Homework: SuperBowl Econometric Analysis

John SIBONY

November 21, 2018

Introduction

Le SuperBowl est l'événement sportif annuel aux Etats-Unis, qui oppose en finale les premiers de chacune des deux ligues de football américain (la National Football Conference NFC et l'American Foot-ball Conference AFC). Krueger et Kennedy (1990) ont remarqué que la victoire d'une équipe de la NFC (resp. AFC) était fortement corrélée avec une hausse (resp. une baisse) du marché américain. Sur la période 1967-1988, lorsqu'une équipe de la NFC (resp. AFC) a remporté le SuperBowl, l'indice SP500 a augmenté (resp. baissé) en moyenne de 15,24% (resp. 10,93%) en moyenne. Or, ces moyennes sont apparue statistiquement significatives. Ils ont d'autre part testé la capacité prédictive du facteur SuperBowl sur l'indice de marché selon le test de Henriksson et Merton (1981) et le test de Reiganum (1981), démontrant la persistance des résultats sur toute la période d'observation.

Ces analyses théoriques ont été backtestés par une stratégie d'investissement active ayant surperformé l'indice SP500 sous hypothèses de forte taxation. Nous allons nuancer ces résultats en s'appuyant sur des outils économétriques, à l'aide des nouvelles données disponibles depuis l'étude menée.

Modèle

Notons Y la variable des rentabilités annuelle de l'indice SP500, X la variable binaire valant 0 si la NFC a perdu le SuperBowl, 1 sinon. Partons de l'hypothèse d'un modèle linaire simple : $Y_t = \beta_0 + \beta_1 X_t + \epsilon_t = \begin{cases} \beta_0 + \epsilon_t & \text{if } X_t{=}0 \\ \beta_0 + \beta_1 + \epsilon_t & \text{if } X_t{=}1 \end{cases}$

avec $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)\ \forall t \in [\![1, T]\!]$. De manière intuitive, la meilleure prédiction $\hat{\beta}_0$ de β_0 est la moyenne des Y lorsque X=0. Il en va de même pour $\hat{\beta}_1$ et on peut démontrer facilement ces résultats :

$$\hat{\beta}_0 = \frac{1}{N} \sum_{t, X_t = 0}^{T} Y_t \qquad \hat{\beta}_0 + \hat{\beta}_1 = \frac{1}{T - N} \sum_{t, X_t = 1}^{T} Y_t \qquad \text{with N=card}(t, X_t = 0) \tag{1}$$

Analyse des Résultats

Notre dataset contient les résultats du Superbowl (X) associés au rendements annuel du SP500 (Y), entre 1967 et 2013. Notons que l'année 1995 est manquante ce qui n'est sans importance puisqu'il n'y a pas de dépendance temporelle entre les variables. Nous avons considérés 3 identiques modèles pour des périodes différentes : 1967-1988 (dataset du papier), 1989-2013 (dataset restant), 1967-2013 (tout le dataset).

On remarque tout d'abord la concordance de nos résultats (voir section "R output") avec ceux du papier sur la période 1967-1988. En effet, toutes les hypothèses de la régression linéaire sont vérifiées à l'exception de la normalité des résidus (inspection graphique à préférer par rapport au test de Shapiro); homoskedasticity selon le test Breusch-Pagan robust (moins sensible à la normalité des résidus) et l'inspection graphique; non autocorrelation selon les test Breusch-Godfrey et Durbin Watson et l'inspection graphique; bonne spécification linéaire selon le test RESET. On peut ainsi valider la significativité de X (pvalue>> $\alpha = 0.01$ fixé) et le coefficient R^2 ajusté indiquant un bon pouvoir explicatif (67% de la variance expliqué par le modèle).

De même pour les 2 autres périodes, une analyse des test montre que l'ensemble des hypothèses sont vérifiées (à l'exception de la normalité qui ne biaise par nos estimateurs !). Le R^2 ajusté de la période 1989-2013 illustre une nette dégradation de la performance de X sur Y allant jusqu'à rejeter sa significativité au taux $\alpha = 0.01$. Bien que sur la totalité de la période $\hat{\beta}_1$ apparaît statistiquement significatif, le R^2 ajusté reste faible (<20%).

Raisonnons par l'absurde et montrons que X est indépendant de Y. Supposons que X est liée linéairement à Y. Les modèles de 1967-1988 et 1989-2013 sont en accord avec notre hypothèse. Cependant, de 1989 à 2013 la baisse du R^2 ajusté ne peut être dû qu'à l'augmentation de la variance empirique de $\hat{\epsilon}$. En effet,

$$R^2 = 1 - \frac{\mathrm{Var}_{\mathrm{emp}}(\hat{\epsilon})}{\mathrm{Var}_{\mathrm{emp}}(Y)} \underset{\text{by (1)}}{=} 1 - \frac{\mathrm{Var}_{\mathrm{emp}}(Y_{X=0}) + \mathrm{Var}_{\mathrm{emp}}(Y_{X=1})}{\mathrm{Var}_{\mathrm{emp}}(Y)}$$

et d'après la première figure 4 p.13, la variance totale des points verts (1967-1988) et rouges (1989-2013) sont identiques. Cependant, la variance des Y sur les sous groupe X=0 et X=1 augmente entre les deux périodes (nette augmentation visuelle de la variance des points rouges comparé aux points verts quand X=1). Le tableau ci dessous résume cette analyse :

1967-1988	1989-2013
$\text{Var}_{\text{emp}}(Y_{X=0}) = 0.015$	$\text{Var}_{\text{emp}}(Y_{X=0}) = 0.037$
$\text{Var}_{\text{emp}}(Y_{X=1}) = 0.005$	$\text{Var}_{\text{emp}}(Y_{X=1}) = 0.034$
$\text{Var}_{\text{emp}}(Y) = 0.24$	$\text{Var}_{\text{emp}}(Y) = 0.033$

On comprend ainsi que de 1967 à 1988 il existait un lien fort entre X et Y : quand X valait 1, la variance de Y était 3 fois moins élevé que lorsque X valait 0. Cependant, ce lien est rompu de 1989 à 2013 ($\text{Var}_{\text{emp}}(Y_{X=0}) \simeq \text{Var}_{\text{emp}}(Y_{X=1})$). Deux cas se présentent alors à nous : soit le modèle de 1967 à 1988 est vraie (X prédit Y), soit celui de 1989 à 2013 (X indépendant de Y). Une simple comparaison des R^2 ajustés sur la totalité du dataset et sur la période 1967 à 1988 montre un forte dégradation de performance (0.2 contre 0.65) en adéquation avec le second cas. On en arrive ainsi à conclure par l'absurde que X ne prédit pas Y.

Quelques remarques

- On a montré que X est indépendant de Y. Cependant, on peut se demander pourquoi de 1967 à 1988 notre modèle arrive à expliquer 67% de la variance de Y. Cela peut s'expliquer par le fruit du hasard. En effet, en se restreignant sur la prédiction de la hausse ou baisse du SP500 seulement, 20/22 des observations sont prédites par X. Si X est indépendant de Y, la probabilité que 20/22 des observations soient correctement prédites est la probabilité qu'une loi Binomiale de paramètre (n, θ) valent 20, avec n=22 et $\theta = P(X = Y) = P(X = 0)P(sgn(Y) = 0) + P(X = 1)P(sgn(Y) = 1)$. En prenant les probabilités à priori selon notre dataset on a : $\theta = 0.76 * 0.74 + 0.24 * 0.26 = 0.6248$. Et alors on obtient 0.27% de chance (ce qui est raisonnablement probable) que 20/22 des observations soient prédite si X est indépendant de Y.

- Bien que le dernier modèle présente un faible R^2, on a $\hat{\beta} \neq 0$ contredisant l'indépendance de X à Y. Ce fait peut être dû à l'insuffisance de donnée biaisant le test asymptotique de Student.

- La pvalue du test RESET vaut 1 quelque soit le modèle. Cela s'explique par le fait qu'on test si un modèle polynomiale est plus adéquate qu'un simple modèle linéaire. Or $X^k = X \ \forall k$ car X binaire. Dans notre cas, le modèle polynomiale et linéaire sont donc confondus. On peut montrer de même que si X n'est pas liée linéairement à Y alors X n'est liée d'aucune façon à Y.

- On a observé une nette dégradation de performance de la régression linéaire au cours des 2 périodes. Il est alors normal de se demander ce qu'il en est de la stabilité des paramètres au cours du temps. Le Chow test ne montre cependant aucun changement au point de l'année de rupture des deux périodes étudiées (1989). De même le CUSUM test ne montre aucune rupture quelque sur la totalité du dataset. Nous tentons d'expliquer en partie ce phénomène dans la section "Pour aller plus loin".

Conclusion

Il est probable que le SuperBowl ait un impact direct sur le rendement de l'indice SP500 le lundi suivant l'événement. Si un nombre suffisant d'agents pensent que d'autres agents réagiront à l'information, il est alors rationnel de suivre cette fausse prédiction. Cependant, l'issue du match n'a évidemment aucune influence économique à l'année. En effet, on a montré que le pouvoir prédicatif de l'événement s'estompe sur la période 1989-2013 postérieur à l'article. Les bonnes performances des critères de sélection sur la période 1967-1988 peuvent quant à elle s'expliquer par le fuit du hasard. Enfin, notons que bien que la variable binaire X soit toujours significative sur la totalité de la période étudié (1967-2013), seul 46 observations sont analysées. Ce nombre est en général insuffisant pour un test asymptotique.

Pour aller plus loin

Dans la section "Analyse des résultats", on a montré que le R^2 a diminué à cause de l'augmentation de la variance empirique de $\hat{\epsilon}$. Cependant, on aurait très bien pu obtenir les mêmes résultats de notre étude si le R^2 avait diminué à cause d'une augmentation de la variance empirique Y d'une période à une autre (impliquant la présence d'hétéroskédasticité). Pour étayer cela, on a simulé un dataset de même taille de façon à obtenir les mêmes résultats (mêmes R^2, même validation des hypothèses de la régression linéaire, X indépendant de Y) mais de façon continue afin d'obtenir un aspect plus visuel. X est ainsi la variable temporelle pour simplification et notre nouveau modèle est :

$$Y = 0.15 * X + \epsilon \qquad \epsilon \overset{iid}{\sim} \mathcal{N}(0, 0.8) \text{ sur la première période}$$

$$Y = 3 + \epsilon \qquad \epsilon \overset{iid}{\sim} \mathcal{N}(0, 1.5) \text{ sur la seconde période}$$

On a bien une augmentation de la variance de Y et donc une baisse du R^2. On a également X indépendant de Y en général comme dans notre étude mais on a forcé à ce qu'il y ait un lien sur la première période pour obtenir les mêmes résultats. On obtient (Modèle additionnel p.11) les mêmes R^2 ajustés que dans notre étude; l'absence d'auto corrélation et l'absence d'hétéroskédasticité (ce qui est faux !); la stabilité des paramètres à la coupure des 2 périodes comme le montre le Chow test; le non rejet de X sur la totalité de la période ! Ce modèle permet donc de prouver que bien que deux variables soient indépendantes mais hasardement liée initialement, l'insuffisance d'observation implique de regarder avec précaution les tests d'hypothèses. Les analyses graphiques sont alors à privilégier.

R output

First period 1967-1988

```
lm(formula = SandP ~ Dum, data = sub1_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.227886 -0.035128  0.003414  0.040172  0.138714

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.12481    0.03466  -3.601  0.00178 **
Dum          0.26577    0.04197   6.332 3.52e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0917 on 20 degrees of freedom
Multiple R-squared:  0.6672,Adjusted R-squared:  0.6505
F-statistic: 40.09 on 1 and 20 DF,  p-value: 3.522e-06


Breusch-Pagan test

data:  ols_sub1
BP = 2.6381, df = 1, p-value = 0.1043

studentized Breusch-Pagan test

data:  ols_sub1
BP = 2.2374, df = 1, p-value = 0.1347

Durbin-Watson test

data:  ols_sub1
DW = 1.9572, p-value = 0.4656
alternative hypothesis: true autocorrelation is greater than 0
```

```
Breusch-Godfrey test for serial correlation of order up to 5

data:  ols_sub1
LM test = 4.5695, df = 5, p-value = 0.4706

Shapiro-Wilk normality test

data:  residuals1
W = 0.95344, p-value = 0.369

RESET test

data:  ols_sub1
RESET = 0, df1 = 2, df2 = 18, p-value = 1

Recursive CUSUM test

data:  SandP[is.element(data$Year, seq(1967, 1988))] ~ Dum[is.element(data$Year,
S = 0.59597, p-value = 0.4212
```

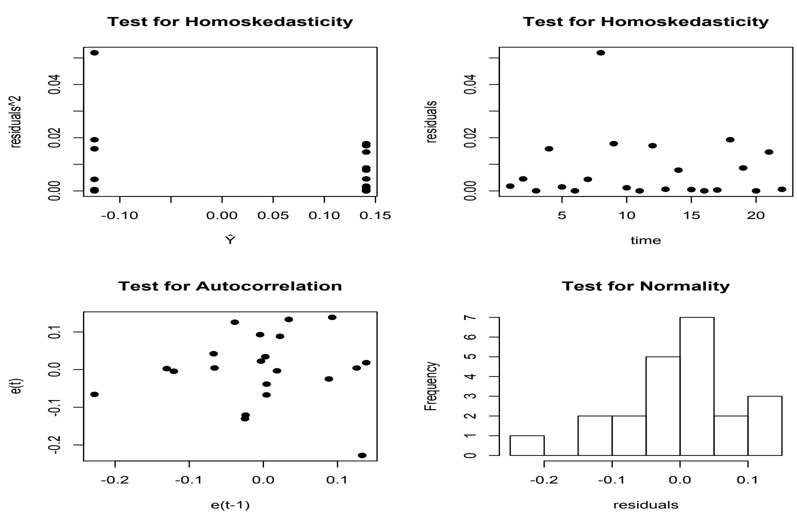

Figure 1: *Graphical inspection*

Second period 1989-2013

```
lm(formula = SandP ~ Dum, data = sub2_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.56460 -0.08258  0.04434  0.15494  0.19140

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.00700    0.09263   0.076    0.940
Dum          0.07171    0.10147   0.707    0.487

Residual standard error: 0.1853 on 22 degrees of freedom
Multiple R-squared:  0.0222,Adjusted R-squared:  -0.02225
F-statistic: 0.4994 on 1 and 22 DF,  p-value: 0.4872

Breusch-Pagan test

data:  ols_sub2
BP = 0.034602, df = 1, p-value = 0.8524

studentized Breusch-Pagan test

data:  ols_sub2
BP = 0.017543, df = 1, p-value = 0.8946

Durbin-Watson test

data:  ols_sub2
DW = 1.8866, p-value = 0.3803
alternative hypothesis: true autocorrelation is greater than 0

Breusch-Godfrey test for serial correlation of order up to 5

data:  ols_sub2
LM test = 5.756, df = 5, p-value = 0.3307

Shapiro-Wilk normality test
```

```
data:  residuals2
W = 0.86918, p-value = 0.00507

RESET test

data:  ols_sub2
RESET = 0, df1 = 2, df2 = 20, p-value = 1

Recursive CUSUM test

data:  SandP[is.element(data$Year, seq(1989, 2013))] ~ Dum[is.element(data$Year,
S = 0.36821, p-value = 0.8982
```

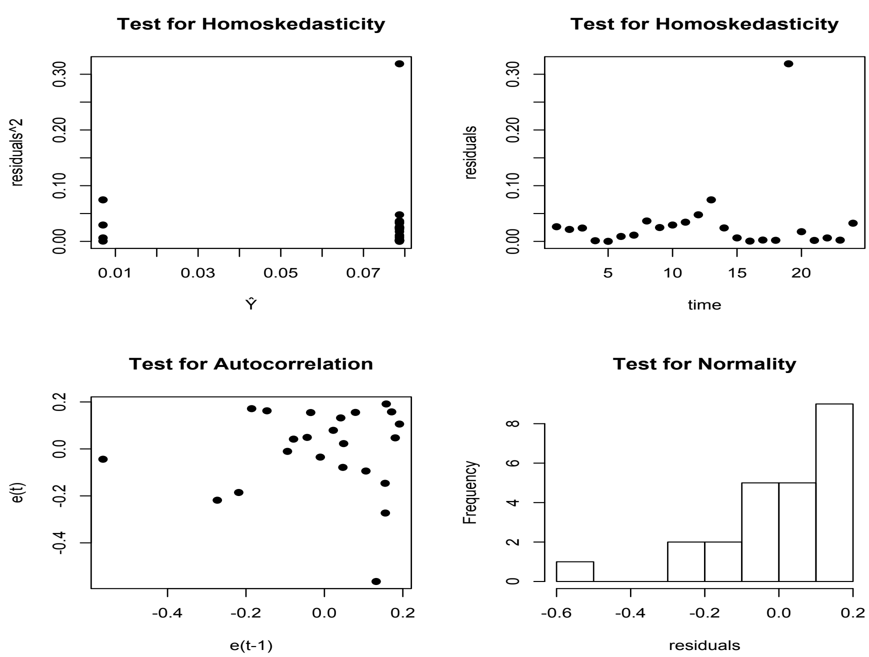

Figure 2: *Graphical inspection*

All period 1967-2013

```
lm(formula = SandP ~ Dum, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59128 -0.06843  0.02137  0.10619  0.25528

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.07688    0.04547  -1.691  0.09794 .
Dum          0.18226    0.05213   3.497  0.00109 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1508 on 44 degrees of freedom
Multiple R-squared:  0.2174,Adjusted R-squared:  0.1997
F-statistic: 12.23 on 1 and 44 DF,  p-value: 0.00109

Breusch-Pagan test

data:  ols_all
BP = 0.0071799, df = 1, p-value = 0.9325

studentized Breusch-Pagan test

data:  ols_all
BP = 0.0025211, df = 1, p-value = 0.96

Durbin-Watson test

data:  ols_all
DW = 1.858, p-value = 0.3123
alternative hypothesis: true autocorrelation is greater than 0

Breusch-Godfrey test for serial correlation of order up to 5

data:  ols_all
LM test = 7.3104, df = 5, p-value = 0.1986
```

```
Shapiro-Wilk normality test

data:  residuals_all
W = 0.89914, p-value = 0.000776

RESET test

data:  ols_all
RESET = 0, df1 = 2, df2 = 42, p-value = 1

Recursive CUSUM test

data:  SandP ~ Dum
S = 0.33685, p-value = 0.9366

Chow test

data:  SandP ~ Dum
F = 1.7615, p-value = 0.1842
```

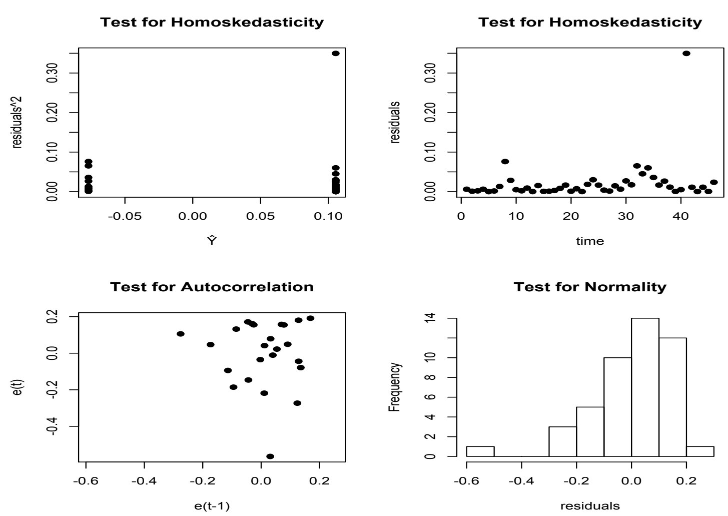

Figure 3: *Graphical inspection*

Modèle additionnel

```
Breusch-Pagan test

data:  my_ols_all
BP = 0.73083, df = 1, p-value = 0.3926

studentized Breusch-Pagan test

data:  my_ols_all
BP = 0.49789, df = 1, p-value = 0.4804

Chow test

data:  Y ~ c(0:46)
F = 1.9647, p-value = 0.1526

Ajusté R^2

0.5638908 (first period 1967-1988)
-0.006283793 (second period 1989-2013)
0.2886848 (all period 1967-2013)

Coefficients (All Dataset):
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.89187    0.40242   2.216   0.0318 *
c(0:46)      0.06684    0.01507   4.435 5.87e-05 ***
```

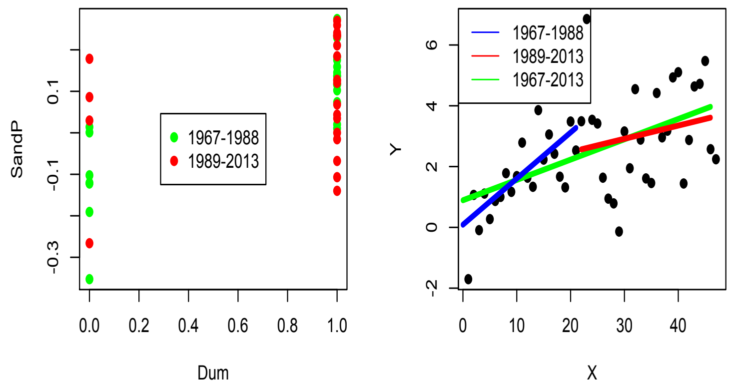

Figure 4: *SuperBowl model vs Simulated model*