# **Methodology and model**

In this section, we discuss about estimation methodology of our model. Then, we explain our different models that give an answer at our research question. And we analyze potential problems of estimation that we have to solve in order to estimate a model without bias.

## **Method Choice**

As mentioned in the previous section, we use daily data for 226 companies between the beginning of 2006 and at the end of 2019.

Our study is about an analysis of a financial data and companies over time period. Then, two possibilities could be possible: Panel Data & Time series.

Panel data seems us to be the better solution because we analyze **i companies** over **t period**.

The use of panel data has many advantages, in terms of the flexibility of econometric modeling and the ability to control for unobserved heterogeneity. It also involves a number of econometric issues that require specific attention.

Moreover, our research tends to respond a dynamic question. By consequent, we need a dynamic model over period of time and company

## **Mode**l

First, we estimate a model without an Instrumental Variable

**Panel Model (fixed effect)**

However, this model suffers of endogeneity bias, more precisely, it suffers of **Reverse causality bias.**

That means :

BEN-David [[1]](#footnote-1)use as Instrumental variable, either the jump between Russel 2000 to Russel 100 or the enter of S&P 500 to illustrate a increase of volatility due to holding of ETF. We choose to use his method to manage our problem of endogeneity.

So, we use 2SLS panel regression to estimate effect of Total ETF growth rate on the growth rate of Companies.

Then, the first stage of regression is written:

Thus, this 2SLS panel regression allows us to isolate the impact of Total NAV ETF growth on growth of PE thanks to our “Dummy” instrument variable: **in S&P 500**.

## **Potential error of estimation**

### Omitted variables bias

Omitted variable bias is very frequent in financial study.

According to this equation, we can see that if we do not care about potential omitted variables, our estimators will be biased.

To solve this problem, we incorporate several control variables in our model. Moreover, we also include fixed effect at our panel regression. These fixed effects remove time-invariant unobserved heterogeneity and thus, our estimators could be strict exogenous of our explanatory variable. That allows to estimate correctly effects of independent variables on dependent variable.

### Multicollinearity

Moreover, we test the potential multicollinearity thanks to VIF Stata command. With this command, further investigation could be necessary if one variable have a VIF larger than 10.

It is not the case and so, our model does not suffer to multicollinearity between our chosen variable.

Thanks to VIF command and economic interpretation, it seems that our model manage potential problem of multicollinearity.

(sortir les graphes)

### Measurement error

In our study, we use daily data. However, daily data are more sensible to extreme value because of different reasons.

Thus, we choose to **winsor** PE and ROE of company. It is two variables that are the most sensible of big variation.

As mentioned in data analysis sections, the objective of this method is to manage the effect of outlier data. We know that OLS regression are very sensible to outliers.

It could be possible to truncate this data. However, truncating data could delete other useful information for our model.

Thus, for us, the better way to manage our measurement error is to do a winsorization or PE and ROE of company.

### Reverse causality

In our study, the major problem of endogeneity is **reverse causality.** That means that the growth of total ETF NAV influence PE of company, but the reverse is also true. Then, it is not possible to estimate correctly the impact of Total ETF NAV growth on company PE.

However, it has to analyze if our instrument is a good instrument for our study. So, we have to see if our instrument is “**relevant**” and “**exogenous**”.

We can consider “in S&P 500” relevant. When a company begin to be part of S&P 500, ETF with physical replication is going to buy this company. Then, the enter in S&P 500 of company i is going to influence NAV of ETF and also NAV ETF growth rate.

Moreover, our instrument could be considered as exogeneous of our dependent variable.

PE of company i do not influence the enter of S&P 500 but at the contrary the inverse is true.

When a company enters in S&P 500, more investors see this company and so the popularity of this stock increase. This popularity could be impacting the price of this stock.

By consequent, we can consider “in S&P 500” exogeneous of PE of company.

1. <https://www.sec.gov/comments/s7-11-15/s71115-1.pdf> [↑](#footnote-ref-1)