

# Frisch-Waugh-Lovell theorem

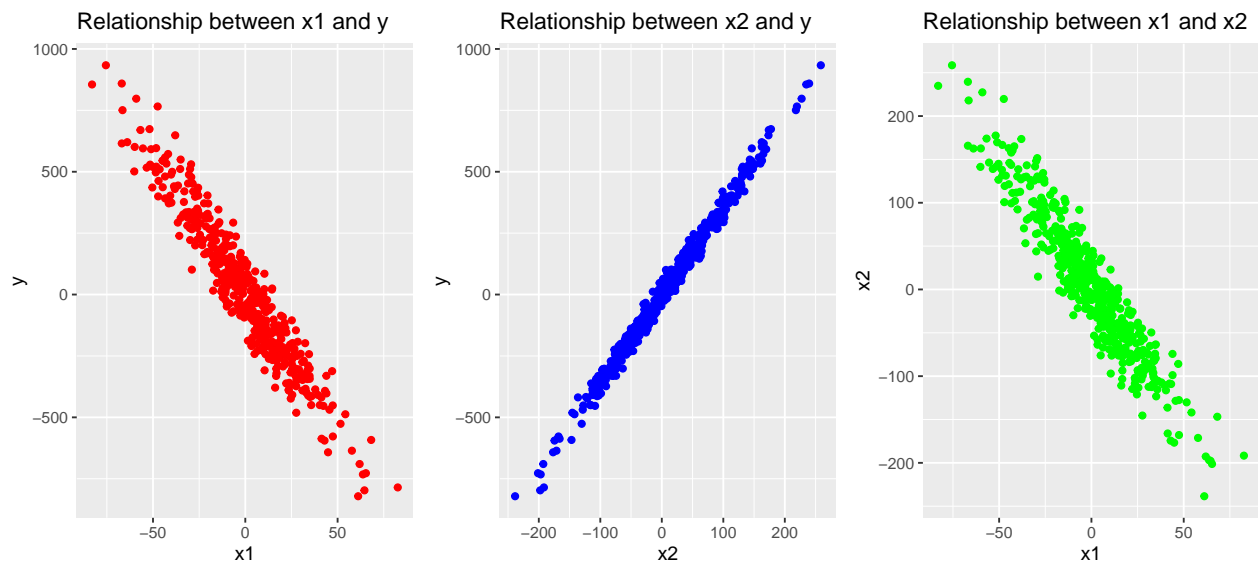
## Brief description

The Frisch–Waugh–Lovell (FWL) theorem allows us to reduce a multivariate regression analysis to an univariate one. The main idea behind is the fact that there are multiple ways to estimate a  $\beta_1$  coefficient in the following regression model:  $y_i = \beta_1 x_i^1 + \beta_2 x_i^2 + \varepsilon_i$ . The main idea behind this is to use the residuals of the regression of  $x^1$  on  $x^2$  as the regressor of  $y$ .

## Let's see it in practice

In order to give the intuition of what's going on, we will start with a concrete example:

```
#create the data
set.seed(100)
n = 500
x1 = rnorm(n=n, sd=25)
x2 = -3*x1 + rnorm(n=n, sd=25)
y = -2*x1 + 3*x2 + rnorm(n=n, sd=25)
data = data.frame(x1,x2,y)
```



We clearly see here that all of our variables are correlated. Now, let's say we want to make a linear regression with  $y$  as a predicted variable and check the estimated coefficients.

```
attach(data)
regression = lm(y ~ x1 + x2)
regression$coefficients
```

```
## (Intercept)          x1          x2
## -0.3094243  -1.9072401   3.0243063
```

For the moment, our estimation seems to go well since the estimated coefficients are near the true value. We said before that the FWL theorem allows us to go from a multivariate regression to an univariate one. What does it mean? You probably know that adding or removing variable to a regression has an impact on the other coefficients. If you're not fully clear with that idea, I encourage you to try to compare these models, especially the  $\hat{\beta}_1$  coefficient:  $\hat{y} = \hat{\beta}_1 x^1$  and  $\hat{y} = \hat{\beta}_1 x^1 + \hat{\beta}_2 x^2$ .

You might say that how is it possible to reduce multivariate regression to an univariate one since adding or removing variable has an impact on the estimation. The answer is the FWL theorem!

**The theorem says that all of these are equivalent for estimate  $\beta_1$ :**

- (1) Estimation by regressing  $y$  on  $x_1$  and  $x_2$
- (2) Estimation by regressing  $y$  on the residual from the regression of  $x_1$  on  $x_2$ , generally called *orthogonalization* or *residualization*
- (3) Estimation by regressing the residual from the regression of  $y$  on  $x_2$  on the residual from the regression of  $x_1$  on  $x_2$

Ok I know it's not fully clear, let's see an example.

**Let's compute for the (1) method:**

```
#compute beta from regressing y on x1 and x2
regression = lm(y ~ x1 + x2)
regression$coefficients
```

```
## (Intercept)      x1      x2
## -0.3094243 -1.9072401  3.0243063
```

Nothing new here. We have the same  $\hat{\beta}_1$  as before.

**Now let's compute with the (2) method (*orthogonalization*):**

```
#compute the residuals of regressing x1 on x2
first_reg = lm(x1 ~ x2)
residuals = first_reg$residuals

#compute beta from regressing y on the residuals from above
second_reg = lm(y ~ residuals)
second_reg$coefficients
```

```
## (Intercept)  residuals
##    15.39855    -1.90724
```

As you can see, we have the exact same value as before with  $\hat{\beta}_1$ ! Obviously, there is no chance here, but let's see the last method.

**And finally, the (3) method:**

```
#compute residuals of regressing y on x2
first_reg = lm(y ~ x2)
residuals1 = first_reg$residuals

#compute the residuals of regressing x1 on x2
second_reg = lm(x1 ~ x2)
```

```
residuals2 = second_reg$residuals

#compute beta from regressing the residuals_1 on residuals_2
last_reg = lm(residuals1 ~ residuals2)
last_reg$coefficients

##      (Intercept)      residuals2
## -5.537335e-15 -1.907240e+00
```

Wow there all the same! You can see that we have reduced a multivariate regression  $y_i = \beta_1 x_i^1 + \beta_2 x_i^2 + \varepsilon_i$  to an univariate one.

What happened is that we isolate the variation of  $x_1$  that is orthogonal from  $x_2$ . Put another way, when computing the residuals from regressing  $x_1$  on  $x_2$ , we **keep only the variation of  $x_1$  that is unexplained by  $x_2$** .

## Why do we care about this theorem?

First, I find it pretty fun to discover things like that, especially when the results are this precise. More seriously, this theorem, even if it seems very theoretical, has quite concrete applications: data visualization, computational speed and for specific inferences. As I am not particularly knowledgeable about this, I refer you to an article (check references at the end) which gives precise leads and sources.

## Proof of the theorem

If you're still here it means that you are ready to go through the demonstration, which is not very funny. It's far from being a very difficult demonstration but I personally had to do it several times before fully understanding what was happening.

### Before starting

Before going any further, you might want to check an excellent article called "*Why Linear Regression is a Projection*" (see references). This article will give you the taste of why we can talk about linear regression with projections, very important for the following demonstration.

### Starting point

We assume the following regression:

$$y = X\hat{\beta}_1 + Z\hat{\beta}_2 + r$$

Where:

- $y$  is the continuous variable we want to predict
- $X$  and  $Z$  are our predictors (vector or matrix, no differences)
- $\hat{\beta}_1$  and  $\hat{\beta}_2$  our estimation using least squared ordinary method
- $r$  the residuals of the estimation

### Useful properties

#### *Propertie 1:*

If there is no collinearity between our explanatory variables, the best fit to the least squares problem is unique.

**Propertie 2:**

Any matrix of variables can be split in its projection. For our regression above, it means that we can re-write  $Z = P_X Z + M_X Z$ , where  $P_X$  is the projection to  $X$  and  $M_X$  the *residual makers* for  $X$ . Also, we have:  $M_X = I - P_X$ .

**Propertie 3:**

A regression on orthogonal sets of regressors can be done on each set at a time while still getting the coefficients from the joint regression.

**A bit more about projection**

In order to go from  $\hat{\beta} = (X^T X)^{-1} X^T y$  into the projection  $\hat{y}$ , we only have to multiple it by  $X$ , which gives us:

$$\begin{aligned} X\hat{\beta} &= X(X^T X)^{-1} X^T y = P_x y = \hat{y} \\ P_x &= X(X^T X)^{-1} X^T \end{aligned}$$

Now, let's say we have the two following estimations:

$$\begin{aligned} y &= X\hat{\beta}_1 + Z\hat{\beta}_2 + r_1 \\ M_Z y &= M_Z X\hat{\beta}_1 + M_Z Z\hat{\beta}_2 + M_Z r_1 \end{aligned}$$

where:

$$M_Z = I - P_Z = I - Z(Z^T Z)^{-1} Z^T = \text{residual makers}$$

What we want to prove with this theorem is that  $\hat{\beta}_1 = \hat{\beta}_3$ .

**We re-write our first estimation by multiplying it by  $M_Z$** 

$$M_Z y = M_Z X\hat{\beta}_1 + M_Z Z\hat{\beta}_2 + M_Z r_1$$

With:

$$\begin{aligned} - M_Z Z\hat{\beta}_2 &= (I - Z(Z^T Z)^{-1} Z^T) Z\hat{\beta}_2 = Z\hat{\beta}_2 - Z(Z^T Z)^{-1} Z^T Z\hat{\beta}_2 = Z\hat{\beta}_2 - Z\hat{\beta}_2 = 0 \\ - M_Z r_1 &= (I - Z(Z^T Z)^{-1} Z^T) r_1 = r_1 - Z(Z^T Z)^{-1} Z^T r_1 = r_1 - Z Z^{-1} (Z^T)^{-1} Z^T r_1 = r_1 - 0 = r_1 \end{aligned}$$

So we now have:

$$M_Z y = M_Z X\hat{\beta}_1 + r_1$$

**We now multiply it by  $X^T$** 

$$X^T M_Z y = X^T M_Z X\hat{\beta}_1 + X^T r_1$$

With:

$$X^T r_1 = X^T (y - X\hat{\beta}_1 - Z\hat{\beta}_2) = X^T y - X^T (X\hat{\beta}_1 + Z\hat{\beta}_2) = X^T y - X^T y = 0$$

So we now have:

$$X^T M_Z y = X^T M_Z X\hat{\beta}_1$$

We multiple the second estimation by  $(M_Z X)^T$

$$\begin{aligned}(M_Z X)^T M_Z y &= X^T M_Z^T M_Z y = X^T M_Z^T (M_Z X \hat{\beta}_3 + r_2) \\ &= X^T M_Z^T M_Z X \hat{\beta}_3 + X^T M_Z^T r_2\end{aligned}$$

With:

$$X^T M_Z^T r_2 = X^T M_Z^T (M_Z y - M_Z X \hat{\beta}_3) = X^T M_Z^T M_Z (y - X \hat{\beta}_3) = 0$$

So we have:

$$X^T M_Z y = X^T M_Z X \hat{\beta}_3$$

### Finishing the proof

We proved that  $X^T M_Z y = X^T M_Z X \hat{\beta}_1$ .

We also proved that  $X^T M_Z y = X^T M_Z X \hat{\beta}_3$ .

So we can conclude that  $y = X_1 \hat{\beta}_1 = X_1 \hat{\beta}_3$  and  $\hat{\beta}_1 = \hat{\beta}_3$ .

As you have seen, the demonstration goes in all directions and is not easy to follow if you don't do the calculations yourself. Although I don't think it's useful to master this demonstration (except in very special cases), I think it's good to manipulate the equations a little bit to improve your intuition.

I hope this has helped you if you are trying to understand this theorem better.

## References

*Why Linear Regression is a Projection* (2017), Vladimir Mikulik

*Understanding the Frisch-Waugh-Lovell Theorem* (2022), Matteo Courthoud

*Partial Time Regressions as Compared with Individual Trends* (1933), Ragnar Frisch and Frederick V. Waugh

*Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis* (1963), Michael Lovell