Are left-handers more successful in tennis ?

Alexandre Dossin and Ganesh Vedarattiname

ENAC

April 8, 2016

Table of contents

What is the aim of our project ?

- Perform an econometrics study on international tennis players
- Create a database with official data
- Propose a model and test it with EViews

Gathering of data

- All of our data come from the official ATP website
- Cross sectional data from March 2016 and April 1998
- For our model : 200 observations with 12 variables

Specificity

We are only interested here in male international tennis players.

What we wanted to show for this project ?

That left-handed tennis players are more successful.

What could be the reasons ?

- Right-handed tennis players are not used to play against left-handed players
- Left-handed players are used to play against right-handed players
- Left-handers often serve right-handers players on their backhand at critical points in the game
- Historically, there has been some very good left-handed tennis players (Rod Laver, John McEnroe, Raphael Nadal, ...)
- Left-handers would more use their left brain and therefore have a better spatial vision and more creativity

What we wanted to show for this project ?

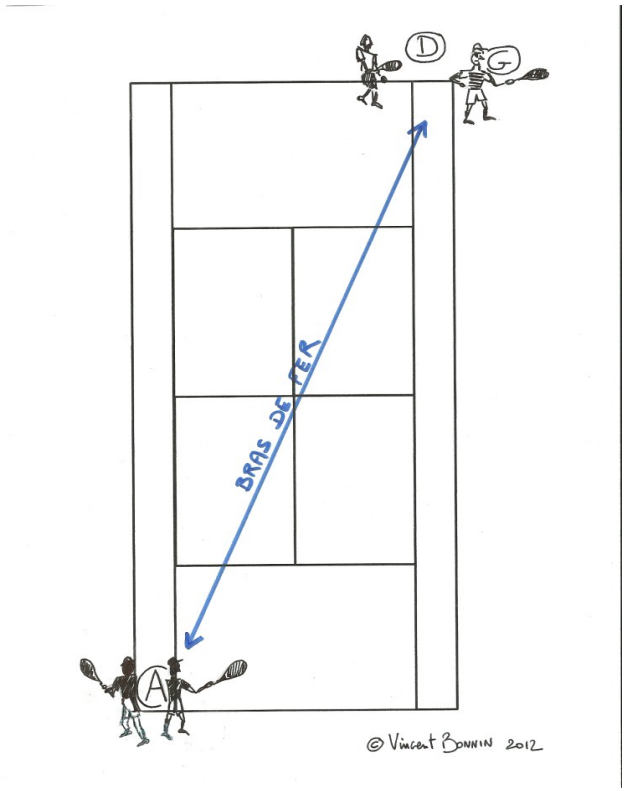

Figure: Advantage for left-handers at critical points for winning the game

What we wanted to show for this project ?

Player	Rank	Hand
Marcelo Rios	1	**Left-hander**
Pete Sampras	2	**Right-hander**
Petr Korda	3	**Left-hander**
Patrick Rafter	4	**Right-hander**
Greg Rusedski	5	**Left-hander**

Table: The ATP top 5 in April 1998.

Definition of the variables

Dependant variables

- POINTS : ATP points, calculated by an objective merit-based method used for determining entry and seeding in all tournaments for both singles on the past 52 weeks
- W_L : win-loss ratio of the player's career, between 0 and 1000
- INV_RNK : inverted ranking of the player based on the number of ATP points (1 is the worst player, 200 is the best one)

Parameters

- AGE : the age of the player in years
- HEIGHT : the height of the player in cm
- WEIGHT : the weight of the player in kg
- EXPER : the player's experience in years at the date of the ranking
- FSPW : First Service Point Win in %
- BPS : Break Points Saved in %
- SGW : Service Games Won in %
- BPC : Break Points Converted in %
- RGW : Return Games Won in %

Parameters

- W_L RIGHT : Win-Loss Ratio of the career's player against right-hander, between 0 and 1000
- W_L LEFT : Win-Loss Ratio of the career's player against left-hander, between 0 and 1000

Introduction

Let's have a quick interest in the best Win-Loss ratios of all ATP tennis players since 1973.

Model

$$W\_L = \beta_0 + \beta_1 * HAND + u$$

Dependent Variable: W_L
Method: Least Squares
Date: 04/07/16 Time: 22:23
Sample: 1 100
Included observations: 100

	Coefficient	Std. Error	t-Statistic	Prob.
C	671.2874	6.128286	109.5392	0.0000
HAND	50.86649	16.99681	2.992709	0.0035

R-squared	0.083738	Mean dependent var	677.9000
Adjusted R-squared	0.074388	S.D. dependent var	59.41338
S.E. of regression	57.16085	Akaike info criterion	10.94941
Sum squared resid	320201.5	Schwarz criterion	11.00152
Log likelihood	-545.4706	Hannan-Quinn criter.	10.97050
F-statistic	8.956304	Durbin-Watson stat	0.163401
Prob(F-statistic)	0.003499		

Figure: First model output

- M.E : Lefties tennis players gain more than 5 % in Win-Loss ratio
- Cross sectional data for 43 years : best tennis players who have ever existed
- Same thing for other cross-sectional data in a one year ranking ?

Model

Proposed model

$$INV\_RNK = \beta_0 + \beta_1 * AGE + \beta_2 * EXPER + \beta_3 * HEIGHT + \beta_4 * WEIGHT + \beta_5 * HAND + \beta_6 * FSPW + \beta_7 * BPS + \beta_8 * SGW + \beta_9 * BPC + \beta_{10} * RGW + u \quad (1)$$

March 2016

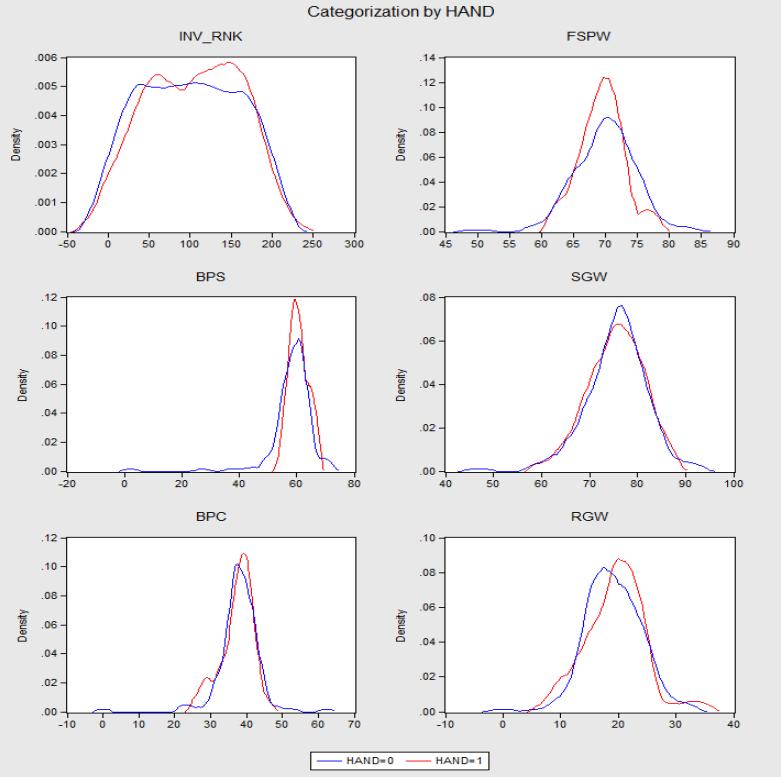

Figure: An intuition

March 2016

Dependent Variable: INV_RNK
Method: Least Squares
Date: 04/07/16 Time: 23:40
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	-440.2657	61.89356	-7.113272	0.0000
EXPER	0.325435	1.071054	0.303846	0.7616
HAND	-2.520565	8.887391	-0.283611	0.7770
FSPW	-0.438874	1.437734	-0.305254	0.7605
BPS	-2.171860	1.430764	-1.517972	0.1307
SGW	6.996218	1.322652	5.289537	0.0000
BPC	2.886870	1.254678	2.300885	0.0225
RGW	2.768715	1.036832	2.670361	0.0082

R-squared	0.438604	Mean dependent var	100.5000
Adjusted R-squared	0.418137	S.D. dependent var	57.87918
S.E. of regression	44.15022	Akaike info criterion	10.45225
Sum squared resid	374254.4	Schwarz criterion	10.58418
Log likelihood	-1037.225	Hannan-Quinn criter.	10.50564
F-statistic	21.42925	Durbin-Watson stat	0.832040
Prob(F-statistic)	0.000000		

Figure: First regression

March 2016

Dependent Variable: INV_RNK
Method: Least Squares
Date: 04/07/16 Time: 23:58
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	-355.8909	67.08952	-5.304716	0.0000
AGE	-4.107654	1.833577	-2.240242	0.0262
EXPER	4.318726	1.869946	2.309546	0.0220
HAND	5.362620	8.822092	0.607863	0.5440
FSPW	2.742032	1.085648	2.525711	0.0124
SGW	3.171690	0.759908	4.173779	0.0000
RGW	4.929020	0.725241	6.796387	0.0000

R-squared	0.464781	Mean dependent var	100.5000
Adjusted R-squared	0.448142	S.D. dependent var	57.87918
S.E. of regression	42.99680	Akaike info criterion	10.39450
Sum squared resid	356803.8	Schwarz criterion	10.50994
Log likelihood	-1032.450	Hannan-Quinn criter.	10.44122
F-statistic	27.93333	Durbin-Watson stat	0.802580
Prob(F-statistic)	0.000000		

Figure: Second regression

March 2016

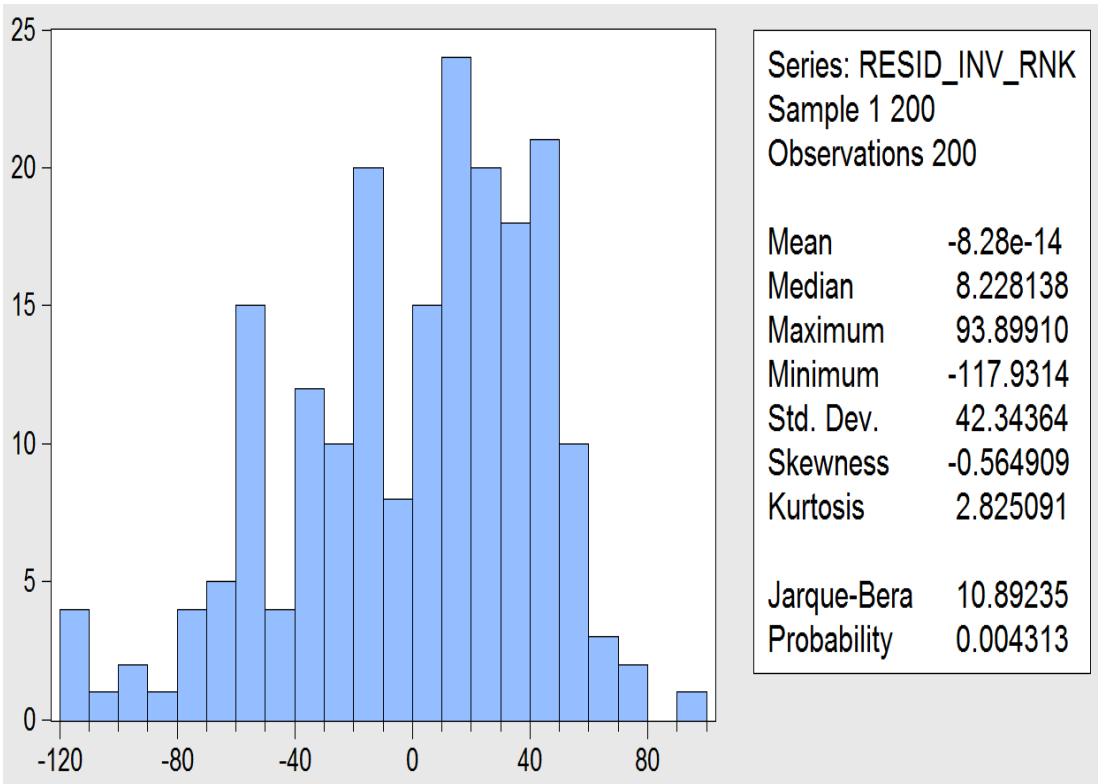

Figure: Error Analysis

March 2016

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.262724	Prob. F(6,193)	0.9536
Obs*R-squared	1.620282	Prob. Chi-Square(6)	0.9511
Scaled explained SS	1.376891	Prob. Chi-Square(6)	0.9672

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/08/16 Time: 00:02
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	1921.236	3812.680	0.503907	0.6149
AGE	56.33165	104.2017	0.540602	0.5894
EXPER	-45.42490	106.2686	-0.427454	0.6695
HAND	-220.0400	501.3571	-0.438889	0.6612
FSPW	1.058626	61.69706	0.017158	0.9863
SGW	-8.570798	43.18539	-0.198465	0.8429
RGW	-32.78499	41.21526	-0.795458	0.4273

R-squared	0.008101	Mean dependent var	1784.019
Adjusted R-squared	-0.022735	S.D. dependent var	2416.185
S.E. of regression	2443.496	Akaike info criterion	18.47462
Sum squared resid	1.15E+09	Schwarz criterion	18.59006
Log likelihood	-1840.462	Hannan-Quinn criter.	18.52134
F-statistic	0.262724	Durbin-Watson stat	1.194542
Prob(F-statistic)	0.953569		

Figure: Homoscedasticity or Heteroscedasticity ?

Conclusion for 2016

- HAND has no influence
- "In my time, not everybody had an excellent backhand, so that was an advantage. But nowadays, there are no weak backhands. Everybody is good on both sides." John McEnroe, The Telegraph, 2015
- There is only one leftie in the top 20 in March 2016 : Nadal.

April 1998

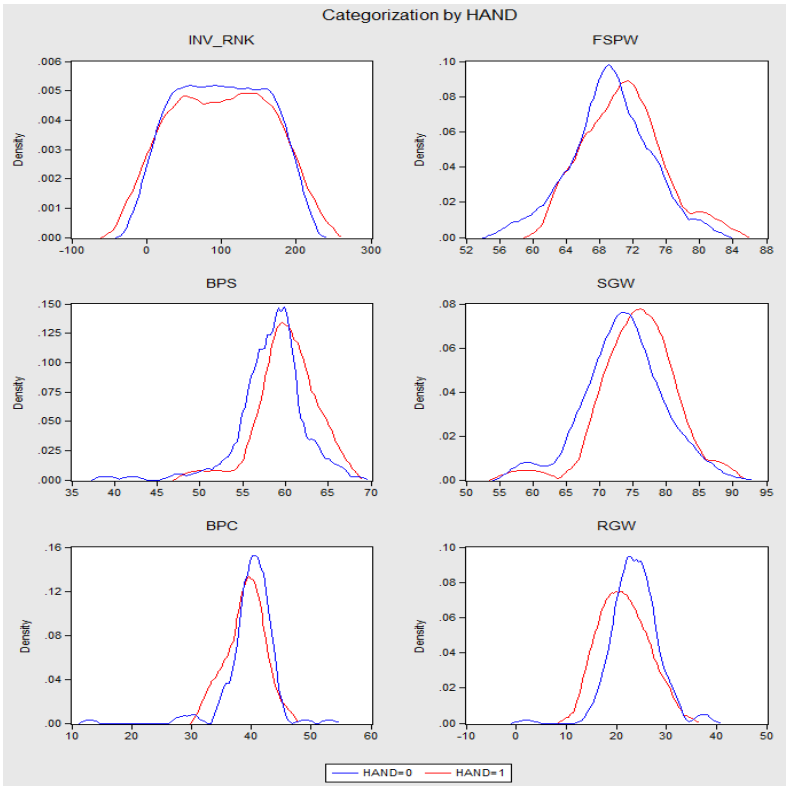

Figure: First approach

April 1998

Dependent Variable: INV_RNK
Method: Least Squares
Date: 04/08/16 Time: 13:55
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	-423.5251	44.49042	-9.519467	0.0000
SGW	5.720060	0.551262	10.37630	0.0000
RGW	4.337849	0.701732	6.181629	0.0000
HAND	-3.034358	8.690329	-0.349165	0.7273

R-squared	0.419900	Mean dependent var	100.5000
Adjusted R-squared	0.411021	S.D. dependent var	57.87918
S.E. of regression	44.41935	Akaike info criterion	10.44502
Sum squared resid	386723.4	Schwarz criterion	10.51099
Log likelihood	-1040.502	Hannan-Quinn criter.	10.47172
F-statistic	47.29100	Durbin-Watson stat	0.756664
Prob(F-statistic)	0.000000		

Figure: First regression

April 1998

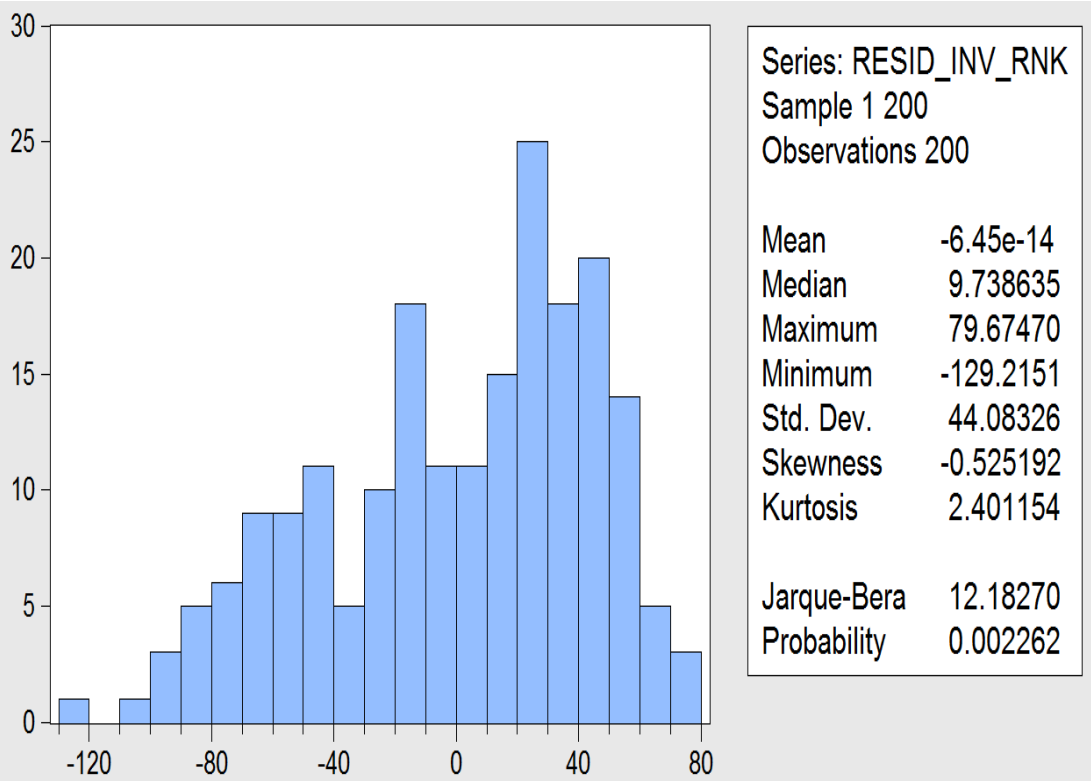

Figure: Errors analysis

April 1998

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	3.578935	Prob. F(3,196)	0.0149
Obs*R-squared	10.38693	Prob. Chi-Square(3)	0.0155
Scaled explained SS	6.988681	Prob. Chi-Square(3)	0.0723

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/08/16 Time: 00:34
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1457.976	2254.830	-0.646601	0.5186
HAND	132.2932	440.4367	0.300368	0.7642
SGW	8.876429	27.93864	0.317712	0.7510
RGW	115.9305	35.56467	3.259709	0.0013

R-squared	0.051935	Mean dependent var	1933.617
Adjusted R-squared	0.037423	S.D. dependent var	2294.573
S.E. of regression	2251.228	Akaike info criterion	18.29614
Sum squared resid	9.93E+08	Schwarz criterion	18.36210
Log likelihood	-1825.614	Hannan-Quinn criter.	18.32283
F-statistic	3.578935	Durbin-Watson stat	1.293915
Prob(F-statistic)	0.014900		

Figure: Homoscedasticity or not ?

April 1998

Dependent Variable: INV_RNK
Method: Least Squares
Date: 04/07/16 Time: 23:47
Sample: 1 200
Included observations: 200
White Heteroskedasticity-Consistent Standard Errors & Covariance

	Coefficient	Std. Error	t-Statistic	Prob.
C	-423.5251	39.69216	-10.67024	0.0000
SGW	5.720060	0.488415	11.71148	0.0000
RGW	4.337849	0.794267	5.461453	0.0000
HAND	-3.034358	8.348024	-0.363482	0.7166

R-squared	0.419900	Mean dependent var	100.5000
Adjusted R-squared	0.411021	S.D. dependent var	57.87918
S.E. of regression	44.41935	Akaike info criterion	10.44502
Sum squared resid	386723.4	Schwarz criterion	10.51099
Log likelihood	-1040.502	Hannan-Quinn criter.	10.47172
F-statistic	47.29100	Durbin-Watson stat	0.756664
Prob(F-statistic)	0.000000		

Figure: Correction for Heteroscedasticity

Conclusion for 1998

- 3 lefties in the top 5
- Yet, HAND has no influence
- "I dont think it makes a big impact if you are a lefty or not",Thomas Muster was one of the best players in the world in the 1990s, The Telegraph

Quick refresher

- Hypothesis 1 : Right-handed tennis players are not used to play against left-handed players
- Hypothesis 2 : Left-handed players are used to play against right-handed players
- Hypothesis 3 : Left-handers often serve right-handers players on their backhand at critical points in the game

March 2016

Dependent Variable: W_L_LEFT
Method: Least Squares
Date: 04/08/16 Time: 15:16
Sample: 1 99
Included observations: 99

	Coefficient	Std. Error	t-Statistic	Prob.
C	483.0000	45.18389	10.68965	0.0000
1-HAND	39.40000	48.76315	0.807987	0.4211

R-squared	0.006685	Mean dependent var	516.8283
Adjusted R-squared	-0.003555	S.D. dependent var	168.7629
S.E. of regression	169.0626	Akaike info criterion	13.11841
Sum squared resid	2772470.	Schwarz criterion	13.17084
Log likelihood	-647.3613	Hannan-Quinn criter.	13.13962
F-statistic	0.652843	Durbin-Watson stat	1.579637
Prob(F-statistic)	0.421075		

Figure: Hypothesis 1 testing

March 2016

Dependent Variable: W_L_RIGHT
Method: Least Squares
Date: 04/08/16 Time: 15:00
Sample: 1 99
Included observations: 99

	Coefficient	Std. Error	t-Statistic	Prob.
C	488.6471	14.97558	32.62959	0.0000
HAND	-22.57563	39.82331	-0.566895	0.5721

R-squared	0.003302	Mean dependent var	485.4545
Adjusted R-squared	-0.006973	S.D. dependent var	137.5892
S.E. of regression	138.0681	Akaike info criterion	12.71337
Sum squared resid	1849090.	Schwarz criterion	12.76579
Log likelihood	-627.3116	Hannan-Quinn criter.	12.73458
F-statistic	0.321370	Durbin-Watson stat	0.991511
Prob(F-statistic)	0.572095		

Figure: Hypothesis 2 testing

March 2016

Dependent Variable: BPS
Method: Least Squares
Date: 04/08/16 Time: 00:42
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	58.57558	0.509140	115.0480	0.0000
HAND	2.102990	1.360734	1.545482	0.1238

R-squared	0.011919	Mean dependent var	58.87000
Adjusted R-squared	0.006929	S.D. dependent var	6.700566
S.E. of regression	6.677311	Akaike info criterion	6.645258
Sum squared resid	8828.125	Schwarz criterion	6.678241
Log likelihood	-662.5258	Hannan-Quinn criter.	6.658605
F-statistic	2.388513	Durbin-Watson stat	1.949394
Prob(F-statistic)	0.123827		

Figure: Hypothesis 3 testing

April 1998

Dependent Variable: W_L_LEFT
Method: Least Squares
Date: 04/08/16 Time: 15:38
Sample: 1 100
Included observations: 99

	Coefficient	Std. Error	t-Statistic	Prob.
C	548.8125	34.45637	15.92775	0.0000
1-HAND	-18.22214	37.63120	-0.484230	0.6293

R-squared	0.002411	Mean dependent var	533.5354
Adjusted R-squared	-0.007873	S.D. dependent var	137.2861
S.E. of regression	137.8255	Akaike info criterion	12.70985
Sum squared resid	1842599.	Schwarz criterion	12.76228
Log likelihood	-627.1375	Hannan-Quinn criter.	12.73106
F-statistic	0.234478	Durbin-Watson stat	1.427054
Prob(F-statistic)	0.629314		

Figure: Hypothesis 1 testing

April 1998

Dependent Variable: W_L_RIGHT
Method: Least Squares
Date: 04/08/16 Time: 15:40
Sample: 1 100
Included observations: 100

	Coefficient	Std. Error	t-Statistic	Prob.
C	519.9048	10.61232	48.99067	0.0000
HAND	5.970238	26.53080	0.225030	0.8224

R-squared	0.000516	Mean dependent var	520.8600
Adjusted R-squared	-0.009682	S.D. dependent var	96.79605
S.E. of regression	97.26353	Akaike info criterion	12.01252
Sum squared resid	927099.0	Schwarz criterion	12.06463
Log likelihood	-598.6261	Hannan-Quinn criter.	12.03361
F-statistic	0.050639	Durbin-Watson stat	0.966428
Prob(F-statistic)	0.822424		

Figure: Hypothesis 2 testing

April 1998

Dependent Variable: BPS
Method: Least Squares
Date: 04/08/16 Time: 00:59
Sample: 1 200
Included observations: 200

	Coefficient	Std. Error	t-Statistic	Prob.
C	58.37126	0.291363	200.3389	0.0000
HAND	1.871167	0.717285	2.608680	0.0098

R-squared	0.033228	Mean dependent var	58.68000
Adjusted R-squared	0.028345	S.D. dependent var	3.819758
S.E. of regression	3.765233	Akaike info criterion	5.499446
Sum squared resid	2807.043	Schwarz criterion	5.532430
Log likelihood	-547.9446	Hannan-Quinn criter.	5.512794
F-statistic	6.805211	Durbin-Watson stat	1.561546
Prob(F-statistic)	0.009783		

Figure: Hypothesis 3 testing

Conclusion

- Being left hander is not an advantage in March 2016 nor in April 1998 regarding the ATP ranking
- But lefties have a mild advantage regarding critical points such as break points.
- Panel data needed to follow the evolution of left-handed players through time.