Analysis of Tourism Demand in the Basque Country with Daily Data.

Ignacio Díaz-Emparanza and Paz Moral,

Departamento de Economía Aplicada III (Econometría y Estadística)
Universidad del País Vasco - Euskal Herriko Unibertsitatea

June 14, 2015

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

- This is a work in progress, in a very preliminary stage.
- A team in our department is analyzing the tourism demand in the Basque Country.
- They have several daily series from Google Analytics (https://www.google.com/analytics/) which we will analyze. Some of them are web site traffic from particular firms, or from an Small to Medium-sized Enterprise Association.

Universidad del País Vasco Euskal Herriko Unibertsitatea

Universidad del País Vasco Euskal Herriko Unibertsitatea

Some 2012 statistics:

- About 60% of the tourism to the Basque Country come from the rest of Spain.
- In particular, the number of tourists who come to Bizkaia may be divided in three equal parts: 33% from Gipuzcoa-Araba, 33% from the rest of Spain 33% from abroad.
- A 20% of the tourists who come to Bizkaia go to the coast.

We present here the study of the data of an active sports tourism firm. XXXXX, was born like ‘Active Tourism and Adventure Sports’ company in 1999. The firm’s website pursues the dissemination of information in the field of the company’s products:

BIG SUP or KING SUP (Stand Up Paddel)

Universidad del País Vasco Euskal Herriko Unibertsitatea

CANOEING

BODYBOARD

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

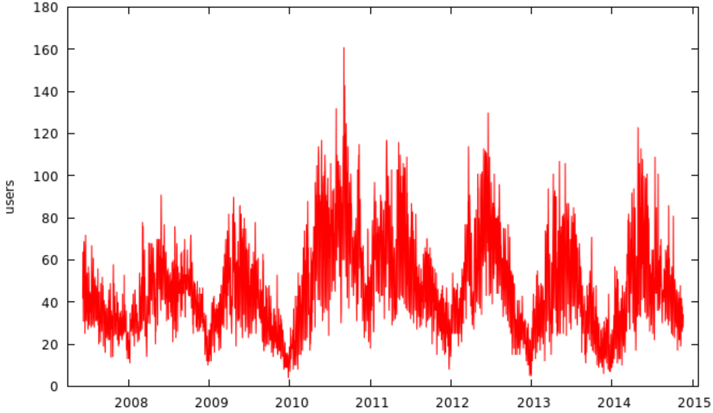
CLIMBING

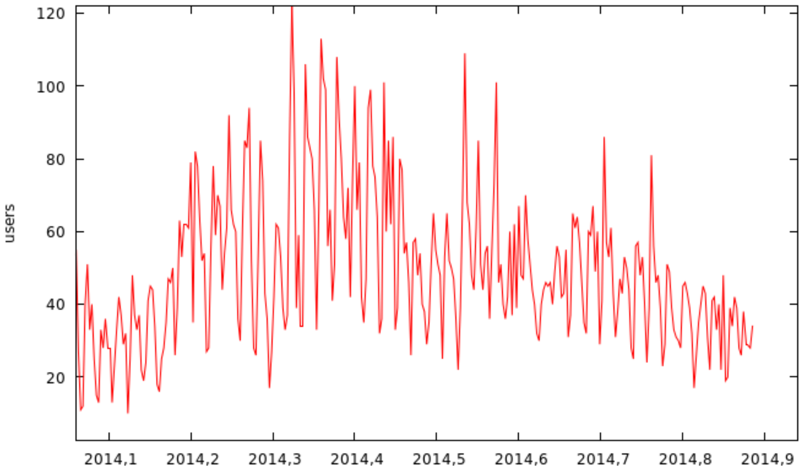
COASTEERING

… etc.

...and they also maintain a Hostel which is associated to Hostelling International (with 55 beds).

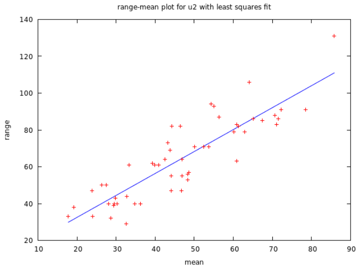
In June 2007 the Webmaster started to analyse web traffic using Google Analytics. We analyze the series of **number of visitors** to the web page from june 6, 2007 to november 20, 2014 (2725 obs.).





Universidad del País Vasco Euskal Herriko Unibertsitatea

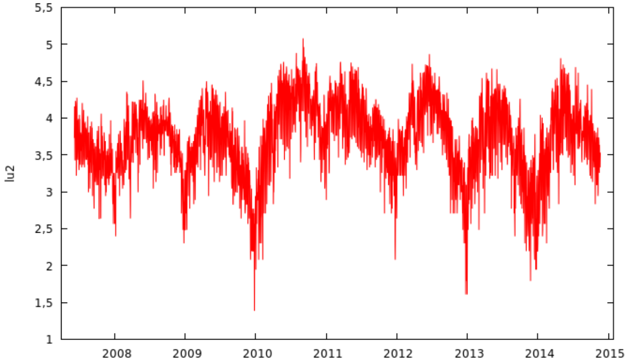
- The series is not affected by advertising campaigns (so as for example Google AdWords).
- The server did not suffer of web attacks, but there were two observations with zeroes, probably because of problems with the server or maintenance shutdown. This observations were substituted with those of the previous week.
- At first glance we may observe: heteroskedasticity (high variance for periods of high mean), annual cycles and weekly cycles.



Universidad del País Vasco Euskal Herriko Unibertsitatea

Stationarity Analysis

We start by taking logs having then the series



Stationarity Analysis

Now we test for unit roots. The seasonal periodicity defined in gretl is S=7. Using HEGY-GLS tests we obtain:

Statistic	Period
t1= -3.50 **	∞
F1= 81.75 **	7
F2= 101.31 **	3.50
F3= 135.49 **	2.33

(Deterministic components removed are det. seasonals + linear trend).
AR order = 14 (determined by MAIC with max.order=14)
Dof (T-k) = 2683
5% asymptotic critical values for t1=-2,87, t2=-1.97
and individual Fk=3.15 (Rodrigues and Taylor, 2007)

Universidad del País Vasco Euskal Herriko Unibertsitatea

The previous results are incorrect, because the annual cycle has not been considered. We tried to apply the HEGY-GLS statistics for S=365 but we had rare results, because of precission problems in the multiplication of polynomials. For even seasonality:(S=4)

$$(1-\gamma_0 L)(1+\gamma_\pi L)(1-2\gamma_k\cos(\omega_k)L+\gamma_k L^2)X_t$$

For odd seasonality:(S=3)

$$(1-\gamma_0 L)(1-2\gamma_k\cos(\omega_k)L+\gamma_k L^2)X_t$$

Where the γ values depend on the type of dterministic part we want to remove. For example for const+seasonals+trend we have$\gamma_0 = (1-13.5/T)$, $\gamma_\pi = (1-7/T)$, $\gamma_k = (1-3.75/T)$

Universidad del País Vasco Euskal Herriko Unibertsitatea

For S=12 this implies

$$(1-\gamma_0 L)(1+\gamma_\pi L)\prod_{i=1}^{5}(1-2\gamma_k\cos(\omega_i)L+\gamma_k L^2)X_t$$

where ω_i are the seasonal frequencies $\omega_i = (2\pi i)/S$ and $i = 1, \ldots, [S-1/2]$. For $S = 365$ the polynomials to multiply are

$$(1-\gamma_0 L)\prod_{i=1}^{182}(1-2\gamma_k\cos(\omega_i)L+\gamma_k L^2)X_t$$

and gretl (by means of the fft) loses too much precission with so many multiplications (until S=52 the precission is acceptable)

Universidad del País Vasco Euskal Herriko Unibertsitatea

Changing the seasonal periodicity to S=365. HEGY test of seasonal unit roots for series lu2:(GHegy gretl package)

Statistic	p-value	Period
t1= -2.85	0.02121 **	∞
F1= 3.23	0.09607	365
F2= 6.62	0.00118 ***	182.50
...	...	...
F11= 7.45	0.00039 ***	33.18
F12= 4.48	0.02118 **	30.42
F13= 7.82	0.00025 ***	28.08
...	...	...
F52= 11.04	0.00000 ***	7.02
...	...	...

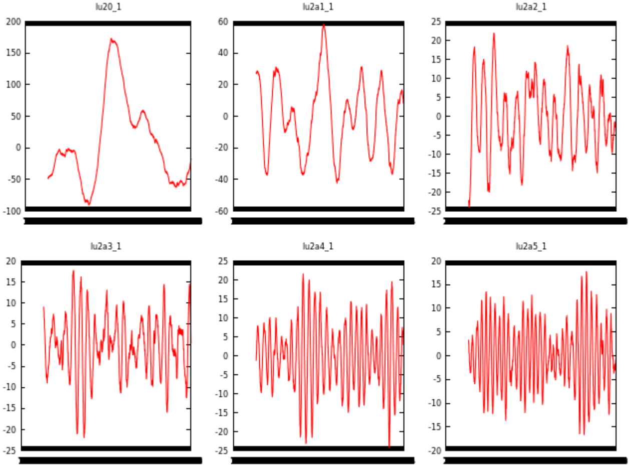
AR order = 1 (determined by HQC with max.order=365)
Deterministic comp: const + (S-1) trigon. terms + trend
Dof (T-k) = 1627

Universidad del País Vasco Euskal Herriko Unibertsitatea

HEGY with incomplete GLS de-trending for S=365. In fact this uses the cycles in S=7 plus monthly and annual cycles (periods 30.43 and 365.2564)

Statistic	Period
t1= -1.91	∞
F1= 0.30	365
F2= 0.70	182.50
...	...
F11= 2.52	33.18
F12= 2.09	30.42
F13= 4.03 **	28.08
...	...
F52= 6,07 **	7,02
F104= 7,93 **	3,51
F156= 12,47 **	2,34
...	...

AR order = 137 (determined by MAIC with max.order=365)
Deterministic: const. Dof (T-k) = 1721
5% asymptotic critical values for t1=-2,87, t2=−1.97
and individual Fk=3.15 (Rodrigues and Taylor, 2007)



Universidad del País Vasco
Euskal Herriko Unibertsitatea

We do again the test with S=7 but allowing for AR order=365. Using HEGY-GLS tests we obtain:

Statistic	Period
t1= -1,83	∞
F1= 3,11	7
F2= 4,22 **	3,50
F3= 9,23 **	2,33

(Deterministic components removed are det. seasonals + linear trend).
AR order = 365 (determined by MAIC with max.order=365)
Dof (T-k) = 1981
5% asymptotic critical values for t1=-2,87, t2=−1.97
and individual Fk=3.15 (Rodrigues and Taylor, 2007)

Canova-Hansen tests of seasonal stability for Δlu2.
Regressors in the auxiliary regression: Trigonometric terms + linear trend.
Degrees of freedom (T-k) = 2716, lag order = 15.

	Statistic	p-value		Ang. Frequency	Period
$L_1 =$	1,8145	0,00018	***	$\pm 2\pi/7$	7,00
$L_2 =$	1,6352	0,00047	***	$\pm 4\pi/7$	3,50
$L_3 =$	0,1990	0,71867		$\pm 6\pi/7$	2,33
$L_f =$	2,8087	0,00029	***	Joint test	

Universidad del País Vasco Euskal Herriko Unibertsitatea

Canova-Hansen tests of seasonal stability for Δlu2.
Regressors in the auxiliary regression: Trigonometric terms + linear trend.
Degrees of freedom (T-k) = 2358, lag order = 54.

	Statistic	p-value	Period
L_1 =	0,1074	0,97709	365,00
L_2 =	0,0582	0,99962	182,50
L_3 =	0,0505	0,99989	121,67
...	...	...	
L_{12} =	0,5843	0,15868	30,42
L_{13} =	0,1940	0,80207	28,08
...	...	...	
L_{52} =	4,0453	0,00000 ***	7,02
L_{104} =	0,9428	0,03145 **	3,51
L_{157} =	0,9270	0,03380 **	2,32
...	...	...	

eman ta zabal zazu
Universidad del País Vasco Euskal Herriko Unibertsitatea

KPSS test for lu2 (including trend and seasonals) (filtering by unattended unit roots with S=7 we obtain a very similar result)
T = 2725
Lag truncation parameter = 54
Test statistic = 0.17373 (Interpolated p-value 0.035)

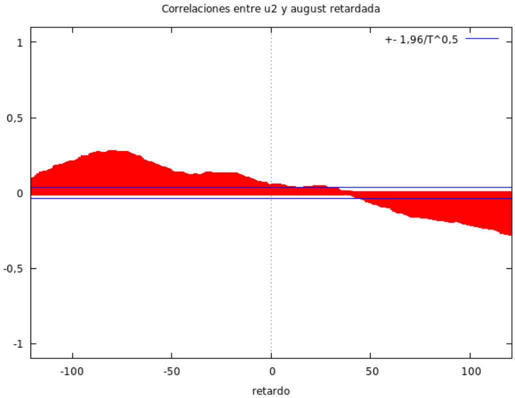
	10%	5%	1%
Critical values:	0.119	0.148	0.218

Universidad del País Vasco Euskal Herriko Unibertsitatea

Conclusions:

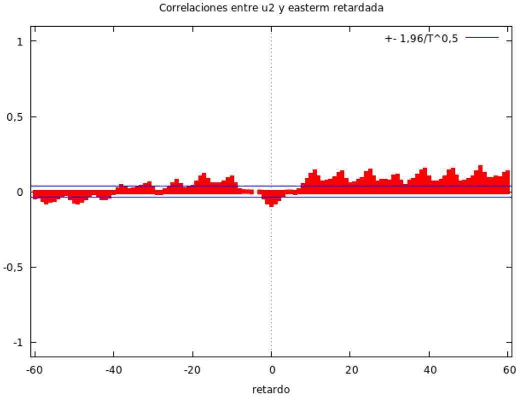
- we have unit roots at period 7 and the long run.
- Both tests do not reject the null for the annual cycle (period 365). This is probably because we only observe 7.5 cycles and this does not imply much information about the structure of this cycle. With this data this may be modelled as a deterministic (+ a possible transitory) cycle.
- The same occurs for the monthly cycle although it seems to be more support for the deterministic cycle.
- Results for cycles of period 3.5 and 2.33 are contradictory. Both tests reject the null.

Including some explanatory variables. August dummy, Easter dummy(we should look at the left part of the graph)



It seems potential visitors consult their server between 2–3 months before the month of august.

But they only consult this server between 1–3 weeks before easter holidays.



This graphs should be considered with caution, because given that we have unit roots we don't have the assumptions of the ergodic theorems so the sample correlations may be inconsistent estimators.

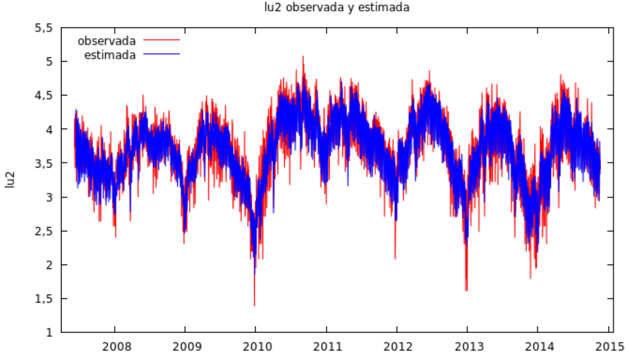
Modelo 1: ARMAX, usando las observaciones 2007-06-13–2014-11-20 (T = 2718)
Variable dependiente: $(1 - L^s)$lu2
Desviaciones típicas basadas en el Hessiano

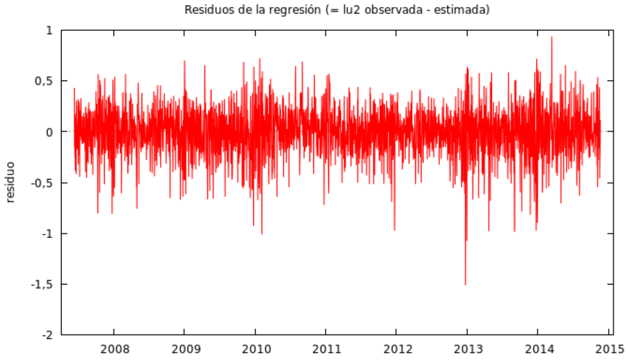
	Coeficiente	Desv. Típica	z	Valor p
const	0,000385282	0,00107124	0,3597	0,7191
ϕ_1	0,962502	0,00830886	115,8405	0,0000
Φ_1	0,112895	0,0204520	5,5200	0,0000
θ_1	−0,709909	0,0227044	−31,2675	0,0000
Θ_1	−0,975796	0,00720615	−135,4116	0,0000
easterm	−0,391447	0,0488828	−8,0079	0,0000
sinpd2	0,0455374	0,0506135	0,8997	0,3683
cospd2	0,410642	0,0510342	8,0464	0,0000
sinpd3	0,0287433	0,0105494	2,7246	0,0064
cospd3	−0,0153349	0,0104535	−1,4670	0,1424

Media de la vble. dep.	−0,001205	D.T. de la vble. dep.	0,327207
media innovaciones	0,000926	D.T. innovaciones	0,237506
Log-verosimilitud	41,44384	Criterio de Akaike	−60,88768
Criterio de Schwarz	4,096484	Hannan–Quinn	−37,39541

			Real	Imaginaria	Módulo	Frecuencia
AR						
	Raíz	1	1,0390	0,0000	1,0390	0,0000
AR (estacional)						
	Raíz	1	8,8578	0,0000	8,8578	0,0000
MA						
	Raíz	1	1,4086	0,0000	1,4086	0,0000
MA (estacional)						
	Raíz	1	1,0248	0,0000	1,0248	0,0000

Universidad del País Vasco Euskal Herriko Unibertsitatea


lu2 observada y estimada
observada
estimada
lu2
5,5
5
4,5
4
3,5
3
2,5
2
1,5
1
2008
2009
2010
2011
2012
2013
2014
2015


Residuos de la regresión (= lu2 observada - estimada)
1
0,5
0
-0,5
-1
-1,5
-2
residuo
2008
2009
2010
2011
2012
2013
2014
2015