

Variables de control - Agentes libres

Antonio Huerta Montellano

15 de junio del 2022

Importando las librerías:

Veamos cuál es el directorio de trabajo

```
getwd()
```

```
## [1] "/home/usuario/Documentos/Github/Proyectos/MLB_HN/Models/Instrumental_variables/Free_agent"
```

Cambiamos el directorio de trabajo y carguemos las bases de datos para el modelo lineal en el mismo chunk:

```
setwd("~/Documentos/Github/Proyectos/MLB_HN/")
free_agents <- read.csv('Data/New_Data/Models/Article/article_iv_linear_regression_fa.csv')
```

Observemos el contenido de las bases de datos de los agentes libres:

```
head(free_agents)
```

```
##      Jugador      Y      X Posicion Equipo Equipos_estado Playoffs
## 1  A.J. Burnett -471.0402 0.6333714      SP  PHI              2       14
## 2  A.J. Pierzynski 199.1753 -0.5233840      C  ATL              1       23
## 3  A.J. Pierzynski 1218.5682 -0.6992366      C  STL              2       27
## 4  A.J. Pierzynski -2515.7319 0.7312696      C  TEX              2        6
## 5  Aaron Harang  509.2721 -0.6291943      SP  LAD              5       26
## 6  Adam LaRoche  293.4260 0.6825397      DH  WSH              1        2
## Pennants WS
## 1      7  2
## 2     18  3
## 3     23 11
## 4      2  0
## 5     25  6
## 6      1  0
```

Dummificación

Creemos las variables dummy correspondientes a las variables categóricas de la posición del jugador y al estad

```
dummy <- dummy_cols(free_agents, select_columns = c('Posicion', 'Equipo'),
                    remove_selected_columns = TRUE)
```

Notemos que ahora hay muchas columnas debido a la dummificación y la gran cantidad de categorías tanto para la posición que puede ocupar un jugador como para el equipo al que pertenecen (29):

```
head(dummy)
```

```
##      Jugador      Y      X Equipos_estado Playoffs Pennants WS
## 1  A.J. Burnett -471.0402 0.6333714      2      14      7  2
## 2  A.J. Pierzynski 199.1753 -0.5233840      1      23     18  3
## 3  A.J. Pierzynski 1218.5682 -0.6992366      2      27     23 11
## 4  A.J. Pierzynski -2515.7319 0.7312696      2       6      2  0
## 5  Aaron Harang   509.2721 -0.6291943      5      26     25  6
## 6  Adam LaRoche   293.4260 0.6825397      1       2      1  0
##  Posicion_1B Posicion_2B Posicion_3B Posicion_C Posicion_CF Posicion_DH
## 1      0      0      0      0      0      0
## 2      0      0      0      1      0      0
## 3      0      0      0      1      0      0
## 4      0      0      0      1      0      0
## 5      0      0      0      0      0      0
## 6      0      0      0      0      0      1
##  Posicion_LF Posicion_RF Posicion_RP Posicion_RP/CL Posicion_SP Posicion_SS
## 1      0      0      0      0      1      0
## 2      0      0      0      0      0      0
## 3      0      0      0      0      0      0
## 4      0      0      0      0      0      0
## 5      0      0      0      0      1      0
## 6      0      0      0      0      0      0
##  Equipo_ARI Equipo_ATL Equipo_BAL Equipo_BOS Equipo_CHC Equipo_CHW Equipo_CIN
## 1      0      0      0      0      0      0      0
## 2      0      1      0      0      0      0      0
## 3      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0
##  Equipo_COL Equipo_DET Equipo_HOU Equipo_KC Equipo_LAA Equipo_LAD Equipo_MIA
## 1      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0
## 5      0      0      0      0      0      1      0
## 6      0      0      0      0      0      0      0
##  Equipo_MIL Equipo_MIN Equipo_NYM Equipo_NYY Equipo_OAK Equipo_PHI Equipo_PIT
## 1      0      0      0      0      0      1      0
## 2      0      0      0      0      0      0      0
## 3      0      0      0      0      0      0      0
## 4      0      0      0      0      0      0      0
## 5      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0
##  Equipo_SD Equipo_SEA Equipo_SF Equipo_STL Equipo_TB Equipo_TEX Equipo_TOR
## 1      0      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0
```

```
## 3      0      0      0      1      0      0      0
## 4      0      0      0      0      0      1      0
## 5      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0
## Equipo_WSH
## 1      0
## 2      0
## 3      0
## 4      0
## 5      0
## 6      1
```

Creación del modelo

Para ello y evitar errores, lo haremos con los nombres explícitos, obtengamos los nombres de las columnas

```
colnames(dummy)
```

```
## [1] "Jugador"      "Y"            "X"            "Equipos_estado"
## [5] "Playoffs"     "Pennants"     "WS"           "Posicion_1B"
## [9] "Posicion_2B"   "Posicion_3B"   "Posicion_C"    "Posicion_CF"
## [13] "Posicion_DH"   "Posicion_LF"   "Posicion_RF"   "Posicion_RP"
## [17] "Posicion_RP/CL" "Posicion_SP"   "Posicion_SS"   "Equipo_ARI"
## [21] "Equipo_ATL"    "Equipo_BAL"    "Equipo_BOS"    "Equipo_CHC"
## [25] "Equipo_CHW"    "Equipo_CIN"    "Equipo_COL"    "Equipo_DET"
## [29] "Equipo_HOU"    "Equipo_KC"     "Equipo_LAA"    "Equipo_LAD"
## [33] "Equipo_MIA"    "Equipo_MIL"    "Equipo_MIN"    "Equipo_NYM"
## [37] "Equipo_NYY"    "Equipo_OAK"    "Equipo_PHI"    "Equipo_PIT"
## [41] "Equipo_SD"     "Equipo_SEA"    "Equipo_SF"     "Equipo_STL"
## [45] "Equipo_TB"     "Equipo_TEX"    "Equipo_TOR"    "Equipo_WSH"
```

Cambiamos el nombre de la columna “Posicion_RP/CL” por “Posicion_RP_CL” para evitar problemas en los algoritmos:

```
names(dummy)[names(dummy) == 'Posicion_RP/CL'] <- 'Posicion_RP_CL'
```

```
formula_iv <- Y ~ X | Playoffs + Pennants + WS+ Equipos_estado + Posicion_1B + Posicion_2B + Posicion_3B
iv_model <- ivreg(formula_iv, data = dummy)
summary(iv_model, diagnostics=TRUE)
```

```
##
## Call:
## ivreg(formula = formula_iv, data = dummy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3613.5  -582.8   139.0   594.4  2745.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -114.49      64.14  -1.785   0.0751 .
```

```

## X          -54.88      140.57  -0.390   0.6965
##
## Diagnostic tests:
##              df1 df2 statistic          p-value
## Weak instruments 42 320      4.170 0.0000000000000989 ***
## Wu-Hausman      1 360      0.463          0.497
## Sargan          44  NA     53.026          0.165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1014 on 361 degrees of freedom
## Multiple R-Squared: -0.002154, Adjusted R-squared: -0.00493
## Wald test: 0.1524 on 1 and 361 DF, p-value: 0.6965

```

En comparación con el modelo sin instrumentos -donde se obtuvo un $p - value = 0.738$ -, aquí mejoró dicho valor. Sin embargo, no es suficiente para que la variable X se estadísticamente significativa para el modelo. También dio positivo para la prueba de instrumentos débiles.