Size of airports in France

Cassandre Cheniez and Valentin Fina

ENAC

April 8, 2016

What can affect the size of an airport in France?

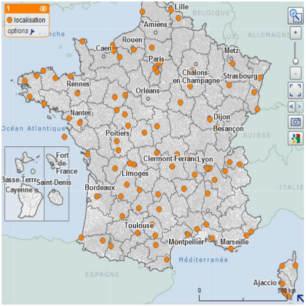

Figure: Maps of the airports in France in 2011.
Source: data.gouv.fr

There are about 160 airports in France, with different sizes such as Charles de Gaulle or Carcassonne.
What can affect the size of airports in France ?

How did we gather our data ?

Sources

Laboratoire d'Economie de l'ENAC, Albatros' database, Insee's database and data.gouv.fr

Dependent variables

Total passenger traffic, domestic passenger traffic, international passenger traffic, Low-Cost passenger traffic, total cargo traffic

Explanatory variables

City's size, city's inhabitants, distance to a major city, region's GDP, number of tourist attractions, high-speed rail link (TGV), airport operator

Database

Year	code airport OAG	name apt city country OAG	mvt	pax_dom	pax_int	pax_tot	fret	mgt	monuments_reg	monuments_dep	nuits	TGV	ville	distane_ville
2000	AJA	AJACCIO-CAMPO DELL'ORO, CORSICA, FRANCE	14 953	1 024 880	44 549	1 069 429	7 031	3	303			0	0	231,1
2005	AJA	AJACCIO-CAMPO DELL'ORO, CORSICA, FRANCE	13 096	949 588	26 205	975 793	6 423	3	303			0	0	231,1
2010	AJA	AJACCIO-CAMPO DELL'ORO, CORSICA, FRANCE	12 781	1 053 129	56 938	1 110 067	5 909	3	303	119	3	0	0	231,1
2013	AJA	AJACCIO-CAMPO DELL'ORO, CORSICA, FRANCE	14 520	1 243 130	102 464	1 345 594	5 988	3	303			0	0	231,1
2014	AJA	AJACCIO-CAMPO DELL'ORO, CORSICA, FRANCE	14 549	1 218 531	135 392	1 353 923	6 065	3	303			0	0	231,1
2015	AJA	AJACCIO-CAMPO DELL'ORO, CORSICA, FRANCE	15 260	1 224 481	131 849	1 356 330	7 268	3	303			0	0	231,1
2000	BIA	BASTIA-PORETTA, CORSICA, FRANCE	13 786	778 769	54 205	832 974	7 321	3	303			0	0	209,3
2005	BIA	BASTIA-PORETTA, CORSICA, FRANCE	12 991	743 977	76 408	820 385	8 093	3	303			0	0	209,3
2010	BIA	BASTIA-PORETTA, CORSICA, FRANCE	13 037	893 450	113 075	1 006 525	6 976	3	303	184	3	0	0	209,3
2013	BIA	BASTIA-PORETTA, CORSICA, FRANCE	13 343	991 058	133 971	1 125 029	8 048	3	303			0	0	209,3
2014	BIA	BASTIA-PORETTA, CORSICA, FRANCE	14 016	1 000 207	161 294	1 161 501	7 849	3	303			0	0	209,3
2015	BIA	BASTIA-PORETTA, CORSICA, FRANCE	14 043	1 022 874	167 769	1 190 643	7 447	3	303			0	0	209,3
2000	BVA	BEAUVAIS-TILLE, FRANCE	4 502	822	386 347	387 169	747	3	2966			0	0	69,4
2005	BVA	BEAUVAIS-TILLE, FRANCE	13 228	6	1 847 378	1 847 384	196	3	2966			0	0	69,4
2010	BVA	BEAUVAIS-TILLE, FRANCE	20 528	124 434	2 805 134	2 929 568	0	3	2966	637	5,5	0	0	69,4
2013	BVA	BEAUVAIS-TILLE, FRANCE	26 916	157 098	3 794 890	3 951 988	0	3	2966			0	0	69,4
2014	BVA	BEAUVAIS-TILLE, FRANCE	26 649	79 711	3 943 911	4 023 622	0	3	2966			0	0	69,4
2015	BVA	BEAUVAIS-TILLE, FRANCE	27 712	87 960	4 241 011	4 328 971	0	3	2966			0	0	69,4
2005	EGC	BERGERAC-ROUMANIERE, FRANCE	2 745	3 192	230 568	233 760	0	2	6083			0	0	86,4
2010	EGC	BERGERAC-ROUMANIERE, FRANCE	3 375	3 244	256 479	259 723	0	2	6083	884	8,4	0	0	86,4
2013	EGC	BERGERAC-ROUMANIERE, FRANCE	3 451	2 206	284 020	286 226	0	2	6083			0	0	86,4
2014	EGC	BERGERAC-ROUMANIERE, FRANCE	3 296	2 803	274 509	277 312	0	2	6083			0	0	86,4
2015	EGC	BERGERAC-ROUMANIERE, FRANCE	2 461	1 098	280 358	281 456	0	2	6083			0	0	86,4
2010	BZR	BEZIERS-VIAS, FRANCE	1 109	195	130 179	130 374	0	3	4630	569	7,4	0	0	53,1
2013	BZR	BEZIERS-VIAS, FRANCE	1 851	52 406	175 618	228 024	0	3	4630			0	0	53,1
2014	BZR	BEZIERS-VIAS, FRANCE	1 859	53 391	190 589	243 980	0	3	4630			0	0	53,1
2015	BZR	BEZIERS-VIAS, FRANCE	1 731	57 737	187 267	245 004	0	3	4630			0	0	53,1
2000	BIQ	BIARRITZ-BAYONNE/ANGLET, FRANCE	10 001	686 673	92 140	778 813	644	3	6083			0	0	170,
2005	BIQ	BIARRITZ-BAYONNE/ANGLET, FRANCE	9 959	652 556	163 843	816 399	107	3	6083			0	0	170,
2010	BIQ	BIARRITZ-BAYONNE/ANGLET, FRANCE	8 918	785 884	203 268	989 152	0	3	6083	388	8,4	0	0	170,
2013	BIQ	BIARRITZ-BAYONNE/ANGLET, FRANCE	9 547	859 777	237 399	1 097 176	0	3	6083			0	0	170,
2014	BIQ	BIARRITZ-BAYONNE/ANGLET, FRANCE	9 584	839 936	223 802	1 063 738	0	3	6083			0	0	170,
2015	BIQ	BIARRITZ-BAYONNE/ANGLET, FRANCE	9 742	812 590	226 756	1 039 346	0	3	6083			0	0	170,
2000	BOD	BORDEAUX-MERIGNAC, FRANCE	50 084	2 322 401	685 309	3 007 710	17 753	2	6083			1	1	10,3
2005	BOD	BORDEAUX-MERIGNAC, FRANCE	49 354	2 145 838	916 555	3 062 393	15 707	2	6083			1	1	10,3
2010	BOD	BORDEAUX-MERIGNAC, FRANCE	46 607	2 263 869	1 348 458	3 612 327	11 410	2	6083	975	8,4	1	1	10,3
2013	BOD	BORDEAUX-MERIGNAC, FRANCE	48 519	2 794 092	1 780 254	4 574 346	12 100	2	6083			1	1	10,3
2014	BOD	BORDEAUX-MERIGNAC, FRANCE	49 975	2 764 242	2 150 250	4 914 492	12 016	2	6083			1	1	10,3
2015	BOD	BORDEAUX-MERIGNAC, FRANCE	52 836	2 824 777	2 464 861	5 289 638	11 864	2	6083			1	1	10,3

Figure: Database used.

Sample size

42 airports, 6 years (2000, 2005,2010, 2013, 2014, 2015).

Expected effects

Explanatory variable	Expected effect
City (=0 or 1)	+
City's inhabitants	+
Region's GDP	+
Distance to the city	?
TGV	-
Number of tourist attractions	+
Airport operator	?

What do our data give when focusing only on the year 2010?

We are looking into a linear problem of the following form:

$$Y = \beta_1 POP\_VILLE + \beta_2 VILLE + \beta_3 PIB\_REG + \beta_4 DISTANCE\_VILLE \\ + \beta_5 TGV + \beta_6 (MGT = 1) + \beta_7 (MGT = 2) \\ + \beta_8 (MGT = 4) + \beta_9 (MGT = 5) + u$$

Results thanks to EViews

Equation: EQ_TOT Workfile: DATA_2010::Data_2010\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PAX_TOT
Method: Least Squares
Date: 04/05/16 Time: 11:10
Sample: 1 43
Included observations: 42

	Coefficient	Std. Error	t-Statistic	Prob.
C	-628678.2	4945229.	-0.127128	0.8997
DISTANCE_VILLE	3554.418	18426.68	0.192895	0.8484
VILLE	3339828.	2975391.	1.122484	0.2709
MONUMENTS_REG	-63.43750	536.9048	-0.118154	0.9068
MONUMENTS_DEP	241.7575	3812.971	0.063404	0.9499
NUITS	254629.9	279174.7	0.912081	0.3692
PIB_REG	10.87157	31.25903	0.347790	0.7305
POP_REGION	-0.614026	1.324479	-0.463598	0.6464
TGV	228907.8	2317966.	0.098754	0.9220
MGT=1	22257689	18233067	1.220732	0.2320
MGT=2	227803.7	2189697.	0.104034	0.9179
MGT=4	-1056725.	2505064.	-0.421836	0.6763
MGT=5	-1294647.	3718799.	-0.348136	0.7303

R-squared	0.839127	Mean dependent var	3357686.
Adjusted R-squared	0.772559	S.D. dependent var	9674015.
S.E. of regression	4613618.	Akaike info criterion	33.77560
Sum squared resid	6.17E+14	Schwarz criterion	34.31345
Log likelihood	-696.2875	Hannan-Quinn criter.	33.97274
F-statistic	12.60551	Durbin-Watson stat	3.007259
Prob(F-statistic)	0.000000		

Conclusion

None of our variables is significant.

What do our data give when focusing only on the year 2010?

We are now looking into a linear problem with variables non linearly entered:

$$Y = \beta_1 \log(POP\_VILLE) + \beta_2 VILLE + \beta_3 \log(PIB\_REG) + \beta_4 \log(DISTANCE\_VILLE) + \beta_5 TGV + \beta_6(MGT = 1) + \beta_7(MGT = 2) + \beta_8(MGT = 4) + \beta_9(MGT = 5) + u$$

Results thanks to EViews

Equation: LOG_TOT Workfile: DATA_2010::Data_2010\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PAX_TOT)
Method: Least Squares
Date: 04/06/16 Time: 08:31
Sample: 1 43
Included observations: 42

	Coefficient	Std. Error	t-Statistic	Prob.
C	14.22075	4.432292	3.208441	0.0030
LOG(DISTANCE_VILLE)	-0.216534	0.317244	-0.682547	0.4998
VILLE	1.599334	0.858476	1.862992	0.0717
LOG(MONUMENTS_REG)	-0.301130	0.304930	-0.987537	0.3308
LOG(MONUMENTS_DEP)	-0.513651	0.401251	-1.280123	0.2097
LOG(POP_REGION)	0.358125	0.326446	1.097042	0.2808
MGT=1	2.844002	1.109702	2.562851	0.0153
MGT=2	0.019762	0.478253	0.041320	0.9673
MGT=4	-0.711651	0.546759	-1.301581	0.2024
MGT=5	-0.905113	0.753012	-1.201991	0.2382

R-squared	0.705424	Mean dependent var	13.36189
Adjusted R-squared	0.622574	S.D. dependent var	1.648424
S.E. of regression	1.012709	Akaike info criterion	3.067392
Sum squared resid	32.81854	Schwarz criterion	3.481122
Log likelihood	-54.41522	Hannan-Quinn criter.	3.219040
F-statistic	8.514518	Durbin-Watson stat	1.790454
Prob(F-statistic)	0.000002		

Conclusion

Lot of insignificant variables : only two variables are significant (VILLE and (MGT=1)).

Model taking data from each year into account

Model (variables entered non linearly)

$$Y = \beta_1 log(POP\_VILLE) + \beta_2 VILLE + \beta_3 log(PIB\_REG)$$
$$+\beta_4 log(DISTANCE\_VILLE) + \beta_5 log(MONUMENTS\_REG)$$
$$+\beta_6 TGV + \beta_7(MGT = 1) + \beta_8(MGT = 2)$$
$$+\beta_9(MGT = 4) + \beta_{10}(MGT = 5) + u$$

Results thanks to EViews

Equation: LOG_TOT Workfile: APT_FRANCE00_15_REGION-1::Apt_france00_15_re...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PAX_TOT)
Method: Least Squares
Date: 04/06/16 Time: 09:25
Sample (adjusted): 1 251
Included observations: 118 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.59837	1.703963	7.980435	0.0000
TGV	-0.032496	0.250791	-0.129575	0.8971
VILLE	0.812102	0.447356	1.815337	0.0723
LOG(DISTANCE_VILLE)	-0.306837	0.153585	-1.997828	0.0483
LOG(MONUMENTS_REG)	-0.588445	0.142139	-4.139931	0.0001
LOG(POP_VILLE)	0.373419	0.096129	3.884561	0.0002
LOG(PIB_REG)	0.125557	0.137278	0.914617	0.3624
MGT=1	1.746085	0.581068	3.004955	0.0033
MGT=2	-0.063406	0.241908	-0.262109	0.7937
MGT=4	-0.551257	0.260377	-2.117149	0.0366
MGT=5	-1.104908	0.422884	-2.612794	0.0103

R-squared	0.744920	Mean dependent var	13.42777
Adjusted R-squared	0.721081	S.D. dependent var	1.613946
S.E. of regression	0.852370	Akaike info criterion	2.606994
Sum squared resid	77.73927	Schwarz criterion	2.865278
Log likelihood	-142.8126	Hannan-Quinn criter.	2.711865
F-statistic	31.24759	Durbin-Watson stat	0.263705
Prob(F-statistic)	0.000000		

Figure: Screenshot of the results given by EViews.

Marginal effects

Signs

Explanatory variable	Expected effect	Observed effect
City (=0 or 1)	+	+
City's inhabitants	+	+
Region's GDP	+	+
City's distance	?	-
TGV	-	-
Number of tourist attractions	+	-
Airport operator	?	+/-

Case of Airport operator

The marginal effects are observed comparing airports operated by CCI to other forms of management.

Validity of the first model

Fitness

$$R^2 = 72,1\%$$

Wald's test

We have to test if we can reject the hypothesis :

$$H_0 : \beta_i = 0, i = 1, 2, \ldots, 8$$

with a level of certainty of 90%.

Wald's tests

Variable	t-Statistic	Prob.
C	7,9804	0
TGV	-0,1295	0,8971
VILLE	1,8115	0,0723
LOG(DISTANCE_VILLE)	-1,9978	0,0483
LOG(MONUMENTS_REG)	-4,1399	0,00001
LOG(POP_VILLE)	3,8845	0,002
LOG(PIB_REG)	0,9146	0,3624
MGT=1	3,0049	0,0033
MGT=2	-0,2621	0,7937
MGT=4	-2,1171	0,0366
MGT=5	-2,96127	0,0103

Figure: Results of the Wald's tests

We can drop the variables with the highest probability (do not reject the null hypothesis) :

- TGV
- LOG(PIB_REG)
- (MGT=2)

Conclusion on the validity of the model

Possible improvements

We can drop the variables TGV and PIB_REG.

Equation: LOG_TOT_RESTRICTED Workfile: APT_FRANCE00_...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PAX_TOT)
Method: Least Squares
Date: 04/06/16 Time: 14:33
Sample (adjusted): 1 251
Included observations: 118 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	14.49416	1.412155	10.26386	0.0000
VILLE	0.784899	0.440558	1.781602	0.0776
LOG(DISTANCE_VILLE)	-0.338571	0.149137	-2.270198	0.0252
LOG(MONUMENTS_REG)	-0.516634	0.116172	-4.447156	0.0000
LOG(POP_VILLE)	0.375269	0.094273	3.980658	0.0001
MGT=1	1.963181	0.487333	4.028416	0.0001
MGT=2	-0.057874	0.236894	-0.244303	0.8075
MGT=4	-0.493372	0.251068	-1.965091	0.0519
MGT=5	-1.041208	0.364292	-2.858169	0.0051

R-squared	0.742767	Mean dependent var	13.42777
Adjusted R-squared	0.723888	S.D. dependent var	1.613946
S.E. of regression	0.848070	Akaike info criterion	2.581499
Sum squared resid	78.39528	Schwarz criterion	2.792822
Log likelihood	-143.3084	Hannan-Quinn criter.	2.667302
F-statistic	39.34260	Durbin-Watson stat	0.263571
Prob(F-statistic)	0.000000		

Figure: Screenshot of the results given by EViews.

Improvements

Fitness

$$R^2 = 72.4\% \ \geq R^2_{firstmodel}$$

Insignificant variables

We only have (MGT=2)(operated by airports' companies) as an insignificant variable

Correlation

Correlation				
	DISTANCE_...	MONUMENT...	POP_VILLE	VILLE
DISTANCE_...	1.000000	-0.308120	-0.439889	-0.714227
MONUMENT...	-0.308120	1.000000	0.062244	0.032905
POP_VILLE	-0.439889	0.062244	1.000000	0.576017
VILLE	-0.714227	0.032905	0.576017	1.000000

Figure: Screenshot of the correlation matrix.

Problem of the variable DISTANCE_VILLE

Correlated with MONUMENTS_DEP,VILLE and POP_VILLE, but needed in order to observe the effect of the proximity of a major city.

Classical assumptions (1/4)

Linearity

The model fits the data (fitted-R squared) and the model is meaningful (global F-test).

Classical assumptions (2/4)

Strict exogeneity

$$E(u) = 0$$

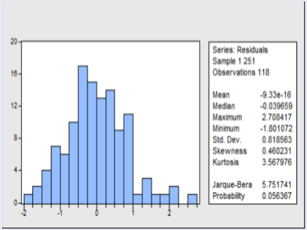

Figure: Histogram of residuals given by EViews.

We have :

$$E(u) \approx 0$$

The classical assumption is held.

Classical assumptions(3/4)

Homoscedasticity

Heteroskedasticity Test: White

F-statistic	1.465926	Prob. F(29,88)	0.0891
Obs*R-squared	38.43635	Prob. Chi-Square(29)	0.1129
Scaled explained SS	42.11066	Prob. Chi-Square(29)	0.0549

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/06/16 Time: 14:36
Sample: 1 251
Included observations: 118
Collinear test regressors dropped from specification

Figure: White's test with EViews

We do not reject the null hypothesis (no heteroscedasticity) at a level of significance of 95 %.

Homoscedasticity of the model.

Classical assumptions(4/4)

Normality of the residuals (Jarque-Bera test)

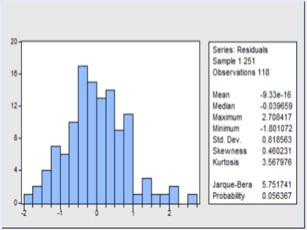

Figure: Histogram of residuals given by EViews.

$$Prob = 5.6\%$$

We do not reject the null hypothesis of normality at a level of significance of 95%.

Strengths and weaknesses

Strengths

- Good model which fits the data while respecting classical assumptions
- Almost all marginal effects correspond to intuition

Weaknesses

- Some variables collected do not seem relevant (TGV and PIB_REG)
- Problem of the sign of the marginal effect MONUMENTS_REG

Explanations

- Different politics of airports' companies
- For the number of tourist attractions, categorization not precise enough
- A lot of other factors which are difficult to estimate have to be taken into account

Comparison between the first and the logarithm model

Variable	t-Statistic	Prob.
C	-0,3458	0,7302
TGV	-0,5264	0,5997
VILLE	0,8349	0,4056
DISTANCE_VILLE	0,3554	0,723
MONUMENTS_REG	-0,5033	0,6158
POP_VILLE	2,075	0,0404
PIB_REG	1,558	0,1223
MGT=1	2,3382	0,0212
MGT=2	0,2967	0,7673
MGT=4	-0,9205	0,3594
MGT=5	-0,1827	0,8554

Figure: Results for the first model

Variable	t-Statistic	Prob.
C	7,9804	0
TGV	-0,1295	0,8971
VILLE	1,8115	0,0723
LOG(DISTANCE_VILLE)	-1,9978	0,0483
LOG(MONUMENTS_REG)	-4,1399	0,00001
LOG(POP_VILLE)	3,8845	0,002
LOG(PIB_REG)	0,9146	0,3624
MGT=1	3,0049	0,0033
MGT=2	-0,2621	0,7937
MGT=4	-2,1171	0,0366
MGT=5	-2,96127	0,0103

Figure: Results for the log model

Conclusion

With the exception of TGV, MONUMENTS_REG and MGT=2, we have the same signs for the marginal effect.
The logarithm model presents fewer insignificant variables.