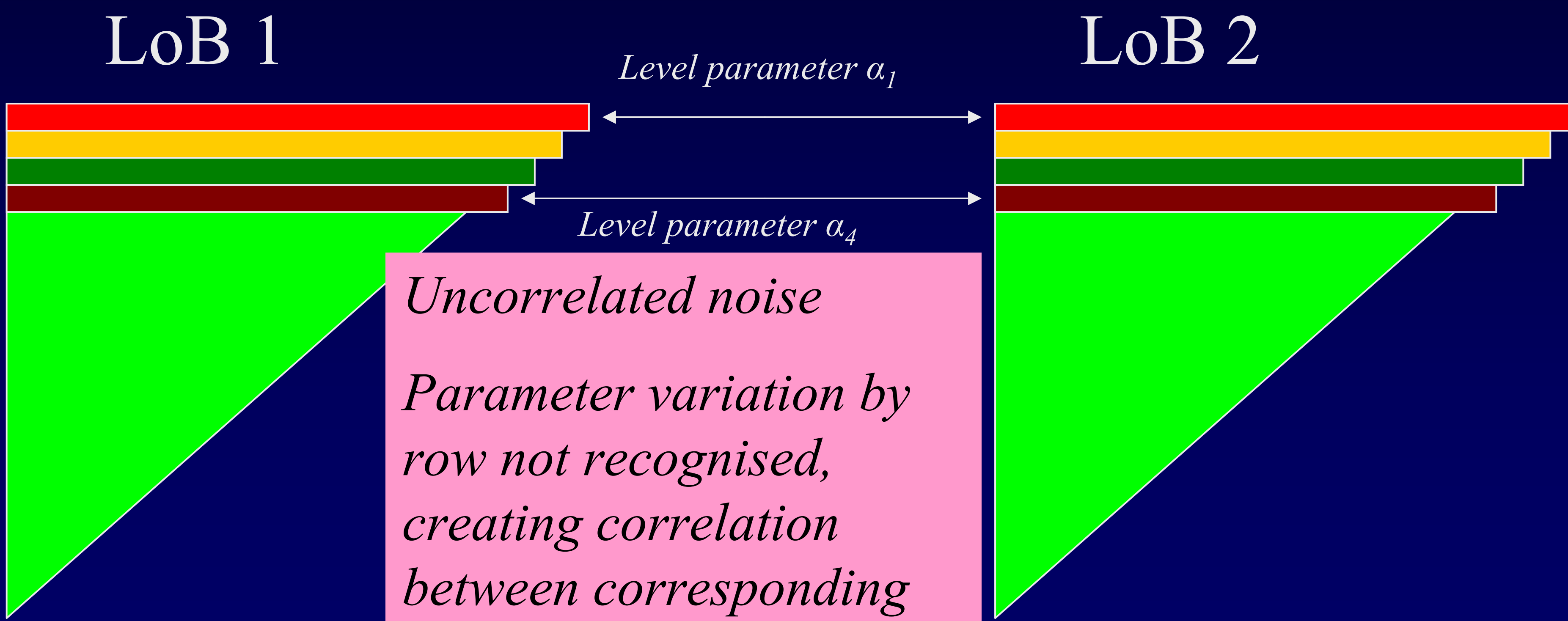


Assume:

- *Background correlation between noise terms of non-corresponding cells*
- *Different correlation for corresponding cells*

TAYLOR FRY
CONSULTING ACTUARIES

Synchronous bootstrap – correlated (shared) row parameters

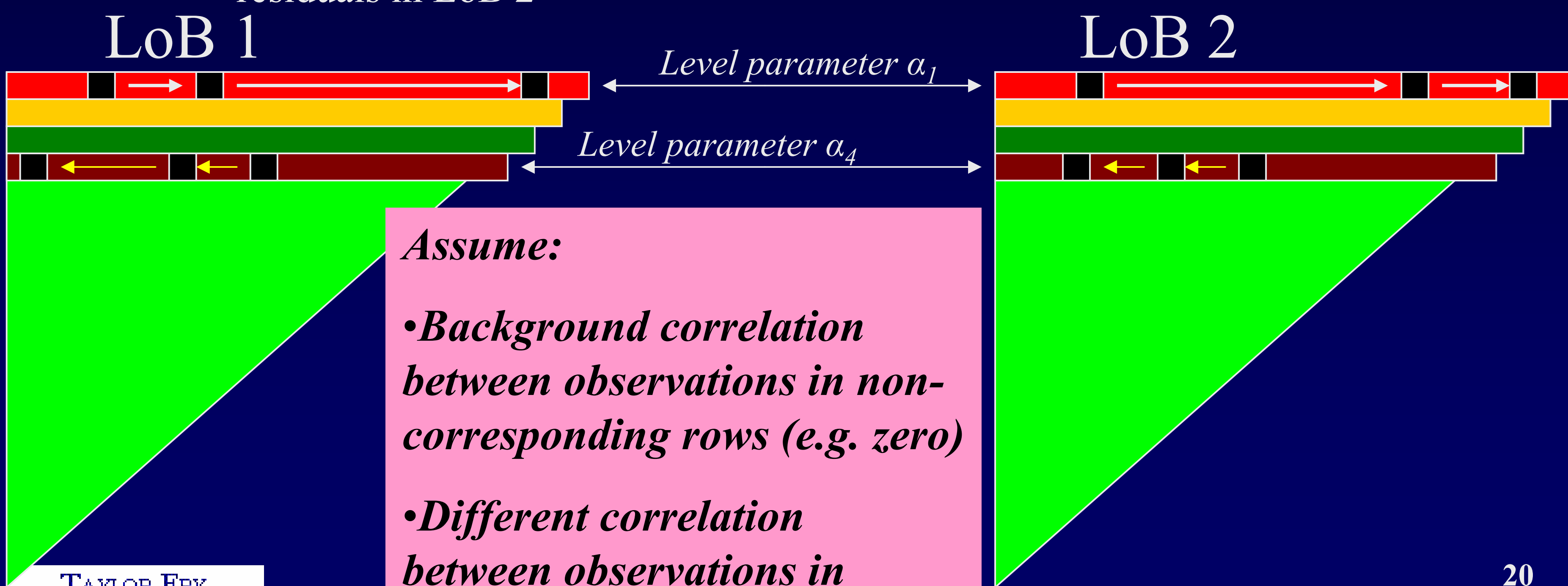


Uncorrelated noise

Parameter variation by row not recognised, creating correlation between corresponding rows of different data sets

Synchronous bootstrap – correlated (shared) row parameters

- Synchronise by restricting permutations to within rows in each data set (specific permutations need not be synchronised)
  - High (low) residuals in LoB 1 will tend to be associated with high (low) residuals in LoB 2



Assume:

•Background correlation between observations in non-corresponding rows (e.g. zero)

•Different correlation between observations in corresponding rows

Numerical results

Point-wise bootstrap

- 3 triangles
  - Each 20x20
  - Triangles have identical expectations
  - Rows within triangles have identical expectations
  - Each row's expectation follows a Hoerl curve (PPCI)
  - Individual cells gamma distributed about expectations
  - Gamma distributions for corresponding cells of different triangles subject to correlation of about 80%
  - Otherwise independent (within and between triangles)
  - Correlated noise
  - Point-wise synchronised bootstrap

Point-wise bootstrap (cont'd)

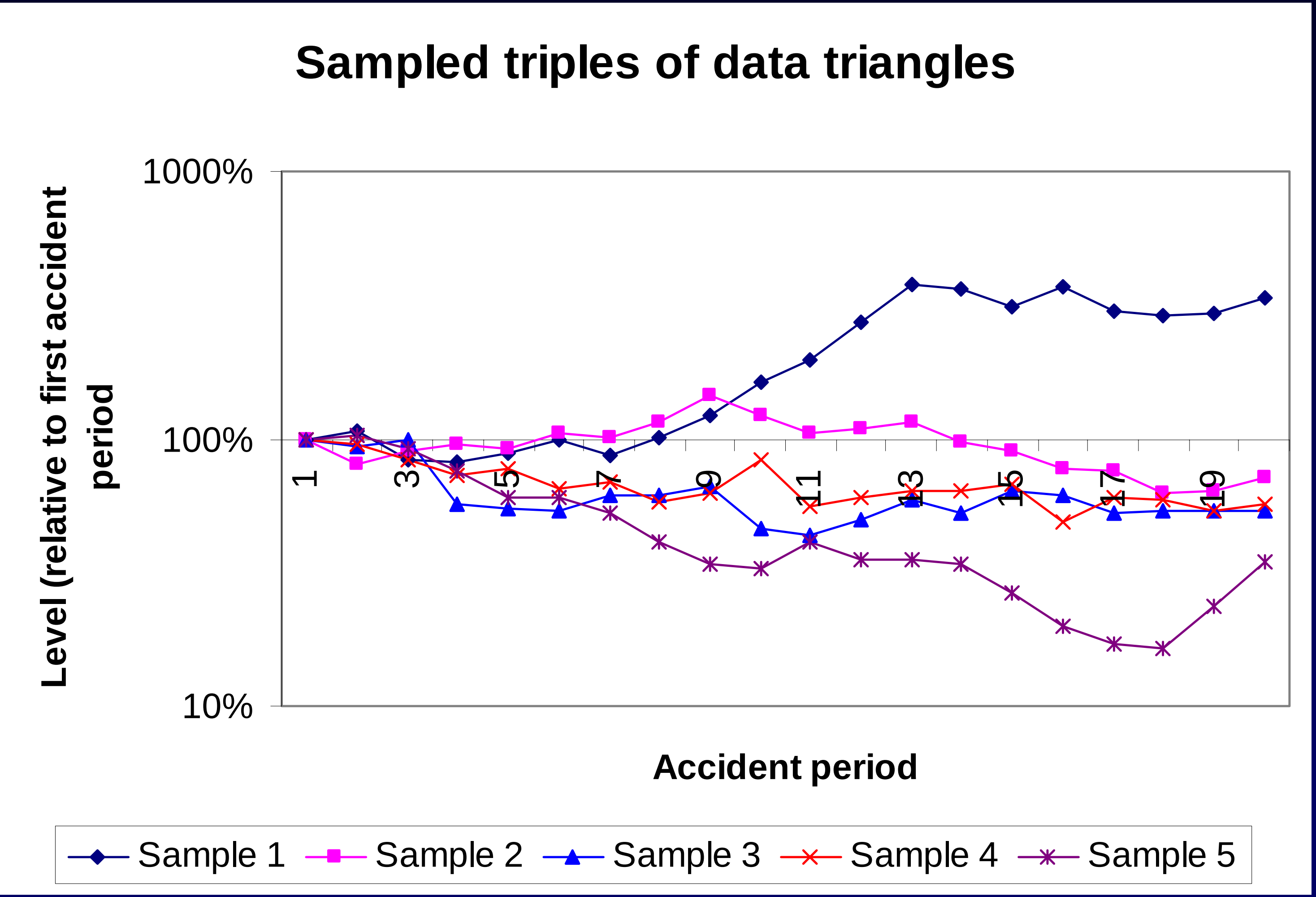
Basis of estimation	Pair-wise correlation of LoB loss reserves	CoV of aggregate loss reserve across 3 LoBs
True (simulation)	0.81	5.4%
Independent bootstrap	-0.00	3.0%
Synchronous point-wise bootstrap	0.79	5.0%

TAYLOR FRY
CONSULTING ACTUARIES

Row-wise bootstrap

- Same data generation (of 3 data sets) as before **except that**
  - No correlated noise
  - Expected level of Hoerl curve follows **geometric random walk** through accident periods
    - Same level for each data set for given accident period
- Model specification deliberately overlooks variation in row parameter
  - (Row) parameter correlation
  - Row-wise synchronised bootstrap

Row-wise bootstrap –examples of data sets



Row-wise bootstrap – efficiency measurement

- Descriptor of sample = ratio R

$$= \mathrm{Var}[\Sigma_i Z_i] / \Sigma_i \mathrm{Var}[Z_i]$$

$$= 1 \text{ for independent } Z_i$$

$$= 3 \text{ for fully correlated } Z_i$$

- Efficiency measure =

$$\frac{(\text{Estimated } R - 1)}{(\text{True } R - 1)}$$

Row-wise bootstrap (cont'd)

Sample	Descriptor	Efficiency measure
1	2.98	73%
2	2.48	60%
3	2.77	24%
4	2.71	26%
5	2.86	103%