# Nonparametric Project of Agricultural Productivity in the U.S.

Permutational Tests

Sofia Moroni*

2023-06-27

## Contents



## 1 Load libraries and data

```
library(pbapply)
```

```
## Warning: il pacchetto 'pbapply' è stato creato con R versione 4.1.3
```

```
library(mgcv)
```

```
data_path = file.path('data')
output_path = file.path('results')
data =
    read.table(
        file.path(data_path, 'agricultural_indices.csv'),
        header = T,
        sep = ';'
    )

# Sostituzione delle virgole con punti
data<- data.frame(lapply(data, function(x) gsub(",", ".", x)))
```

*sofia.moroni@mail.polimi.it

```
data <- as.data.frame(lapply(data, as.numeric))
data<- data[1:68,]

set.seed(100)
B = 1000
n = nrow(data)
```

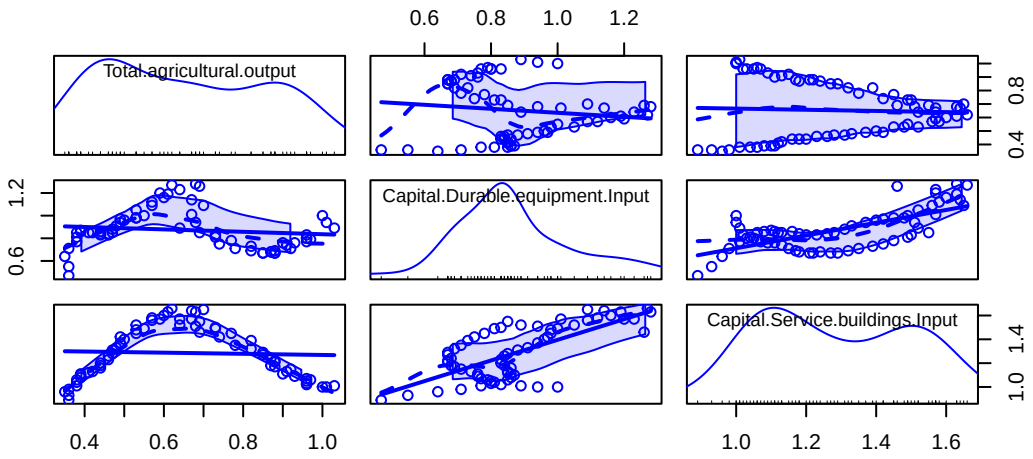
Studying the relationship between input variables and output variable

```
library(car)
```

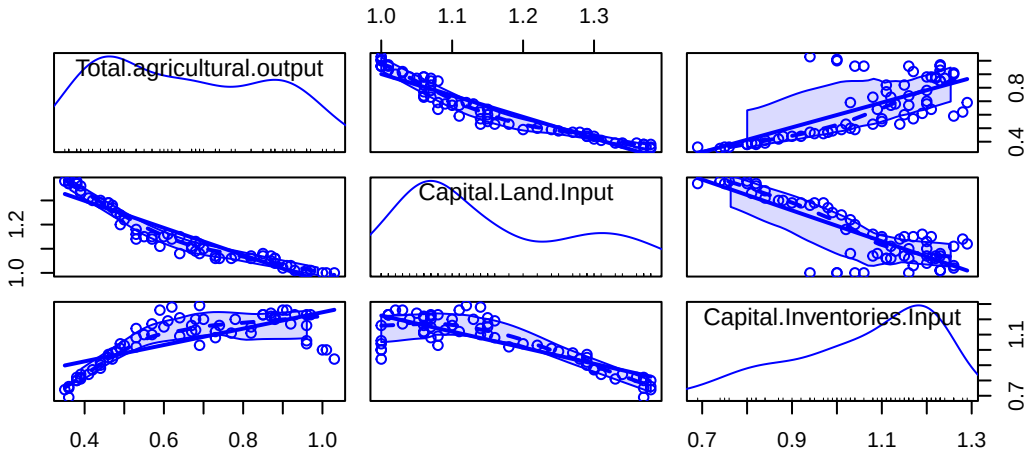
```
## Caricamento del pacchetto richiesto: carData
```

```
## Warning: il pacchetto 'carData' è stato creato con R versione 4.1.3
```

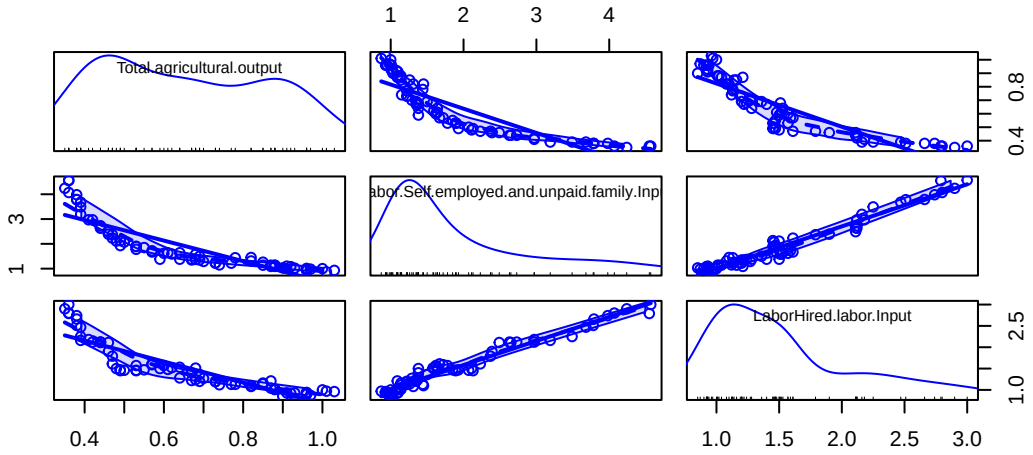
```
with(data  , scatterplotMatrix(data.frame(Total.agricultural.output, Capital.Durable.equipment.Input, Ca
```



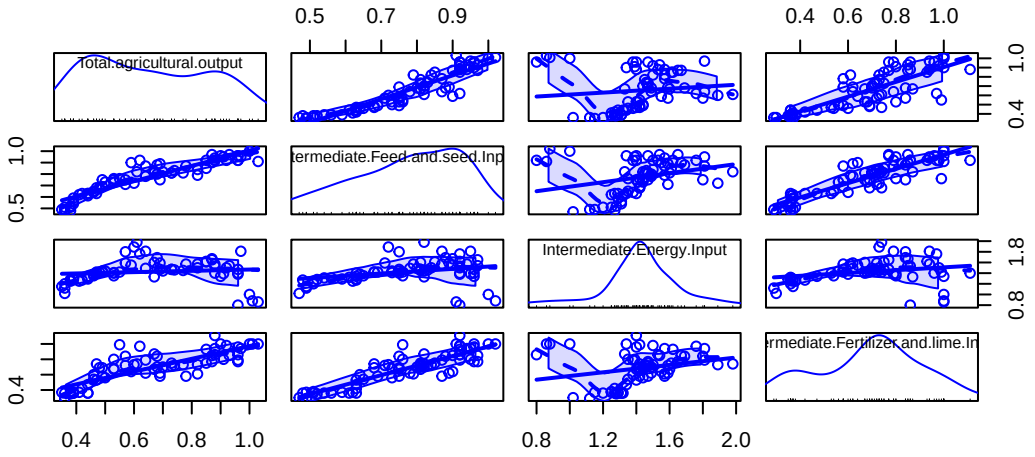

```
with(data ,scatterplotMatrix(data.frame(Total.agricultural.output,Capital.Land.Input,Capital.Inventories
```



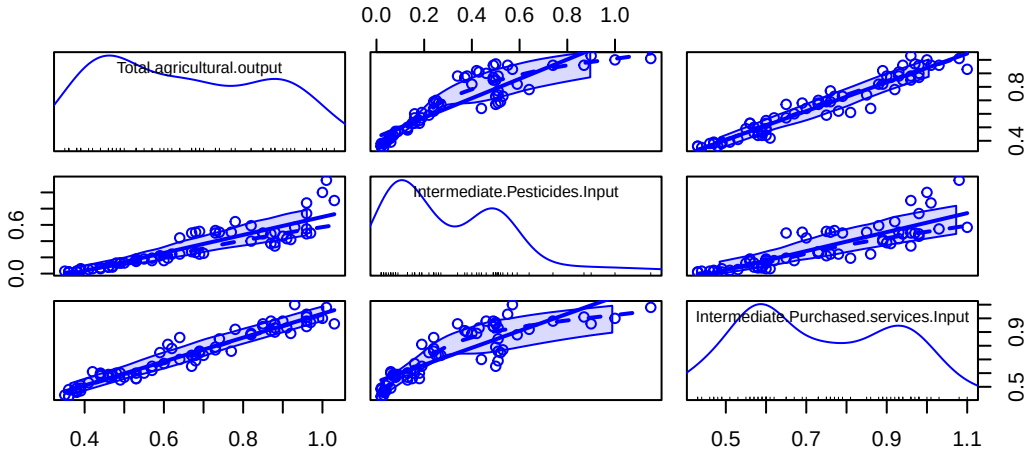

```
with(data , scatterplotMatrix(data.frame(Total.agricultural.output,Labor.Self.employed.and.unpaid.family
```




```
with(data, scatterplotMatrix(data.frame(Total.agricultural.output,Intermediate.Feed.and.seed.Input,Inte
```




```
with(data, scatterplotMatrix(data.frame(Total.agricultural.output,Intermediate.Pesticides.Input,Interme
```



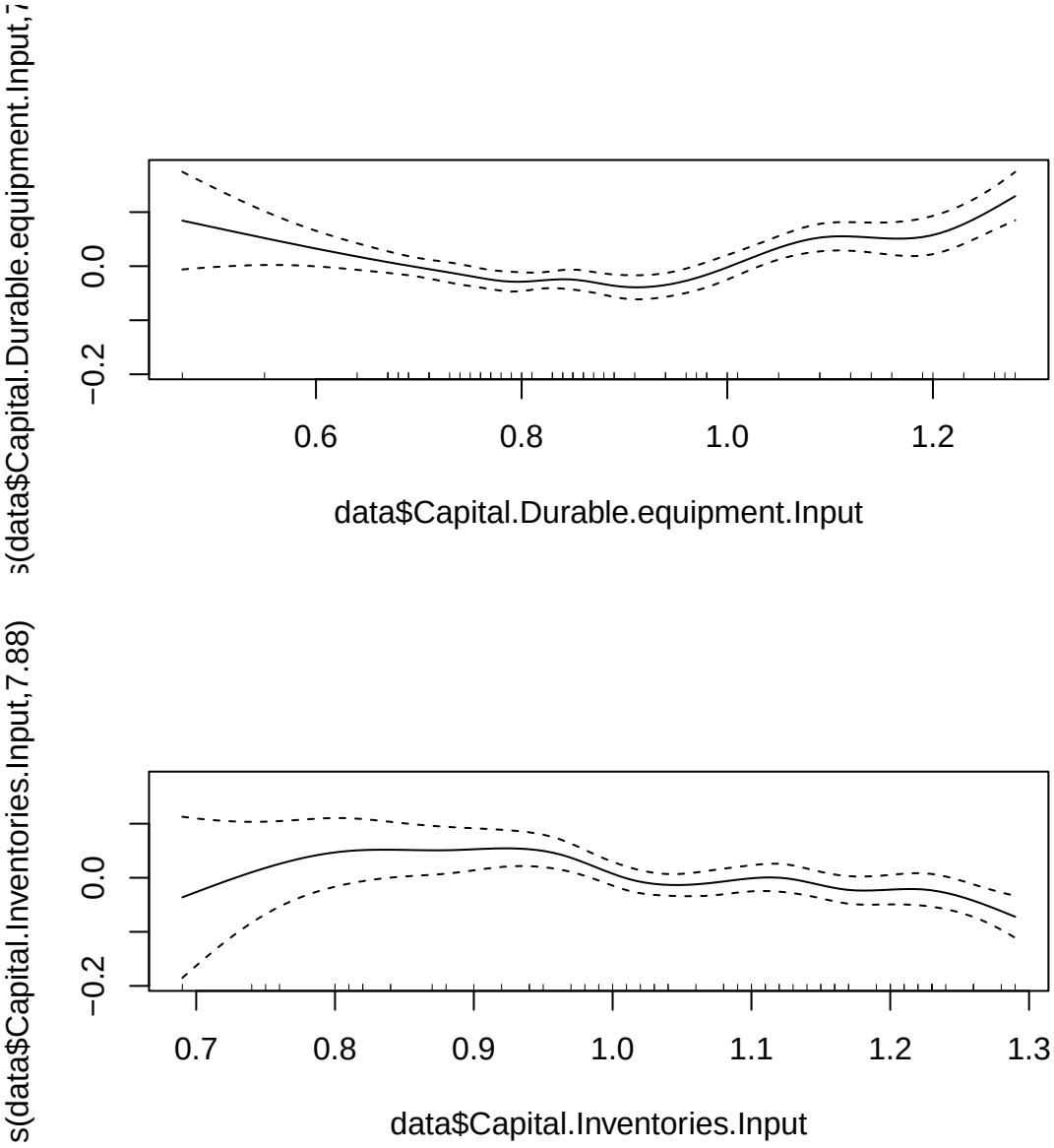

Starting with all the covariates

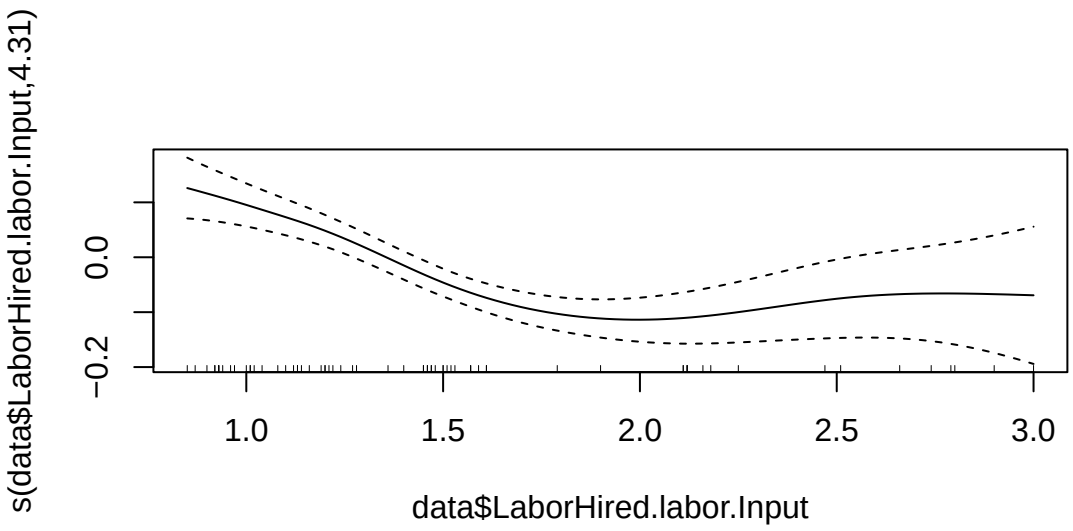
```
model_gam = gam(
    data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Feed.and.seed.Input
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Fertilizer.and.lime.Input
```

```
    + data$Intermediate.Purchased.services.Input
   )

summary(model_gam)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input,
##     bs = "cr") + data$Capital.Service.buildings.Input + data$Capital.Land.Input +
##     s(data$Capital.Inventories.Input, bs = "cr") + data$Labor.Self.employed.and.unpaid.family.Input +
##     s(data$LaborHired.labor.Input, bs = "cr") + data$Intermediate.Feed.and.seed.Input +
##     data$Intermediate.Energy.Input + data$Intermediate.Pesticides.Input +
##     data$Intermediate.Fertilizer.and.lime.Input + data$Intermediate.Purchased.services.Input
##
## Parametric coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                     1.28965    0.27975   4.610
## data$Capital.Service.buildings.Input           -0.26196    0.05012  -5.227
## data$Capital.Land.Input                        -0.27853    0.22092  -1.261
## data$Labor.Self.employed.and.unpaid.family.Input -0.09032    0.03164  -2.855
## data$Intermediate.Feed.and.seed.Input          -0.01930    0.11474  -0.168
## data$Intermediate.Energy.Input                  0.06988    0.02280   3.065
## data$Intermediate.Pesticides.Input              0.08415    0.04180   2.013
## data$Intermediate.Fertilizer.and.lime.Input     0.02022    0.02811   0.719
## data$Intermediate.Purchased.services.Input      0.09584    0.08097   1.184
##                                                Pr(>|t|)
## (Intercept)                                    4.27e-05 ***
## data$Capital.Service.buildings.Input           6.15e-06 ***
## data$Capital.Land.Input                         0.21490
## data$Labor.Self.employed.and.unpaid.family.Input  0.00687 **
## data$Intermediate.Feed.and.seed.Input           0.86727
## data$Intermediate.Energy.Input                  0.00394 **
## data$Intermediate.Pesticides.Input              0.05106 .
## data$Intermediate.Fertilizer.and.lime.Input     0.47623
## data$Intermediate.Purchased.services.Input      0.24376
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                        edf Ref.df     F  p-value
## s(data$Capital.Durable.equipment.Input) 7.904  8.568 6.200 2.47e-05 ***
## s(data$Capital.Inventories.Input)       7.877  8.629 2.957  0.00888 **
## s(data$LaborHired.labor.Input)          4.308  5.256 8.573 8.15e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.994   Deviance explained = 99.6%
## GCV = 0.00047364  Scale est. = 0.00027103  n = 68
```

```r
plot(model_gam)
```



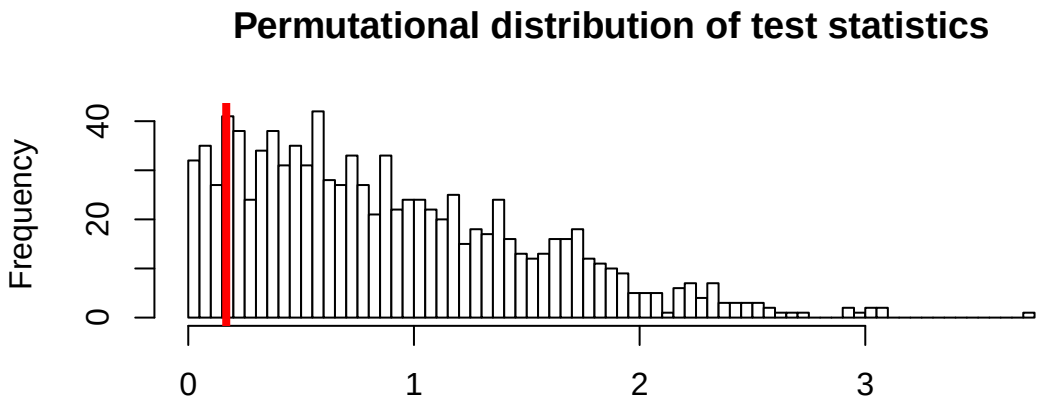



## 2 $H_0$: Feed and seed $= 0$ VS $H_1$: Feed and seed $\neq 0$

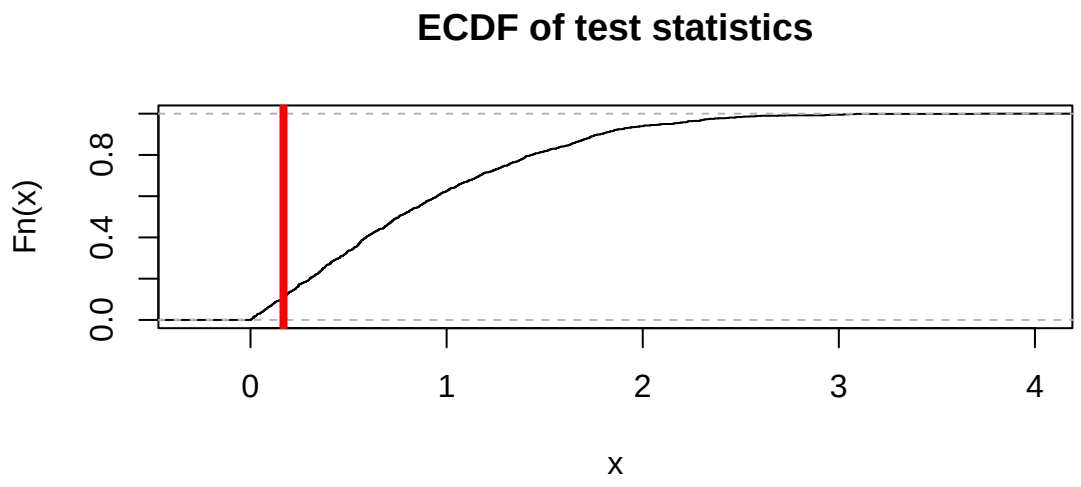
```
T0 = abs(summary(model_gam)$p.table[5, 3])
gam.H0 = gam(
    data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Fertilizer.and.lime.Input
    + data$Intermediate.Purchased.services.Input
   )


res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Feed.and.seed.Input
    + data$Intermediate.Energy.Input
```

```
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Fertilizer.and.lime.Input
    + data$Intermediate.Purchased.services.Input
   )

    return(abs(summary(gam.perm)$p.table[5, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(sort(T_H0)[-1000],
     breaks = 100,
     col = 'white',
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(sort(T_H0)[-1000]), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.892
```

Can accept H0, so I remove the variable

## 3 $H_0$: Fertilizer and lime $= 0$ VS $H_1$: Fertilizer and lime $\neq 0$

```
model_gam = gam(data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Fertilizer.and.lime.Input
    + data$Intermediate.Purchased.services.Input
   )

summary(model_gam)
```

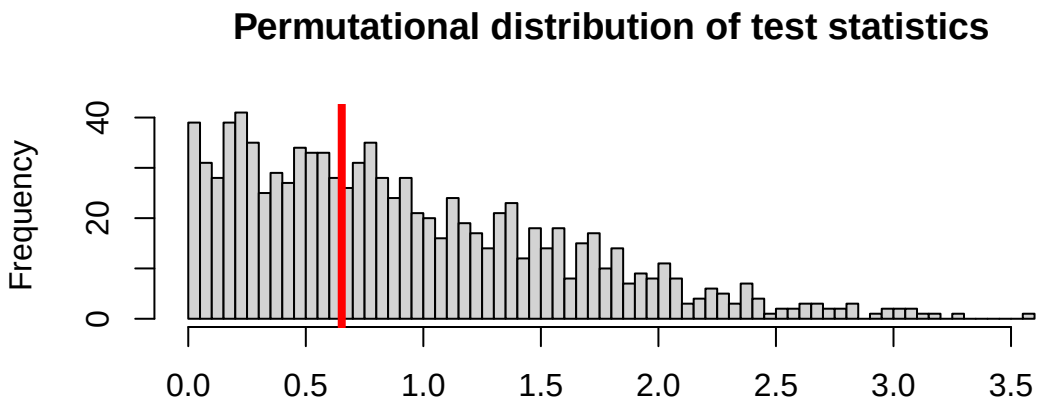
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input,
##     bs = "cr") + data$Capital.Service.buildings.Input + data$Capital.Land.Input +
##     s(data$Capital.Inventories.Input, bs = "cr") + data$Labor.Self.employed.and.unpaid.family.Input +
```

```
##     s(data$LaborHired.labor.Input, bs = "cr") + data$Intermediate.Energy.Input +
##     data$Intermediate.Pesticides.Input + data$Intermediate.Fertilizer.and.lime.Input +
##     data$Intermediate.Purchased.services.Input
##
## Parametric coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                     1.25795    0.23802   5.285
## data$Capital.Service.buildings.Input           -0.26208    0.04745  -5.523
## data$Capital.Land.Input                        -0.26263    0.21634  -1.214
## data$Labor.Self.employed.and.unpaid.family.Input -0.08990    0.03000  -2.997
## data$Intermediate.Energy.Input                  0.07175    0.02232   3.214
## data$Intermediate.Pesticides.Input              0.08500    0.04108   2.069
## data$Intermediate.Fertilizer.and.lime.Input     0.01759    0.02694   0.653
## data$Intermediate.Purchased.services.Input      0.09124    0.07768   1.175
##                                                Pr(>|t|)
## (Intercept)                                    4.87e-06 ***
## data$Capital.Service.buildings.Input           2.27e-06 ***
## data$Capital.Land.Input                         0.23196
## data$Labor.Self.employed.and.unpaid.family.Input  0.00469 **
## data$Intermediate.Energy.Input                  0.00260 **
## data$Intermediate.Pesticides.Input              0.04509 *
## data$Intermediate.Fertilizer.and.lime.Input     0.51766
## data$Intermediate.Purchased.services.Input      0.24717
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                          edf Ref.df     F  p-value
## s(data$Capital.Durable.equipment.Input) 8.122  8.685 6.680 1.05e-05 ***
## s(data$Capital.Inventories.Input)       7.973  8.678 3.555  0.00309 **
## s(data$LaborHired.labor.Input)          4.289  5.248 9.700 2.33e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.994   Deviance explained = 99.6%
## GCV = 0.00045095  Scale est. = 0.00026272  n = 68
```

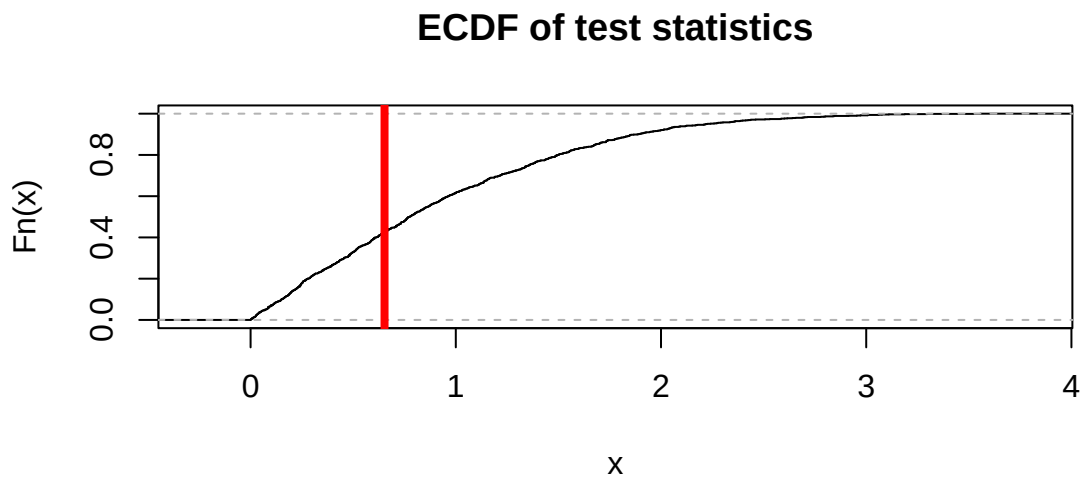
```
T0 = abs(summary(model_gam)$p.table[7, 3])
gam.H0 = gam(data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )

res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
```

```
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Fertilizer.and.lime.Input
    + data$Intermediate.Purchased.services.Input
    )
    return(abs(summary(gam.perm)$p.table[7, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(sort(T_H0)[-1000],
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(sort(T_H0)[-1000]), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.575
```

Can accept H0, so I remove the variable `Feed and Seed`

# 4 $H_0$: Land $= 0$ VS $H_1$: Land $\neq 0$

```
model_gam = gam(data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )

summary(model_gam)
```

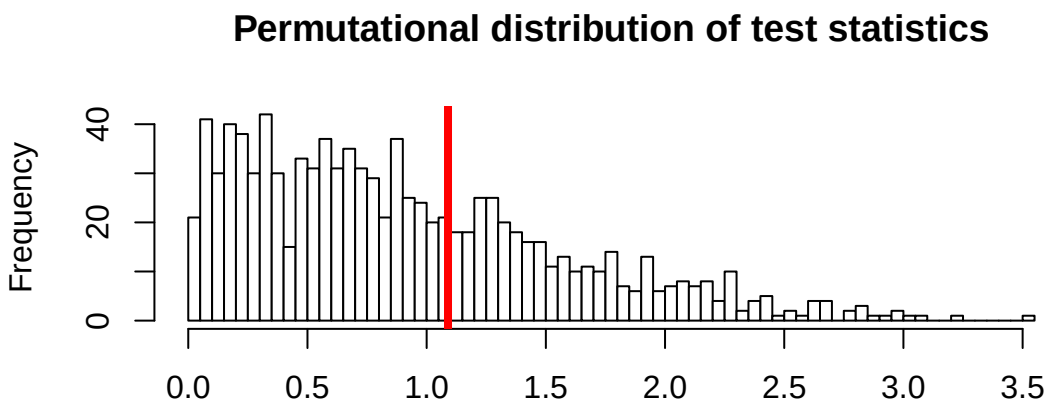
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input,
##     bs = "cr") + data$Capital.Service.buildings.Input + data$Capital.Land.Input +
##     s(data$Capital.Inventories.Input, bs = "cr") + data$Labor.Self.employed.and.unpaid.family.Input +
##     s(data$LaborHired.labor.Input, bs = "cr") + data$Intermediate.Energy.Input +
```

```
##     data$Intermediate.Pesticides.Input + data$Intermediate.Purchased.services.Input
##
## Parametric coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                     1.24819    0.23574   5.295
## data$Capital.Service.buildings.Input           -0.26498    0.04710  -5.626
## data$Capital.Land.Input                        -0.23273    0.21361  -1.090
## data$Labor.Self.employed.and.unpaid.family.Input -0.09624    0.02876  -3.347
## data$Intermediate.Energy.Input                  0.07374    0.02222   3.319
## data$Intermediate.Pesticides.Input              0.08898    0.04049   2.198
## data$Intermediate.Purchased.services.Input      0.08984    0.07752   1.159
##                                                Pr(>|t|)
## (Intercept)                                    4.55e-06 ***
## data$Capital.Service.buildings.Input           1.56e-06 ***
## data$Capital.Land.Input                         0.28240
## data$Labor.Self.employed.and.unpaid.family.Input  0.00178 **
## data$Intermediate.Energy.Input                  0.00193 **
## data$Intermediate.Pesticides.Input              0.03380 *
## data$Intermediate.Purchased.services.Input      0.25334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                          edf Ref.df      F  p-value
## s(data$Capital.Durable.equipment.Input) 8.443  8.845  7.675 2.18e-06 ***
## s(data$Capital.Inventories.Input)       8.175  8.776  4.013  0.00122 **
## s(data$LaborHired.labor.Input)          4.168  5.112 10.540 1.30e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.994   Deviance explained = 99.6%
## GCV = 0.00043442  Scale est. = 0.00025691  n = 68
```

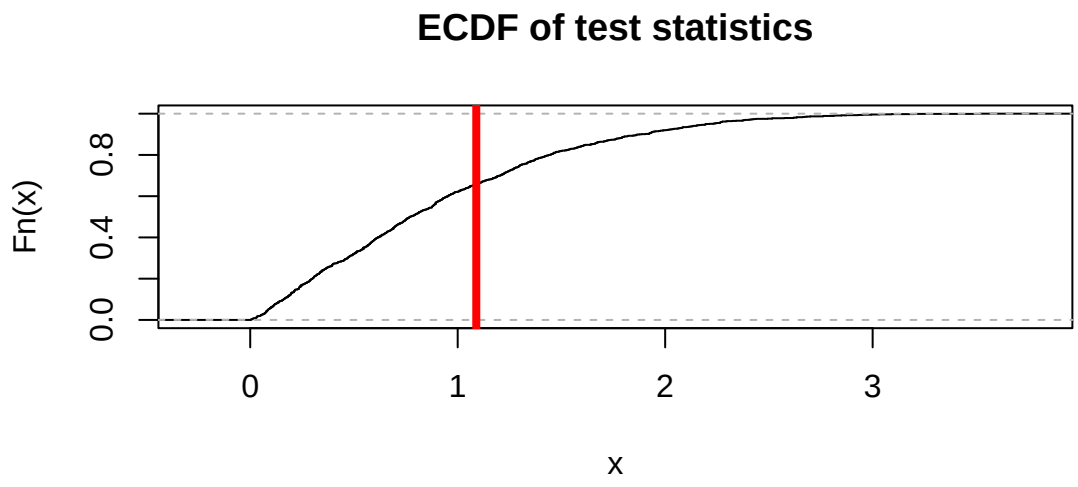
```r
T0 = abs(summary(model_gam)$p.table[3, 3])
gam.H0 = gam(data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )

res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + data$Capital.Land.Input
```

```
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )
    return(abs(summary(gam.perm)$p.table[3, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(sort(T_H0)[-1000],
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     col = 'white',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(sort(T_H0)[-1000]), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```




```
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.344
```

Can accept H0, so I remove the variable

## 5 $H_0$: Purchased Services $= 0$ VS $H_1$: Purchased Services $\neq 0$

```
model_gam = gam(data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )

summary(model_gam)
```

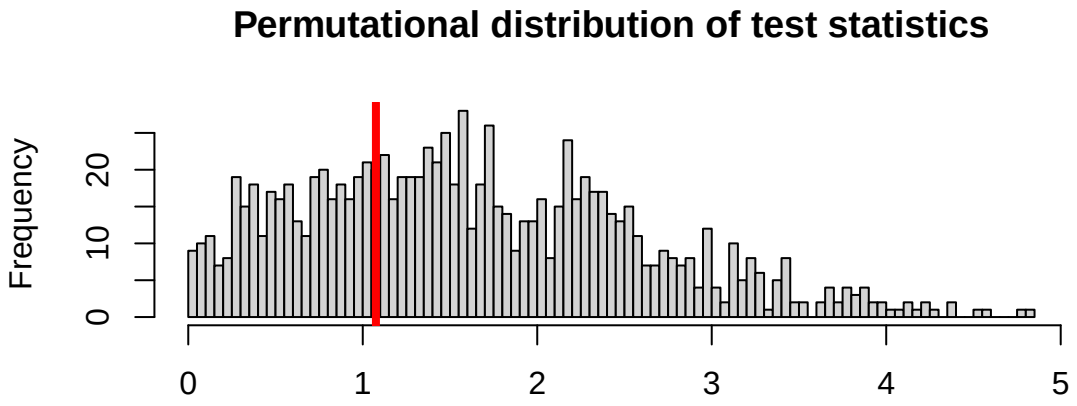
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input,
##     bs = "cr") + data$Capital.Service.buildings.Input + s(data$Capital.Inventories.Input,
##     bs = "cr") + data$Labor.Self.employed.and.unpaid.family.Input +
##     s(data$LaborHired.labor.Input, bs = "cr") + data$Intermediate.Energy.Input +
##     data$Intermediate.Pesticides.Input + data$Intermediate.Purchased.services.Input
```

```
##
## Parametric coefficients:
##                                                 Estimate Std. Error t value
## (Intercept)                                      1.02299    0.10810   9.463
## data$Capital.Service.buildings.Input            -0.27125    0.04676  -5.801
## data$Labor.Self.employed.and.unpaid.family.Input -0.11614    0.02221  -5.229
## data$Intermediate.Energy.Input                   0.07427    0.02214   3.355
## data$Intermediate.Pesticides.Input               0.10920    0.03608   3.027
## data$Intermediate.Purchased.services.Input       0.08310    0.07727   1.075
##                                                 Pr(>|t|)
## (Intercept)                                     7.19e-12 ***
## data$Capital.Service.buildings.Input            8.30e-07 ***
## data$Labor.Self.employed.and.unpaid.family.Input 5.35e-06 ***
## data$Intermediate.Energy.Input                   0.00172 **
## data$Intermediate.Pesticides.Input               0.00426 **
## data$Intermediate.Purchased.services.Input       0.28845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                           edf Ref.df      F  p-value
## s(data$Capital.Durable.equipment.Input) 8.601  8.908 10.523  < 2e-16 ***
## s(data$Capital.Inventories.Input)       8.292  8.821  4.209 0.000788 ***
## s(data$LaborHired.labor.Input)          4.092  5.028 17.341  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.994   Deviance explained = 99.6%
## GCV = 0.00042026  Scale est. = 0.00025349  n = 68
```

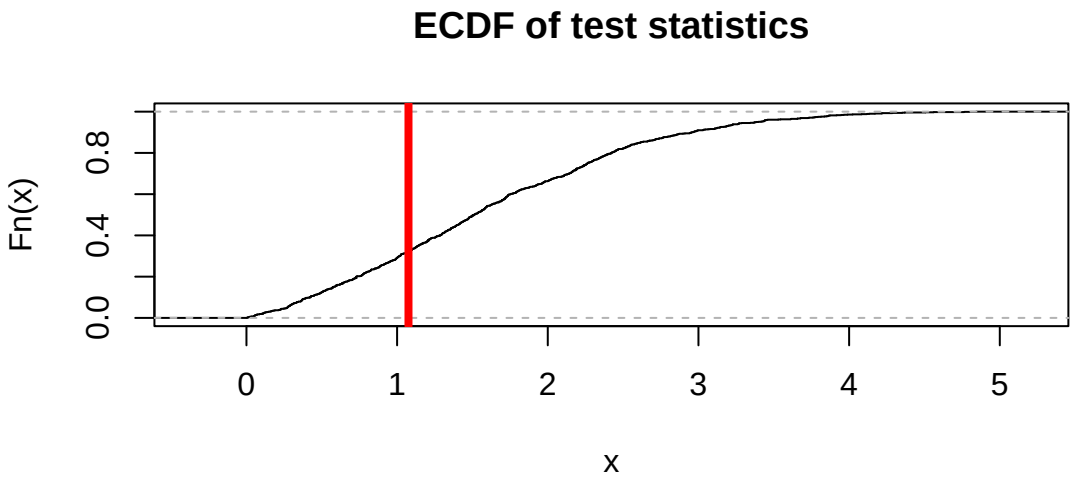
```
T0 = abs(summary(model_gam)$p.table[6, 3])
gam.H0 = gam(data$Total.agricultural.output ~ s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )

res.H0 = gam.H0$residuals

wrapper = function() {
    permutation = sample(n)
    res.H0.perm = res.H0[permutation]
    Y.perm.H0 = gam.H0$fitted + res.H0.perm
    gam.perm = gam(
        Y.perm.H0 ~  s(data$Capital.Durable.equipment.Input, bs = 'cr')
    + data$Capital.Service.buildings.Input
    + s(data$Capital.Inventories.Input,bs = 'cr')
    + data$Labor.Self.employed.and.unpaid.family.Input
    + s(data$LaborHired.labor.Input, bs = 'cr')
    + data$Intermediate.Energy.Input
```

```
    + data$Intermediate.Pesticides.Input
    + data$Intermediate.Purchased.services.Input
   )
    return(abs(summary(gam.perm)$p.table[6, 3]))
}
T_H0 = pbreplicate(B, wrapper(), simplify = 'vector')
```

```
hist(sort(T_H0)[-1000],
     breaks = 100,
     main = 'Permutational distribution of test statistics',
     xlab = '')
abline(v = T0, col = 'red', lwd = 4)
```




```
plot(ecdf(sort(T_H0)[-1000]), main = 'ECDF of test statistics')
abline(v = T0, col = 'red', lwd = 4)
```



```r
P = sum(T_H0 >= T0) / B
P
```

```
## [1] 0.68
```