Project: Causal Discovery between Manufacturer-Retailer Price Channels

Project Goal

- Identify **causal relationships** between three variables: Manufacturer's Price (MPRICE), Retailer's Price (RPRICE) and Quantity of soft drinks sold (QTY).

Dataset

- A time series data set containing three variables namely QTY, RPRICE, and MPRICE. The data was collected on a weekly basis at the Dominick's Finer Foods supermarket in the Chicago area and made available by the University of Chicago's Kilts Center.

Variable name	Description
QTY	Number of packs sold.
RPRICE	Retailer's selling price of a product.
MPRICE	Selling price of a product decided by the manufacturer.

Procedure: Build a Vector Autoregressive Model

- In order to **remove the interdependencies** among the variables in the time series data set, you will build a **vector autoregressive model (VAR model)**. The **Schwartz Information Criterion** will be used to choose the optimal time lag in the VAR model. While building a VAR model, you can assume absence of trend in the model.

```
# Build a VAR model
# Select the lag order using the Schwarz Information Criterion with a maximum lag of 10
VARselect(input_data, lag.max=10)
var_obj <- VAR(input_data, p=1)
```

Procedure: Test the residuals

- The residuals obtained from the VAR model for each variable will be tested using the Augmented Dickey-Fuller test at a significance level of 0.05 to detect if the time series is stationary or not.

```
# Extract the residuals from the VAR model
residual_data <- residuals(var_obj)

# Check for stationarity using the Augmented Dickey-Fuller test
apply(residual_data, 2, function(x) summary(ur.df(x, type="none")))
# The p-value for all the residuals of three variables indicates that there is evidence
# to reject the null hypothesis. Hence, the residuals follow a stationary pattern.
```

- Test if the residuals follow a normal distribution using the Shapiro Wilk test.

```
# Check whether the variables follow a Gaussian distribution
apply(residual_data, 2, function(x) shapiro.test(x))
# Based on the KS test, we have evidence to reject the null hypothesis i.e., the
# residuals follow a normal (Gaussian) distribution.
```

Procedure: Build causal graphs

- Build two causal graphs using the PC algorithm and LiNGAM algorithm. For the PC algorithm set the significance level to 0.1 and for the LiNGAM algorithm set the prune factor to 1.

```
# PC Algorithm
suffStat<-list(C=cor(residual_data), n=nrow(residual_data))
pc_fit <- pc(suffStat, indepTest=gaussCItest, alpha=0.1, labels=colnames(residual_data), skel.method="original", verbose=TRUE)
plot(pc_fit, main="PC Output")

# LiNGAM Algorithm
lingam_fit <- LINGAM(residual_data, verbose=TRUE)
show(lingam_fit)
```

Output: PC (left) and LiNGAM (right)