Sample size planning using Predictive Accuracy Analysis

For (V)AR(1) models in the context of N=1

10/06/2023

Jordan Revol
Ginette Lafit
Eva Ceulemans

KU LEUVEN

fwo

VAR(1) models

- For N=1 and 2 variables:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_t = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_{t-1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

with: $$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\varepsilon 1} & \sigma_{\varepsilon 1 \varepsilon 2} \\ \sigma_{\varepsilon 2 \varepsilon 1} & \sigma^2_{\varepsilon 2} \end{bmatrix} \right]$$

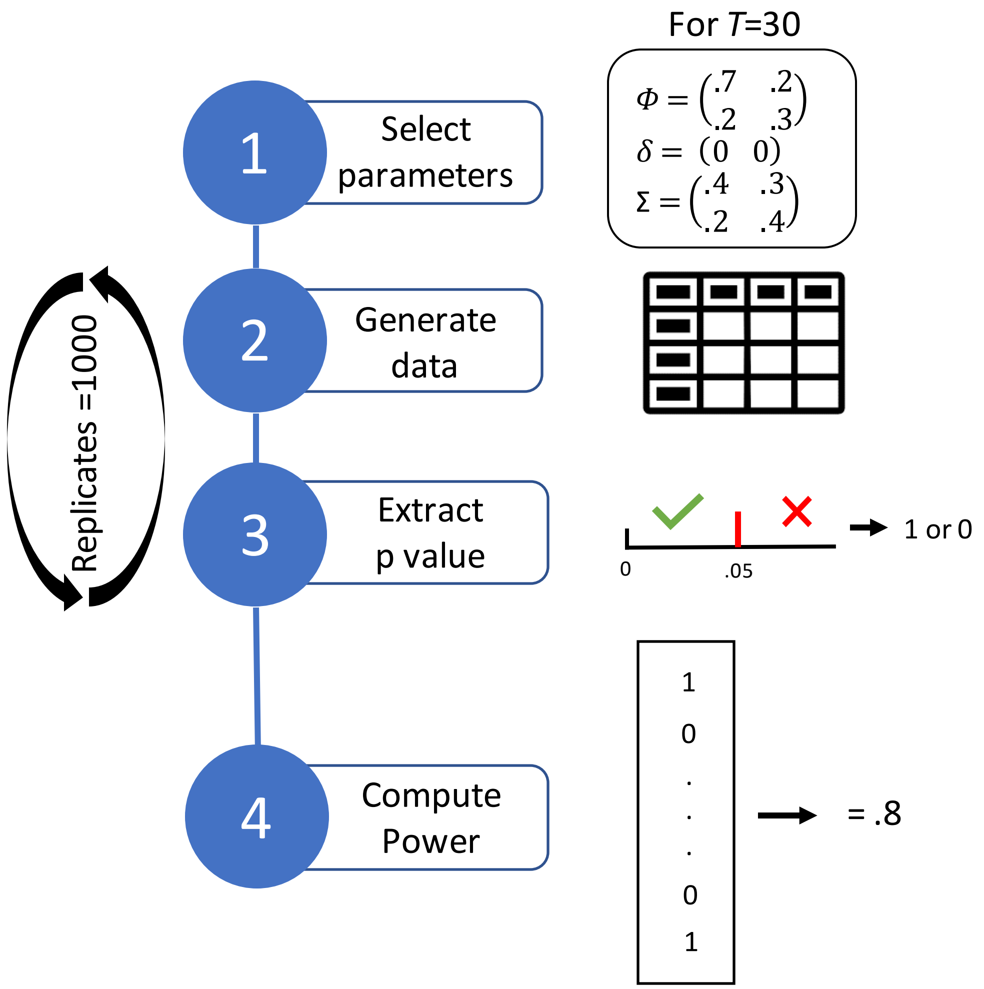
➔Errors follow multivariate normal distribution with **variance-covariance Σ**

How many timepoints?

- Power

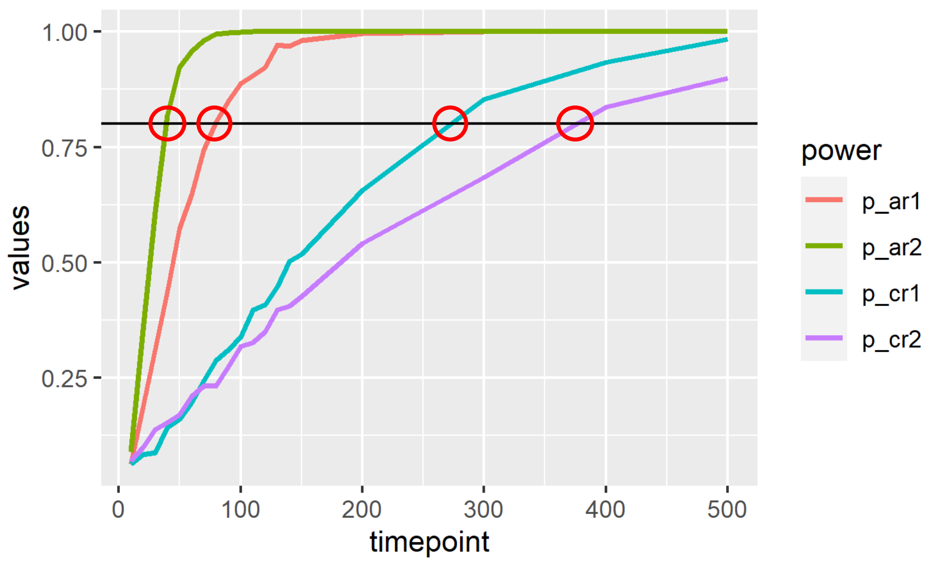
Simulation-based approach

Power analysis



How many timepoints?

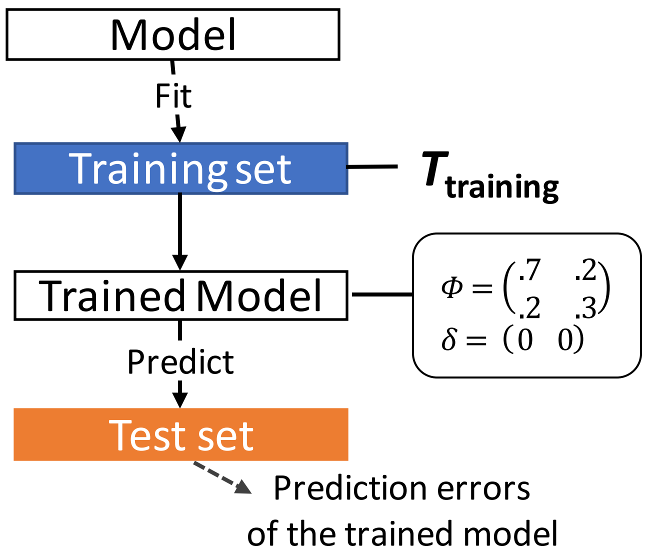
- Power:
  - Parameter specific
  - Focus on the effect(s) of interest



Example for 2 variables

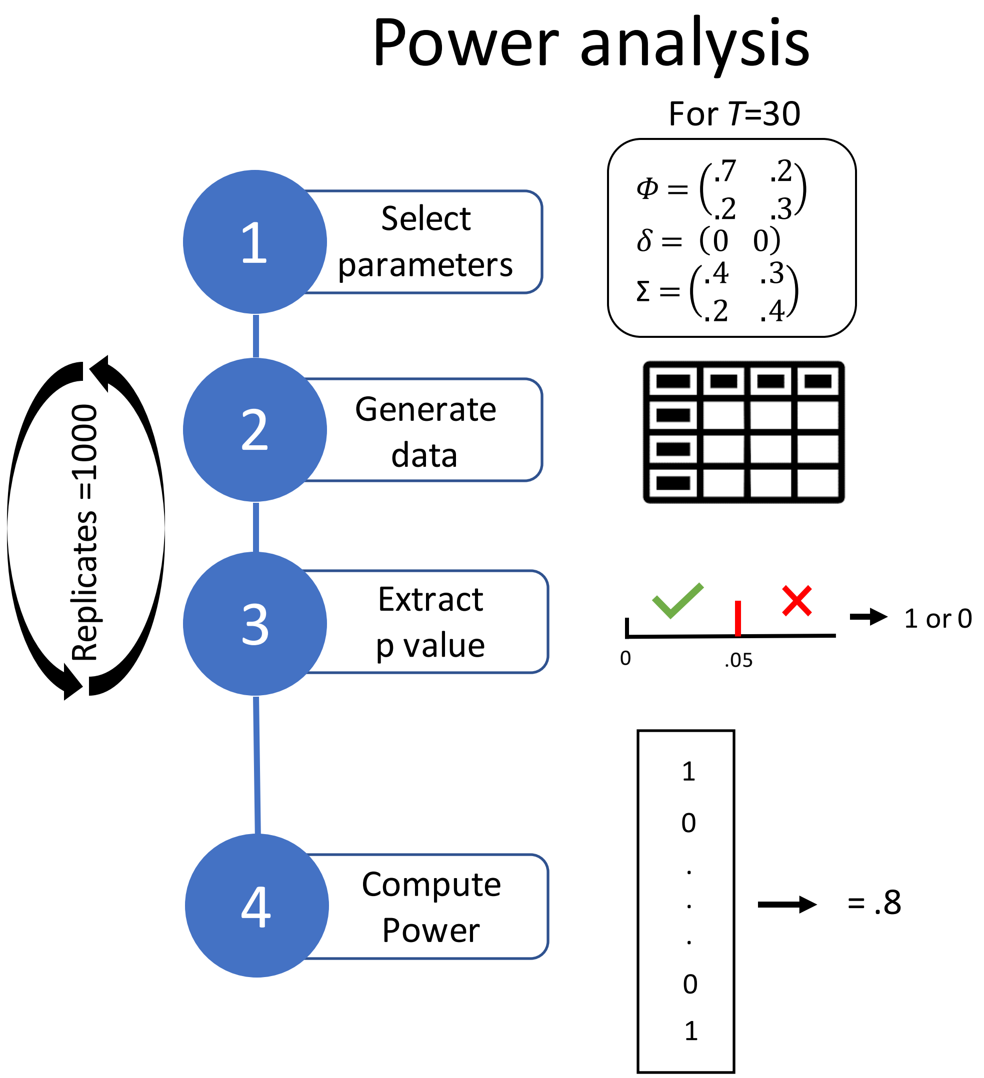
How many timepoint?

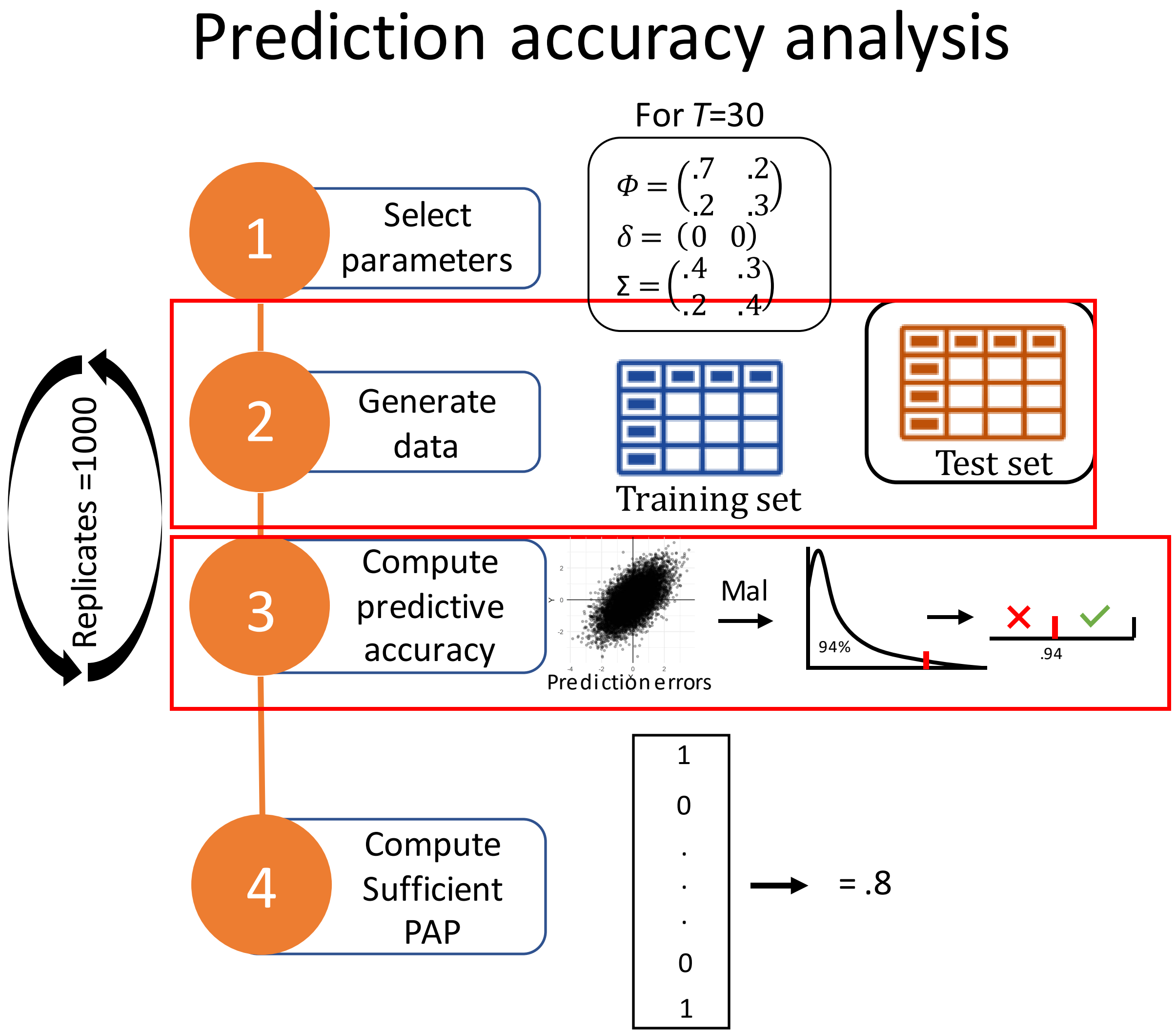
- Power:
  - Parameter specific
  - Focus on the effect(s) of interest
- Prediction accuracy:
  - Focus on the whole model:
    “how well will my model perform on unseen data?”
  - Usually MSPE ➔ Issue
  - ↗ T_{training} = ↗ Predictive accuracy



- PAA: Optimize the number of timepoints ($\boldsymbol{T}_{\textbf{training}}$)
  to have a ***good probability*** to achieve
  a ***good* predictive accuracy**

Simulation-based comparison

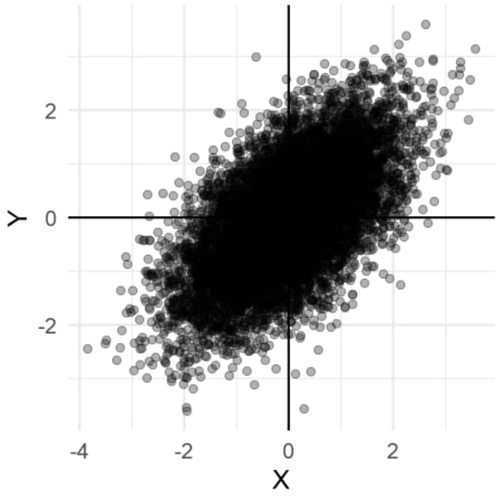




➔ Modification of step 2 and 3

Step 3 of PAA

- Steps:
  - **3.1** Compute Mahalanobis distance using true Σ *(standardization)*
  - **3.2** Compute proportion of prediction errors < 95th percentile of the χ^2 distribution with df = #variables



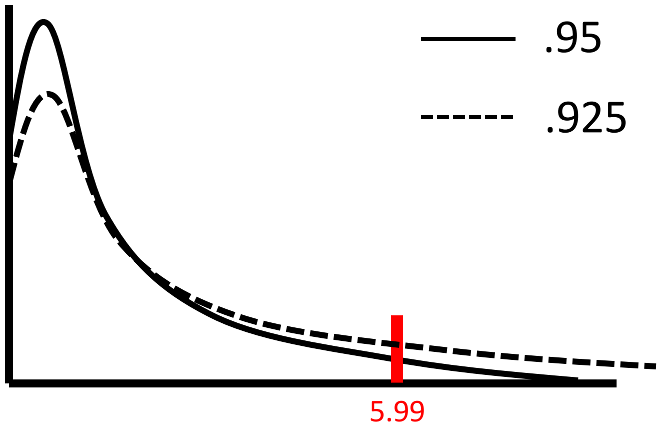
Prediction errors

Step 3.1

Mahalanobis distance

$$D^2 = E \cdot \Sigma^{-1} \cdot E$$

Step 3.2

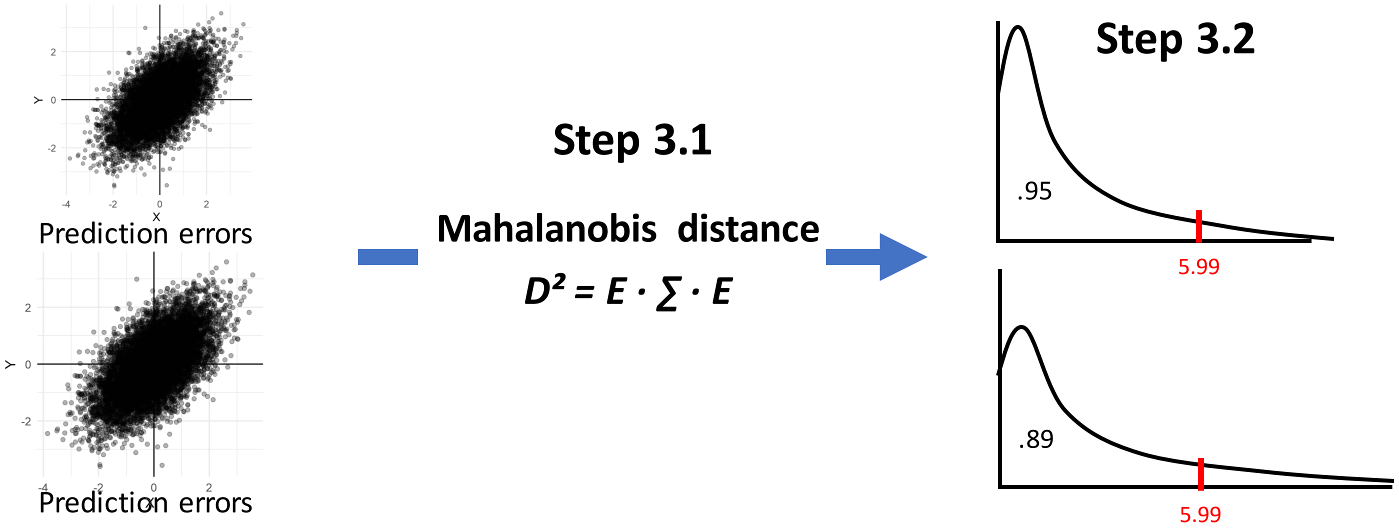


Prediction errors

Step 3 of PAA

- Steps:
  - 3.1 Compute Mahalanobis distance using true Σ (standardization)
  - 3.2 Compute proportion of prediction errors < .95 quantile of χ^2(#vars)

- For high $T_{training}$:
- For smaller $T_{training}$:



Steps 3 and 4 of PAA

- Steps:

Good predictive accuracy

3.1 Compute Mahalanobis distance using true Σ *(standardization)*

3.2 Compute proportion of prediction errors < .95 quantile of $\chi^2(\#\text{vars})$

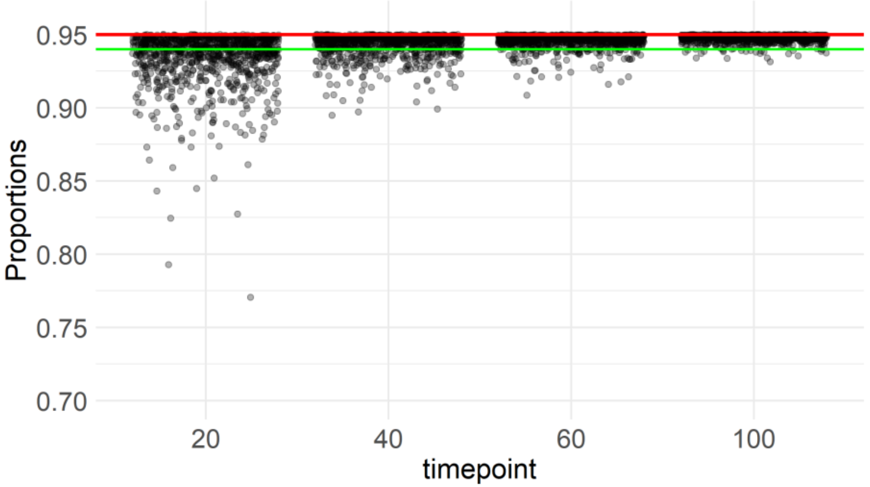
3.3 Define performance threshold: .94

Probability to reach it

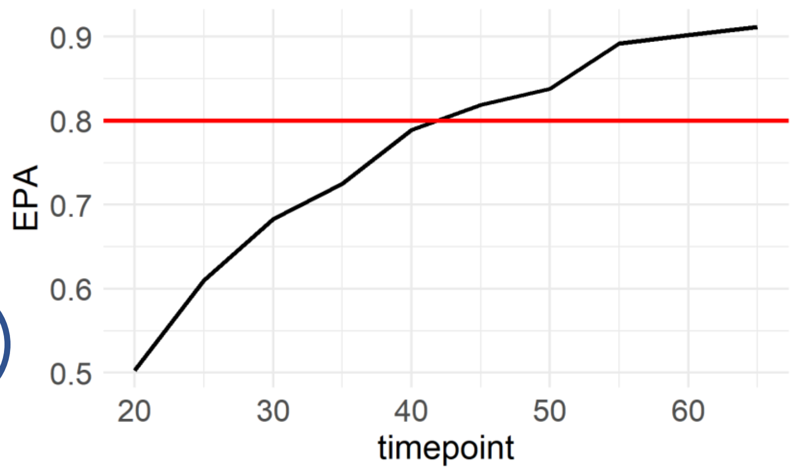
4. Compute expected predictive accuracy (EPA)

- Looking for **.8** proportion of replicates that reach performance

3.3



4.



Results

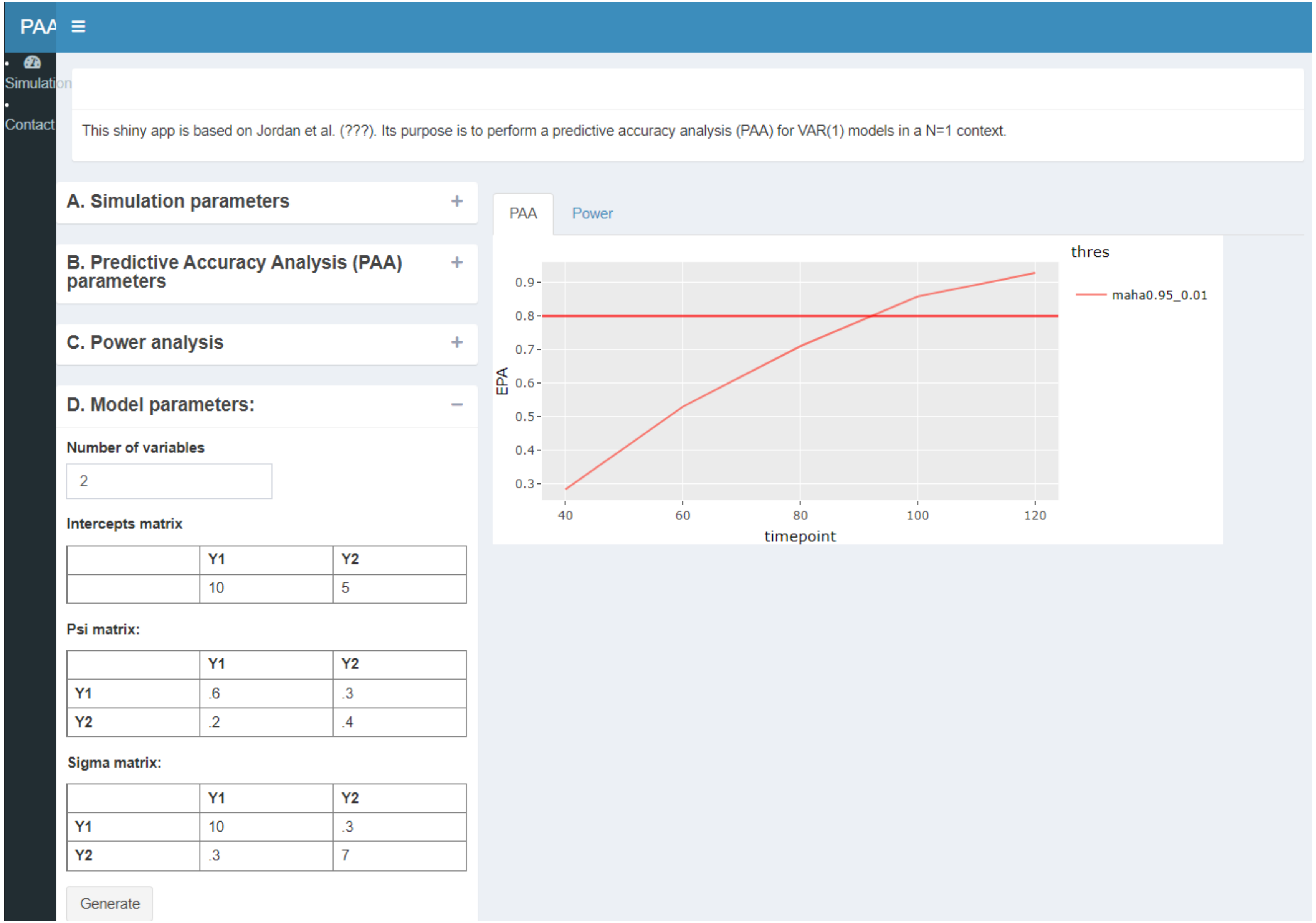
	Power	Predictive accuracy (PAA)
Complexity of the model (#vars)	-	↘°
Auto-regressive	↗*	↘°
Cross-regressive	↗*	↘°
Intercept	↗*	-
Variance	↘	-`
Covariance	↘	-`

- Complement power
- **Warning:** Predictive purpose

° *Whole model*
** Parameter specific*
` *Standardized*

Apps

- R: Shiny app
- Julia: Dash app

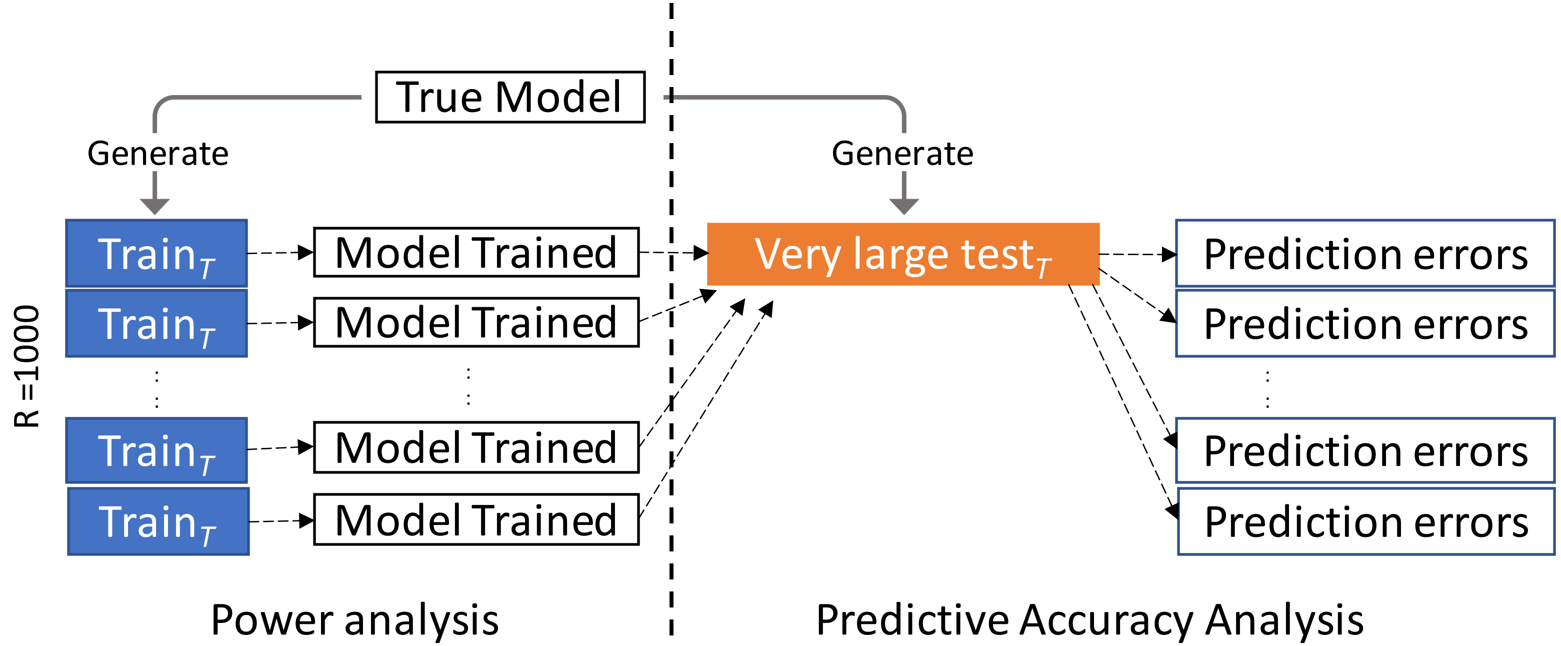


Thanks for your attention

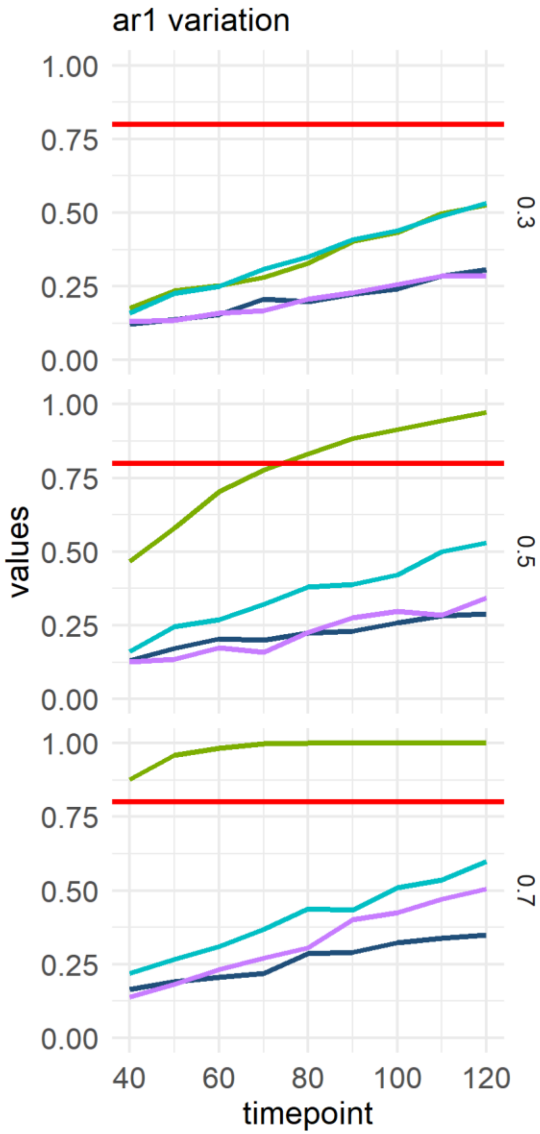
jordan.revol@kuleuven.be

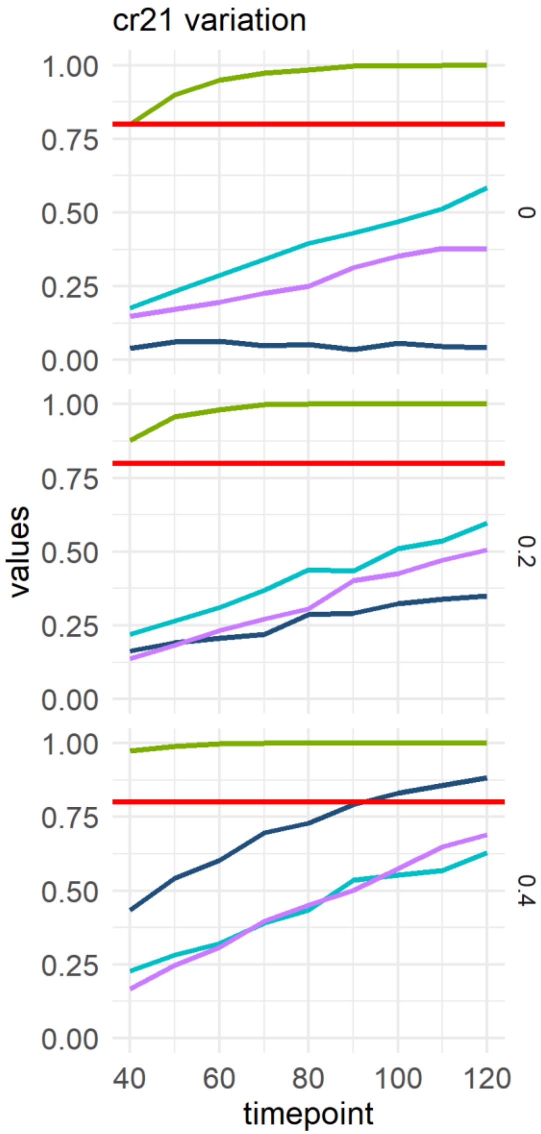
Step 2 of PAA

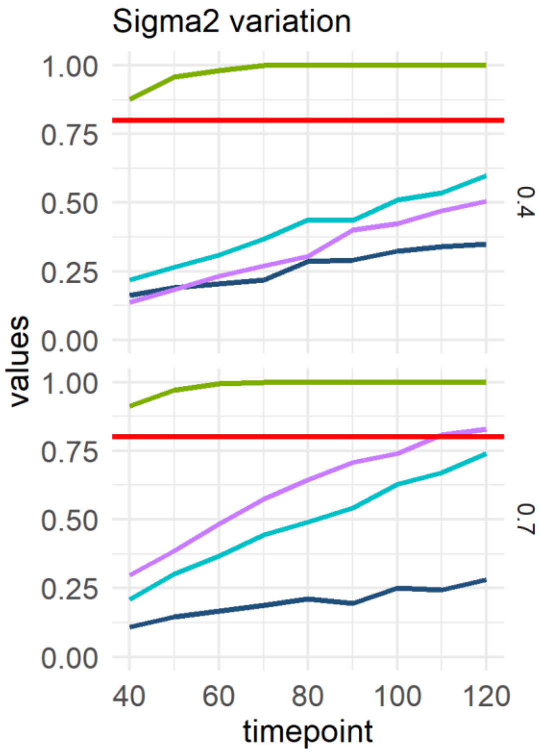
- Generate data

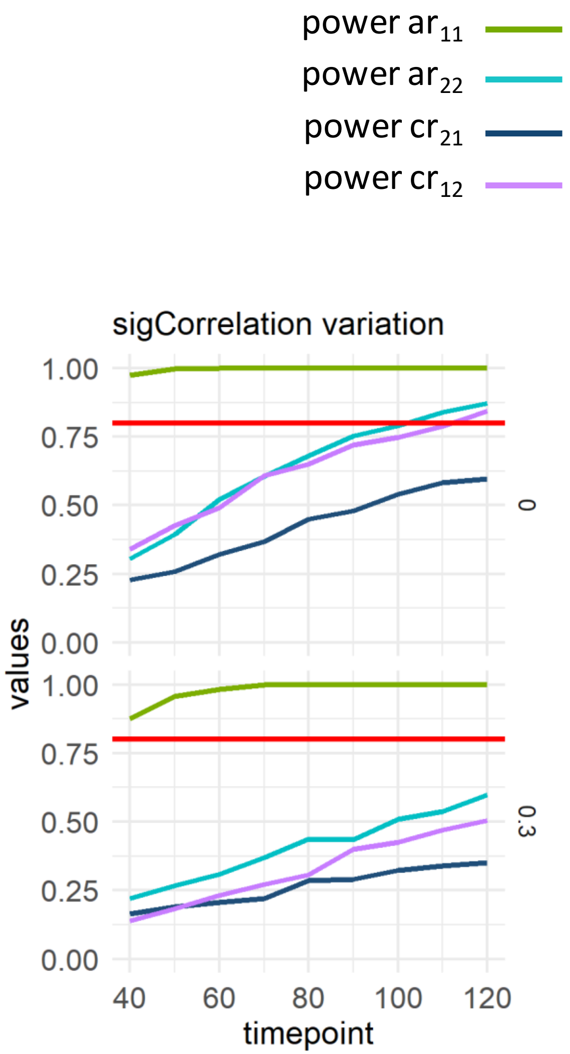


Results: Parameters’ influence





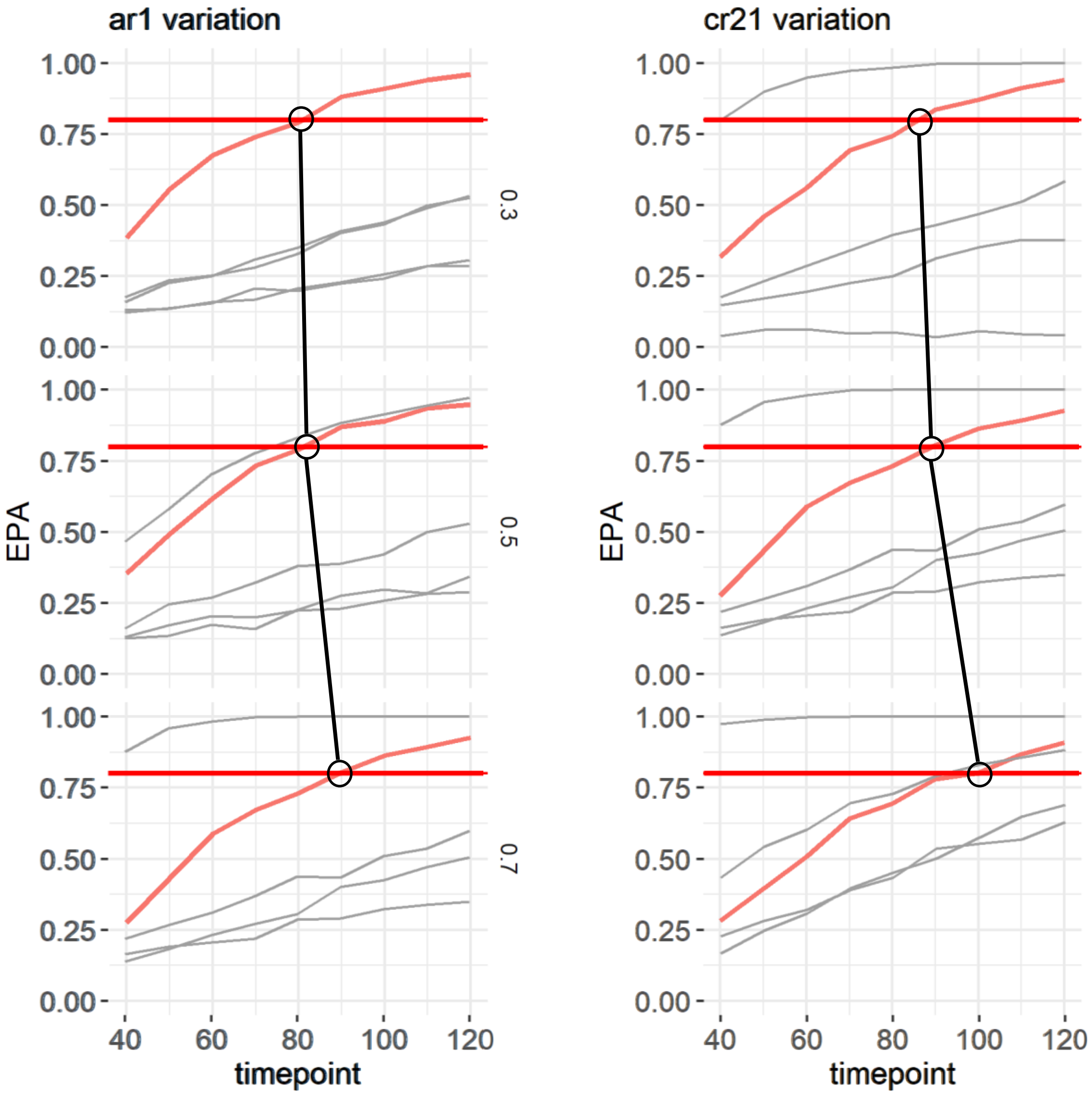


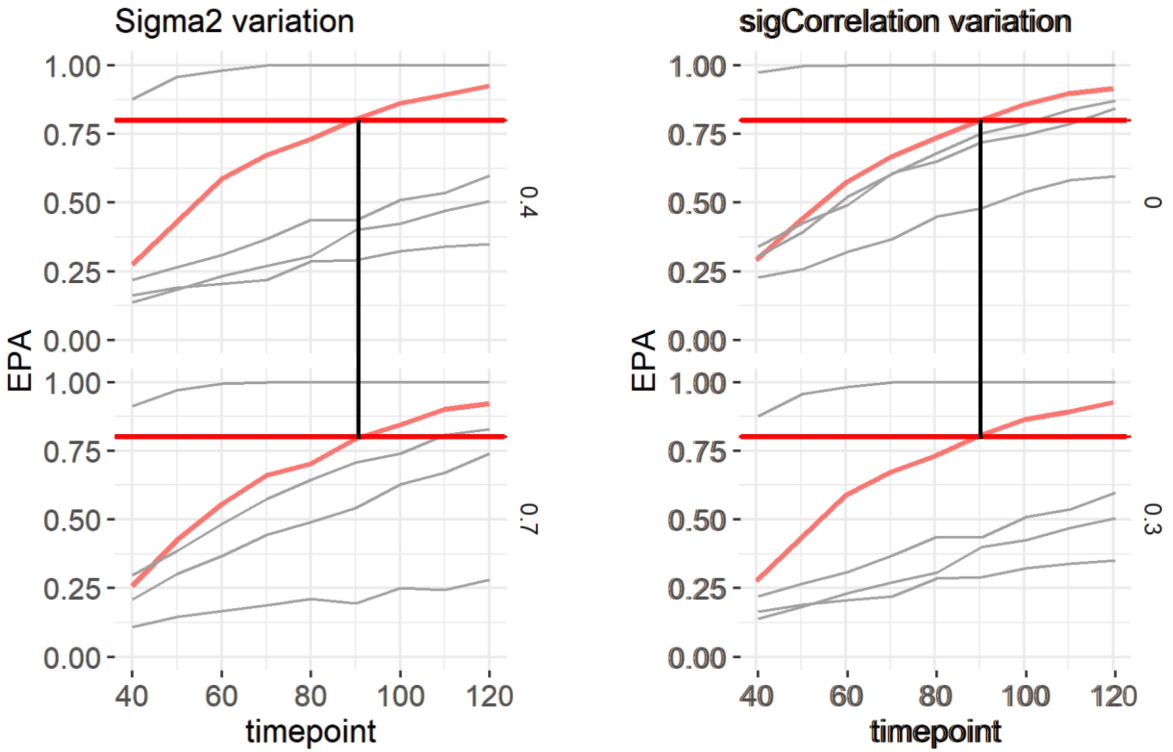


$$\Phi = \begin{pmatrix} .7 & .2 \\ .2 & .3 \end{pmatrix}$$
$$\delta = \begin{pmatrix} 0 & 0 \end{pmatrix}$$
$$\sigma = \begin{pmatrix} .4 & .3 \\ .2 & .4 \end{pmatrix}$$

Results: Parameters' influence

Expected Predictive Accuracy
Parameters' power





$$\Phi = \begin{pmatrix} .7 & .2 \\ .2 & .3 \end{pmatrix}$$
$$\delta = \begin{pmatrix} 0 & 0 \end{pmatrix}$$
$$\sigma = \begin{pmatrix} .4 & .3 \\ .2 & .4 \end{pmatrix}$$