Online Dating Applications

Milad Larkem, Axelle Marcy

ENAC, IENAC16

April 13, 2018

Questions of interest

- Who is more likely to use online dating applications ?
- Does the use of social networks / dating applications make you happier ?

Table of contents

1. Introduction
2. Variables
3. First model
4. Second model
5. Conclusion

Purpose

- Collect data on people's behaviour
- Create an Econometrics model of this behaviour
- Analyse the results on Eviews and find an answer to our issues

Data Gathering

Our Survey

- Use of a Google Form shared on social networks
- n = 198 answers collected in 5 days
- Majority of students (> 80%)

Figure: Our survey

dependent Variables

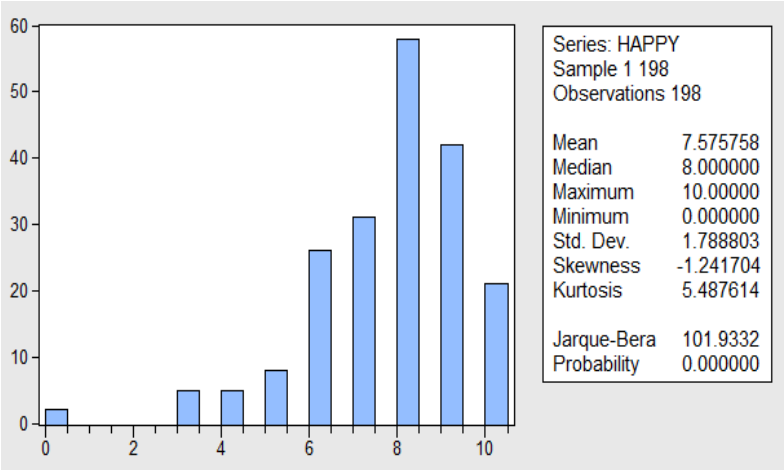
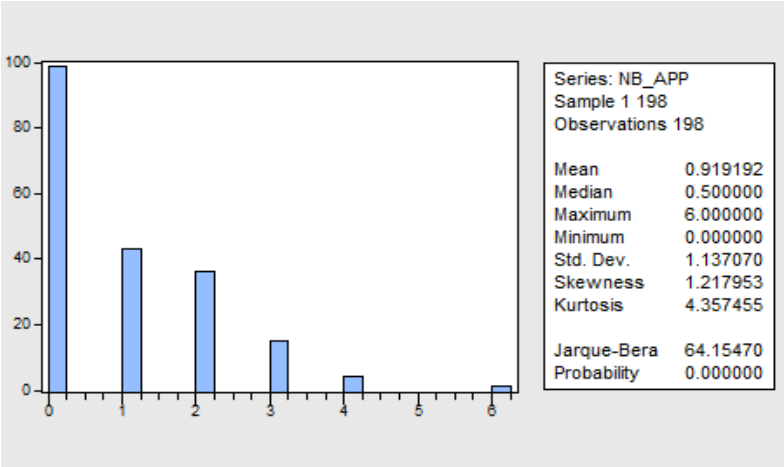


Figure: Histogram of HAPPY

- $\frac{n}{6}SK^2 = 50.9 > \chi^2_{0.95} = 3.8$

- HAPPY : degree of happiness (from 0 to 10)

Dependent Variables



- $\frac{n}{6} SK^2 = 49.0 > \chi^2_{0.95} = 3.8$

Figure: Histogram of NB_APP

- NB_APP : number of online dating apps used

Explanatory Variables

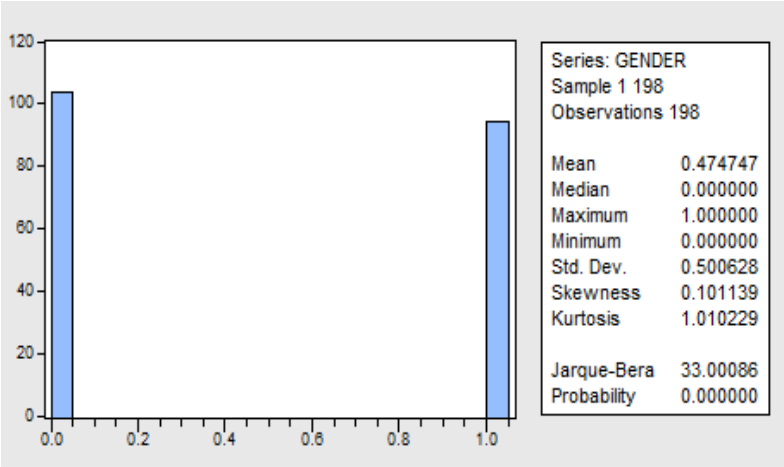


Figure: Histogram of GENDER

Balanced variable

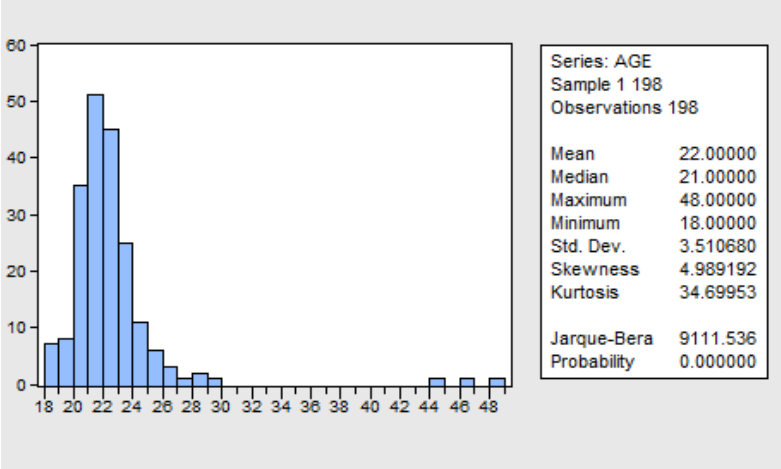


Figure: Histogram of AGE

Low variation

Choice of variables

Purpose

- Analysis of the correlation matrix to avoid multicollinearity
- Identify insignificant variables (with $p > 0.15$)
- Try to improve the ajusted R-Squared

Analysis of the correlation matrix

Variables	**Correlated with**	**Coefficient**
AGE	FRIEND	+ 0,50
AGE	PROF	+ 0,42
FIELD	GENDER	+ 0,51
GENDER	HEIGHT	- 0,72
GENDER	WEIGHT	- 0,57
HEIGHT	WEIGHT	+ 0,57

Table: Main significant coefficients in the correlation matrix

First model

Question of interest

Users of dating applications : What do they have in common ?

Linear Regression

$$NB_APP = \beta_0 + \beta_1 AGE + \beta_2 FIELD + \beta_3 FRIEND$$

$$+ \beta_4 GENDER + \beta_5 HAPPY + \beta_6 NB_PUBLICATION$$
$$+ \beta_7 NB_STORIES + \beta_8 RELATION + \beta_9 SOCNET$$
$$k = 9$$

Variables and marginal effect

Variables	**Expected Effect**	**Real Effect**	**Significant?**
AGE	=	+	NO
FIELD	-	+	NO
FRIEND	-	-	NO
GENDER	-	-	YES
HAPPY	-	-	YES
NB_PUBLICATION	=	+	YES
NB_STORIES	=	+	NO
RELATION	-	-	YES
SOCNET	-	+	YES

Table: Comparison of expected and real marginal effects

First model

Regression output

Dependent Variable: NB_APP
Method: Least Squares
Date: 04/11/18 Time: 17:09
Sample: 1 198
Included observations: 178

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.680148	0.805808	0.844057	0.3998
AGE	0.043719	0.033561	1.302696	0.1945
FIELD	0.072501	0.046468	1.560237	0.1206
FRIEND	-0.000441	0.000280	-1.573225	0.1175
GENDER	-0.534787	0.178343	-2.998638	0.0031
HAPPY	-0.126004	0.043347	-2.906891	0.0041
NB_PUBLICATION	0.027798	0.006428	4.324850	0.0000
NB_STORIES	0.036237	0.026342	1.375663	0.1708
RELATION	-0.297161	0.152070	-1.954109	0.0523
SOCNET	0.190483	0.077845	2.446973	0.0154

R-squared	0.296326	Mean dependent var	0.910112
Adjusted R-squared	0.258629	S.D. dependent var	1.136358
S.E. of regression	0.978437	Akaike info criterion	2.848820
Sum squared resid	160.8331	Schwarz criterion	3.027572
Log likelihood	-243.5450	Hannan-Quinn criter.	2.921309
F-statistic	7.860751	Durbin-Watson stat	1.479323
Prob(F-statistic)	0.000000		

Homoscedasticity ?

- H0 : Homoscedasticity
- H1 : Heteroscedasticity

Heteroskedasticity Test: White

F-statistic	0.828401	Prob. F(53,124)	0.7785
Obs*R-squared	46.54493	Prob. Chi-Square(53)	0.7220
Scaled explained SS	42.97228	Prob. Chi-Square(53)	0.8358

Figure: White Test for Heteroscedasticity

Conclusion : Homoscedasticity

What about classical assumptions ?

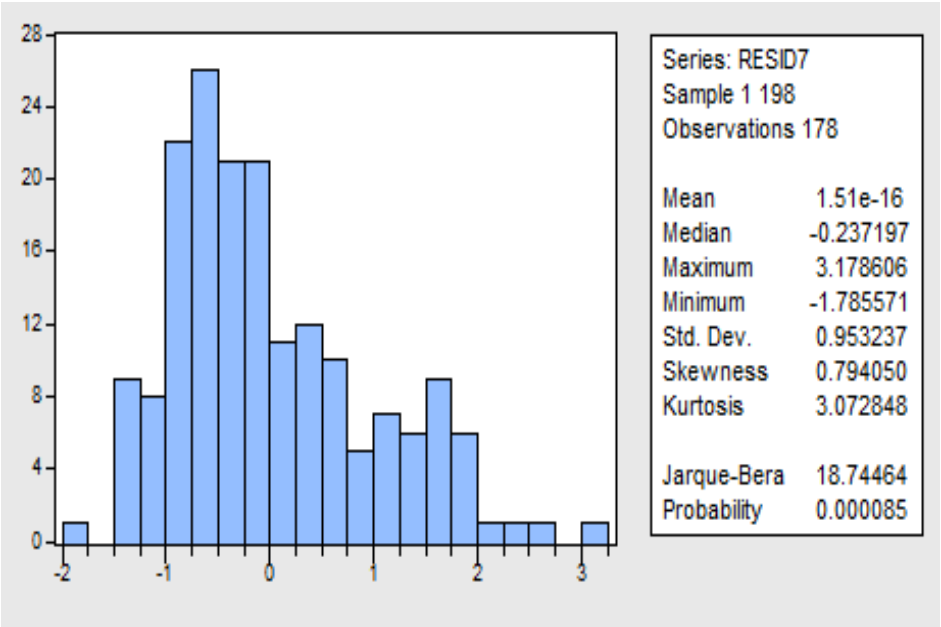


Figure: Graph of Residuals

- Mean $= 0$
- JBcorrected $= 18{,}7$
- We can NOT assume the normality of the errors

Conclusion

- Variables are jointly significant
- No heteroscedasticity
- Adjusted $R^2 \approx 25\%$: good model ?

Introduction
First Equation
Variables and marginal effect
Second Equation
What about Classical Assumptions ?
What about homoscedasticity ?
Conclusion

Introduction

Question of interest

- Users of dating applications and social networks : Are you happy ?
- k = 17
- dependent variable : HAPPY

Introduction
First Equation
Variables and marginal effect
Second Equation
What about Classical Assumptions ?
What about homoscedasticity ?
Conclusion

First Equation

Linear Regression

$$HAPPY = \beta_0 + \beta_1 AGE + \beta_2 FIELD + \beta_3 FRIEND + \beta_4 GENDER$$

$$+ \beta_5 HEIGHT + \beta_6 INC + \beta_7 NB_APP + \beta_8 NB_PUBLICATION + \beta_9 NB_STORIES + \beta_{10} PARTIES + \beta_{11} PROF + \beta_{12} RELATION + \beta_{13} SOCNET + \beta_{14} SPORT + \beta_{15} TPS_SONET + \beta_{16} WEIGHT$$

$$k = 17$$

Regression Output (1)

Dependent Variable: HAPPY
Method: Least Squares
Date: 04/11/18 Time: 16:36
Sample: 1 198
Included observations: 178

	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.719423	4.111730	-0.174969	0.8613
AGE	0.044942	0.064105	0.701066	0.4843
FIELD	0.025360	0.082087	0.308936	0.7578
FRIEND	-0.000331	0.000509	-0.651655	0.5156
GENDER	-0.429975	0.409388	-1.050287	0.2952
HEIGHT	0.059291	0.023705	2.501220	0.0134
INC	0.072890	0.160440	0.454310	0.6502
NB_APP	-0.358840	0.128900	-2.783874	0.0060
NB_PUBLICATION	-0.006112	0.011919	-0.512824	0.6088
NB_STORIES	0.065135	0.046495	1.400911	0.1632
PARTIES	-0.003701	0.060341	-0.061339	0.9512
PROF	-0.046065	0.112578	-0.409186	0.6829
RELATION	0.798827	0.254869	3.134260	0.0020
SOCNET	0.215563	0.138895	1.551986	0.1226
SPORT	0.024134	0.022146	1.089751	0.2775
TPS_SONET	-0.202146	0.093434	-2.163511	0.0320
WEIGHT	-0.045737	0.016683	-2.741614	0.0068

Figure: First equation's regression output (1)

Regression Output (2)

R-squared	0.260878	Mean dependent var	7.567416
Adjusted R-squared	0.187425	S.D. dependent var	1.840675
S.E. of regression	1.659238	Akaike info criterion	3.941226
Sum squared resid	443.2446	Schwarz criterion	4.245105
Log likelihood	-323.7692	Hannan-Quinn criter.	4.064457
F-statistic	3.551633	Durbin-Watson stat	2.195597
Prob(F-statistic)	0.000018		

- Ajusted R-squared = 18%
- Proba(F-stat) = 0,000018
- Many insignificant variables

Figure: First equation's regression output (2)

Variables and marginal effect (1)

Variables	**Expected Effect**	**Real Effect**	**Significant?**
AGE	-	+	NO
FIELD	?	+	NO
FRIEND	+	-	NO
GENDER	?	-	NO
HEIGHT	+	+	YES
INC	+	+	NO
NB_APP	-	-	YES
NB_PUBLICATION	-	-	NO
NB_STORIES	-	+	NO

Table: Comparison of expected and real marginal effects (1)

Variables and marginal effect (2)

Variables	Expected Effect	Real Effect	Significant
PARTIES	+	-	NO
PROF	?	-	NO
RELATION	+	+	YES
SOCNET	-	+	YES
SPORT	+	+	NO
TPS_SONET	-	-	YES
WEIGHT	-	-	YES

Table: Comparison of expected and real marginal effects (2)

Purpose and Linear Regression

Linear Regression

$$HAPPY = \beta_0 + \beta_4 GENDER + \beta_5 HEIGHT + \beta_7 NB_APP +$$

$$\beta_9 NB_STORIES + \beta_{12} RELATION + \beta_{13} SOCNET + \beta_{14} SPORT +$$
$$\beta_{15} TPS_SONET + \beta_{16} WEIGHT$$
$$k = 10$$

Regression Output - Second Equation

Dependent Variable: HAPPY
Method: Least Squares
Date: 04/12/18 Time: 22:43
Sample: 1 198
Included observations: 194

	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.394325	2.502946	-0.557074	0.5782
HEIGHT	0.067407	0.017394	3.875285	0.0001
NB_APP	-0.274820	0.109045	-2.520251	0.0126
NB_STORIES	0.050639	0.040333	1.255537	0.2109
RELATION	0.761867	0.228099	3.340072	0.0010
SOCNET	0.183097	0.118153	1.549655	0.1229
SPORT	0.038449	0.019124	2.010442	0.0458
TPS_SONET	-0.225225	0.084243	-2.673510	0.0082
WEIGHT	-0.046045	0.014521	-3.170970	0.0018

R-squared	0.240139	Mean dependent var	7.577320
Adjusted R-squared	0.207280	S.D. dependent var	1.805777
S.E. of regression	1.607771	Akaike info criterion	3.832855
Sum squared resid	478.2114	Schwarz criterion	3.984457
Log likelihood	-362.7870	Hannan-Quinn criter.	3.894243
F-statistic	7.308180	Durbin-Watson stat	2.325620
Prob(F-statistic)	0.000000		

- Ajusted R-squared = 21%
- Proba(F-stat) = 0,000000
- Only one insignificant variable !

What about classical assumptions ?

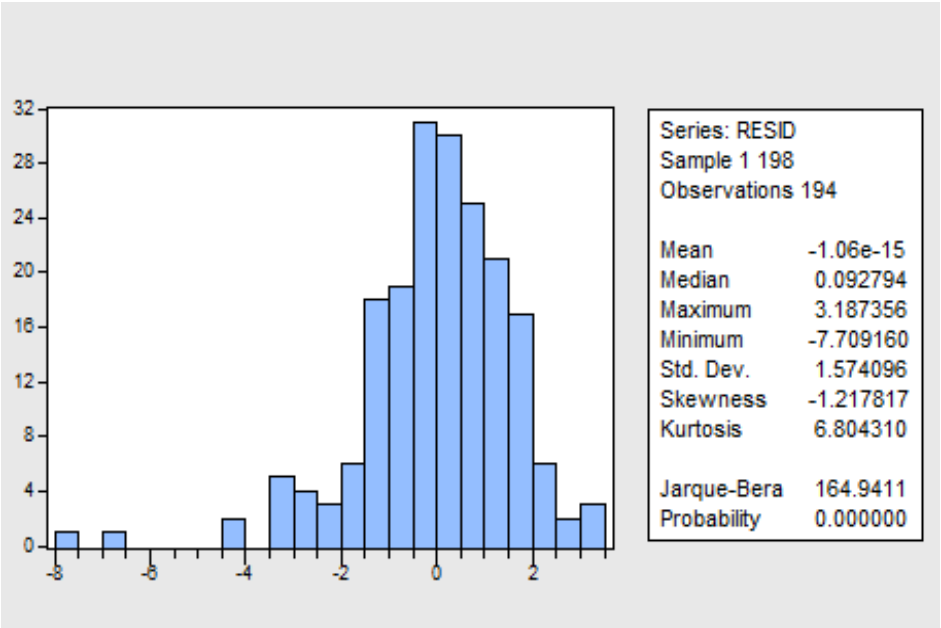


Figure: Graph of Residuals

- Mean $= 0$
- JBcorrected $= 156{,}4$
- We can NOT assume the normality of the errors

White's Test

Heteroskedasticity Test: White

F-statistic	1.073985	Prob. F(44,149)	0.3672
Obs*R-squared	46.71227	Prob. Chi-Square(44)	0.3616
Scaled explained SS	123.2797	Prob. Chi-Square(44)	0.0000

- H0 : Homoscedasticity
- H1 : Heteroscedasticity

Conclusion : Homoscedasticity

Figure: White's Test

Introduction
First Equation
Variables and marginal effect
Second Equation
What about Classical Assumptions ?
What about homoscedasticity ?
Conclusion

Conclusion

- Variables are jointly significant
- No heteroscedasticity
- Adjusted $R^2 \approx 21\%$:

Conclusion

Who is more likely to use online dating applications ?

- Male, single, unhappy and active on social networks

Does the use of social networks/dating apps make you happier ?

- Being in a relationship would make you happier
- On the contrary, the use of dating applications and the time spent on social networks could be a sign of dissatisfaction

Limits of our models

- Fairly homogeneous population
- Some significant variables might have been omitted