



ANÁLISE E PREVISÃO DA PROCURA TURÍSTICA NA REGIÃO NORTE DE PORTUGAL

Natália dos Santos (nspink@hotmail.com); Paula Odete Fernandes (pof@ipb.pt).

Instituto Politécnico de Bragança
Escola Superior de Tecnologia e Gestão
Campus de Santa Apolónia, Apartado 1038
5301-854 Bragança - Portugal.
Telefone: +351 273 303103
Fax: +351 273 325405

RESUMO

A região Norte de Portugal, nos últimos anos, tem apresentado relevantes potencialidades de desenvolvimento turístico. Assim, no sentido de criar condições favoráveis aos turistas que se deslocam à Região Norte de Portugal, será de todo proveitoso construir mecanismos que permitam obter uma visão antecipada da evolução da procura turística.

Neste contexto surgiu o presente estudo centrando-se o seu principal objectivo na construção de modelos econométricos, assentes no modelo linear geral estocástico, estático e dinâmico, que permitissem explicar o comportamento da procura turística na Região Norte de Portugal. Para tal, utilizou-se a série temporal de turismo: “Dormidas Mensais, nos estabelecimentos hoteleiros, na região Norte de Portugal”, registadas no período de Janeiro de 1996 a Dezembro 2008. Os resultados obtidos revelaram que o modelo estático foi o que produziu os resultados mais satisfatórios, garantindo as hipóteses básicas do modelo linear geral, evidenciando, ainda, ser adequado para explicar o comportamento da procura turística, na região Norte de Portugal.

Palavras-Chave: Turismo, Procura Turística, Modelos Econométricos, Modelo Linear Geral.

1. Introdução

O turismo é considerado um fenómeno de singular importância na vivência estrutural da sociedade (Cunha, 2003). Pois, desde as remotas viagens do homem pré-histórico, até finais do século XVIII, quando tiveram início os primeiros fluxos turísticos, continuando até aos dias actuais, este fenómeno evoluiu e adquiriu cada vez mais uma significativa importância socioeconómica quer a nível regional, nacional ou mundial.



Neste sentido, o turismo tem um interesse verdadeiramente estratégico para a economia portuguesa em virtude da sua capacidade em criar riqueza e emprego. Trata-se de um sector em que se evidenciam vantagens competitivas claras como sucede com poucos outros. Em termos turísticos, Portugal pode caracterizar-se como um país de sol, mar e de praias. Contudo sendo um país muito diversificado, colocando à disposição de quem o visita uma grande variedade de produtos turísticos existe a necessidade de reunir esforços económicos para o aproveitamento destes atractivos e dinamizar este potencial bruto.

À semelhança de Portugal também a região Norte de Portugal se pauta por ser uma região muito diferenciada que oferece uma alternativa interessante ao chamado ‘turismo de massas’, apostando na oferta de uma grande multiplicidade de produtos turísticos, que vão desde a praia, às montanhas, passando pelas estâncias termais não esquecendo o turismo rural que teve um acréscimo significativo nos últimos anos.

De notar que os principais mercados emissores de turistas que se deslocam à região Norte de Portugal são a Espanha, a Alemanha, a França, a Holanda e o Reino Unido. A Espanha é responsável por quase metade dos turistas estrangeiros e no seu conjunto estes cinco países são responsáveis por mais de quatro quintos dessas entradas. Não se deve esquecer que o mercado doméstico, turistas portugueses, é o principal mercado emissor da região Norte de Portugal, contribuindo, em 2008, com cerca de 60% (INE, 2009).

Assim, e no sentido de compreender o comportamento e as variações da procura turística na Região Norte de Portugal pretende-se com este trabalho construir um modelo econométrico eficaz, assente no modelo linear geral estocástico e dinâmico. Neste sentido, vai utilizar-se a série temporal do turismo: “Dormidas Mensais, nos estabelecimentos hoteleiros, na região Norte de Portugal”, registadas no período de Janeiro de 1996 a Dezembro de 2008, como sendo a variável dependente e para a sua explicação pretende-se utilizar, como variáveis explicativas, as seguintes variáveis: permanência média, índice de preços no consumidor de Espanha, o número de desempregados de Espanha e o número de desempregados de Portugal.

O artigo encontra-se estruturado da seguinte forma: no ponto dois apresenta-se uma breve descrição dos modelos econométricos aplicados, seguidamente no ponto três



descreve-se as variáveis a incluir no modelo. A explicação dos dados e aplicação dos modelos econométricos é apresentada no ponto quatro. A análise e discussão dos resultados apresenta-se no ponto cinco e no ponto seis apresentam-se as principais conclusões desta investigação.

2. Modelo Linear Geral

Modelar e prever a procura turística torna-se de fundamental importância para o planeamento turístico, recorrendo-se a diferentes métodos e instrumentos de modelação e previsão que permitam melhorar o rigor da previsão da procura turística (Preez & Witt, 2003). Vários são os estudos científicos publicados tendo por base a modelação e previsão da procura turística (Witt & Witt, 1995; Lim, 1997; Thomakos & Guerard, 2004). O crescente interesse desta área de estudo tem-se relacionado com o rápido desenvolvimento do turismo Internacional e das economias de um país (Frechtling, 2009). Contudo existe um largo número de factores de índole sócio-culturais, económicos, políticos e tecnológicos, que podem influenciar a procura turística negativa ou positivamente. Planear debaixo destas circunstâncias torna-se singularmente difícil mas importante. Vários são os autores que têm vindo a contribuir para o aparecer de diferentes metodologias, robustas, de modelação e previsão, utilizando diferentes abordagens para solucionar diferentes problemas associados aos modelos, desde os mais simples aos mais complexos (Makridakis & Hibon, 1997; Goh & Law, 2002; Thawornwong & Enke, 2004; Yu & Schwartz, 2006; Moutinho *et al.* 2008; Fernandes *et al.* 2008; Guizzardi & Mazzocchi, 2010; Kairat, 2010).

Neste sentido, surgiu a ideia de contribuir para o enriquecimento da modelação da procura turística para a Região Norte de Portugal, utilizando um modelo linear geral.

“O modelo linear pela sua simplicidade, tanto de tratamento matemático como de interpretação, é usual privilegiar a utilização de modelos lineares, ou linearizáveis, nos parâmetros.” Segundo Chaves *et al.* (2000:100).

O modelo de regressão linear simples (MRLS) define-se pelo seguinte (Johnston & Dinardo, 2000):

$$Y_i = b_1 + b_2 X_i + u_i \quad (1)$$



Estabelece, para cada uma observação i , uma relação linear de dependência – suportada por dois parâmetros b_1 e b_2 , designados de coeficientes de regressão – entre uma variável explicada observável Y , uma variável explicativa X , e um termo de erro U .

A variável dependente (Y) depende de um conjunto de $n+1$ factores conhecidos e um factor desconhecido, sendo uma variável endógena, explicada, estocástica ou aleatória e observável. Esta variável depende das variáveis independentes (X), um conjunto de $n+1$ variáveis, sendo variáveis exógenas, explicativas, não estocásticas e observáveis. Admite-se ainda: a independência das variáveis explicativas $\text{cov}(X_i, X_j) = 0$, hipótese da não-multicolinearidade; e a independência das variáveis explicativas e o termo de erro $\text{cov}(X_i, u) = 0$, hipótese da exogeneidade.

Na estimação dos parâmetros b_1 e b_2 , adopta-se o método dos mínimos quadrados (MQ), que se fundamenta no critério da minimização da soma dos quadrados dos resíduos. São valores constantes que vão ser estimados e os seus valores estimados medem o grau da influência marginal que as variáveis explicativas exercem sobre a variável dependente. O termo de erro ou termo estocástico (u) é uma variável aleatória ou estocástico, uma variável não observável dado que não existem dados que possam permitir a sua representação no modelo.

No sentido de garantir a convergência estatística o estimador dos mínimos quadrados deve satisfazer algumas propriedades estatísticas, tais como: a propriedade de não enviesamento ($E(\hat{b}) = b$), de eficiência (de entre os estimadores não enviesados têm variância mínima) e de consistência ($P \lim_{n \rightarrow \infty}(\hat{b}) = b$). Os estimadores OLS¹⁹ devem ser BLUE²⁰, ou seja, o melhor estimador linear não enviesado (Gujarati, 1995).

2.1. Hipóteses básicas do MLG

O método de estimação dos modelos econométricos, as inferências e as previsões só são válidos quando estas hipóteses são garantidas.

¹⁹ Do inglês *Ordinary Least Squares*.

²⁰ Do inglês *Best Linear Unbiased Estimators*.



Para que se possa estimar os parâmetros do MLG é necessário considerar algumas hipóteses relevantes. As hipóteses básicas consistem na linearidade do MLG, ou seja, o modelo de regressão múltipla é linear pelo menos nos parâmetros como se pode observar na [eq. (1)], devendo garantir a ausência do enviesamento ou erros de especificação (Zhihua & Qihua, 2009).

Em relação à estocasticidade da variável dependente, os seus valores são fixos em amostras repetidas e são variáveis, ou seja, os valores das variáveis independentes não são iguais na mesma amostra, devendo existir uma independência das variáveis explicativas com o termo de erro, também conhecida por ortocolinearidade.

2.1.1. Hipóteses básicas sobre as variáveis explicativas

As hipóteses básicas do MLG mais significativas fazem-se sobre as variáveis explicativas e o termo de erro. Geralmente, a estimação dos modelos e as inferências estatísticas são válidas, só quando as hipóteses básicas são garantidas.

As variáveis explicativas são independentes entre si: $\text{cov}(X_i, X_j) = 0$, o valor de uma variável independente não influencia outra variável independente, ou seja, ausência de multicolinearidade.

A convergência assintótica é o número de observações T superior ao número de parâmetros do modelo, para garantir suficientes graus de liberdade nos testes de hipótese. Esta hipótese é admitida para assegurar a propriedade estatística dos estimadores da consistência e convergência.

2.1.2. Hipóteses básicas sobre o termo de erro

O comportamento do termo de erro do MLG descreve-se estatisticamente através das seguintes hipóteses (Gujarati, 1995):

- Hipótese da normalidade do termo de erro: o termo de erro segue uma lei normal com média e variância conhecida, $u \sim N(0, \delta^2 I_T)$, $\text{cov}(u_i, u_j) = 0$ com $i \neq j$;
- Hipótese das médias zero do termo de erro: implica que o efeito médio dos erros sobre Y é nulo, ou seja, $E(Y) = E(Xb + u) = Xb + E(u) = Xb$;



- Hipótese da variância constante de termo de erro: a variância do termo de erro é a mesma para todas as observações (δ^2). Hipótese da homoscedasticidade, ou seja, a variância é constante de observação para observação. Quando a variância não é constante para todas as observações, o modelo é heteroscedástico, $\text{var}(u_t) = (\delta^2)$, com t diferente de observação para observação;

- Hipótese da independência dos erros (não autocorrelação dos erros). Os erros do MLG são independentes de período para período, ou seja, os erros são independentes não correlacionados entre si $\text{cov}(u_i, u_j) = 0$ com $i = j$. Quando a hipótese é infringida diz-se que tem autocorrelação dos erros $\text{cov}(u_i, u_j) \neq 0$ com $i \neq j$.

2.2. Modelo Estático

O modelo restrito não é mais do que uma representação das relações entre as variáveis no mesmo momento de tempo.

Em termos gerais o modelo restrito apresenta a seguinte expressão:

$$Y_t = f(y_t) \quad (2)$$

Ou então:

$$Y_t = a + b_0 X_t + b_1 X_t + b_2 X_t + b_3 X_t + b_4 X_t + b_5 X_t + u_t \quad (3)$$

2.3. Modelo Dinâmico

Segundo Chaves *et al.* (2000:118), “os modelos dinâmicos são modelos em que o tempo desempenha um papel fundamental.²¹ Tipicamente, esta dependência no tempo é obtida pela inclusão, como variável explicativa, da própria variável desfasada.”

Em termos gerais o Modelo Dinâmico apresenta a seguinte forma:

$$Y_{t-1} = f(y_{t-1}) \quad (4)$$

Ou então:

²¹ Por modelo dinâmico entende-se a classe de modelos que podem ser expressos por equações a diferenças.



$$Y_t = a + b_0 X_t + b_1 X_t + b_2 X_t + b_3 X_t + b_4 X_t + b_5 X_t + b_6 X_{t-1} + \\ + b_7 X_{t-1} + b_8 X_{t-1} + b_9 X_{t-1} + b_{10} X_{t-1} + b_{11} X_{t-1} + u_t \\ (5)$$

Para se poder efectuar a estimação dinâmica do modelo é feita vão introduzir-se as variáveis desfasadas também conhecidas por “lags”²², ou seja, o factor tempo.

3. Análise e Previsão da Procura Turística na Região Norte de Portugal

3.1. Apresentação e Comportamento das Variáveis do Modelo

No presente estudo optou-se por trabalhar variáveis, que influenciam a procura turística, tais como: Permanência Média; Índice Harmonizado de Preços no Consumidor, para o mercado emissor de Espanha; N.º de Desempregados para os mercados emissores de Portugal e Espanha. Seguidamente serão analisadas cada uma destas variáveis bem como a variável dependente Dormidas nas unidades hoteleiras da Região Norte.

De sublinhar que o facto de se ter escolhido Espanha e Portugal como principais mercados emissores deve-se ao peso que ocupam, em 2008, no total das dormidas da Região Norte de Portugal, com um total de cerca de 60% e 13%, respectivamente, ou seja, a respectiva quota de mercado. Importa referir que a introdução da variável permanência aplica-se e é justificável devido ao peso que estes dois mercados ocupam no total das dormidas da Região Norte de Portugal. Ainda, e porque a variável Índice Harmonizado de Preços no Consumidor para o mercado emissor Portugal apresentou problemas de colinearidade, optou-se por não incorporá-la no modelo, logo não será objecto de análise.

As dormidas na Região Norte de Portugal consistem na estada de um indivíduo num estabelecimento que fornece alojamento por um período compreendido entre as 12 horas de um dia e as 12 horas do dia seguinte (INE, 2007), ou seja, é o número total de dormidas registadas na região em estudo. A variável dormidas da Região Norte de Portugal tem como unidade de medida o número de dormidas, o mesmo é dizer o n.º de noites que um hóspede passou na região. Os dados recolhidos e apresentados na Figura 1 consideram o período compreendido entre Janeiro de 1996 e Dezembro de 2008, correspondendo a 156 dados mensais ao longo de 13 anos. Analisando o

²² As “Lags” permitem introduzir o factor tempo como factor explicativo da variação das variações ocorridas na variável dependente.



16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

comportamento da série verifica-se a presença de sazonalidade (valores máximos nos meses de Verão e mínimos nos meses de Inverno), isto é, os valores das dormidas vêm em função da altura do ano. Constata-se ainda a existência de uma tendência crescente ao longo do horizonte temporal em estudo. Pode verificar-se que a série apresenta um crescimento acentuado a partir de 1998 até 2001, com um decréscimo não muito significativo até 2004, e um significativo crescimento entre os anos de 2005 e 2008. Este crescimento pode ser resultado de investimentos realizados em variáveis de marketing que permitiram projectar a região não só a nível nacional mas também a uma dimensão internacional.

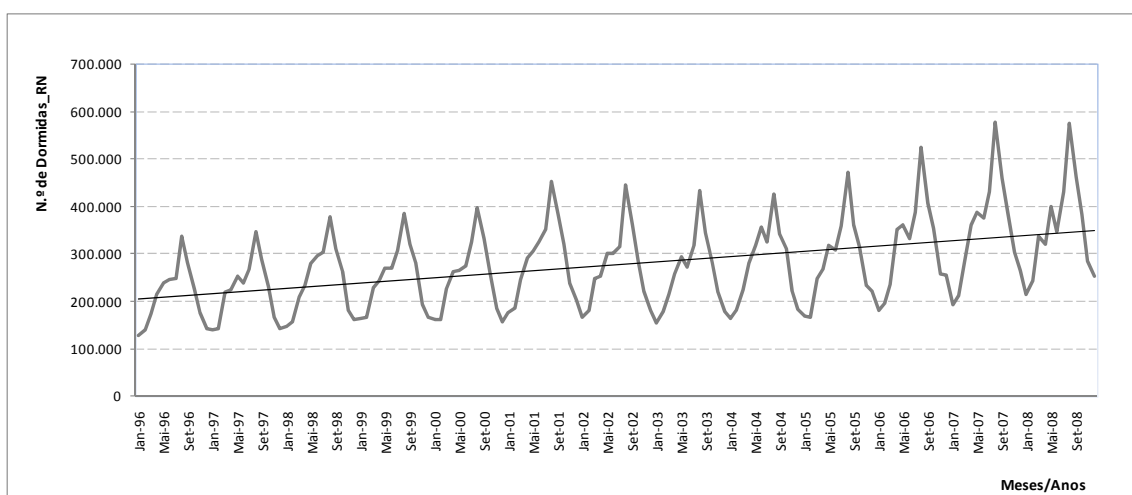


Figura 1: Dormidas nas Unidades de Alojamento, no Norte de Portugal, [Jan-96:Dez-08].

A permanência média consiste numa relação entre o número de dormidas na região Norte de Portugal e o número de hóspedes que deram motivo a essas dormidas. A unidade de medida desta variável é o número de dias que um determinado indivíduo permanece num estabelecimento hoteleiro (INE, 2010). Assim e analisando a Figura 2, para o período em análise verifica-se que os hóspedes passam em média aproximadamente 2 noites na região em estudo (1,77 noites), embora os valores encontram-se compreendidos entre 1,6 noites, em Janeiro de 2006, valor mais baixo que a série apresenta e 2,1 noites, em Junho de 2004, valor mais alto. Esta última situação pode justificar-se pela coincidência com o evento desportivo, EURO2004, referindo-se que 5 dos 10 estádios de futebol encontram-se na região Norte.



16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

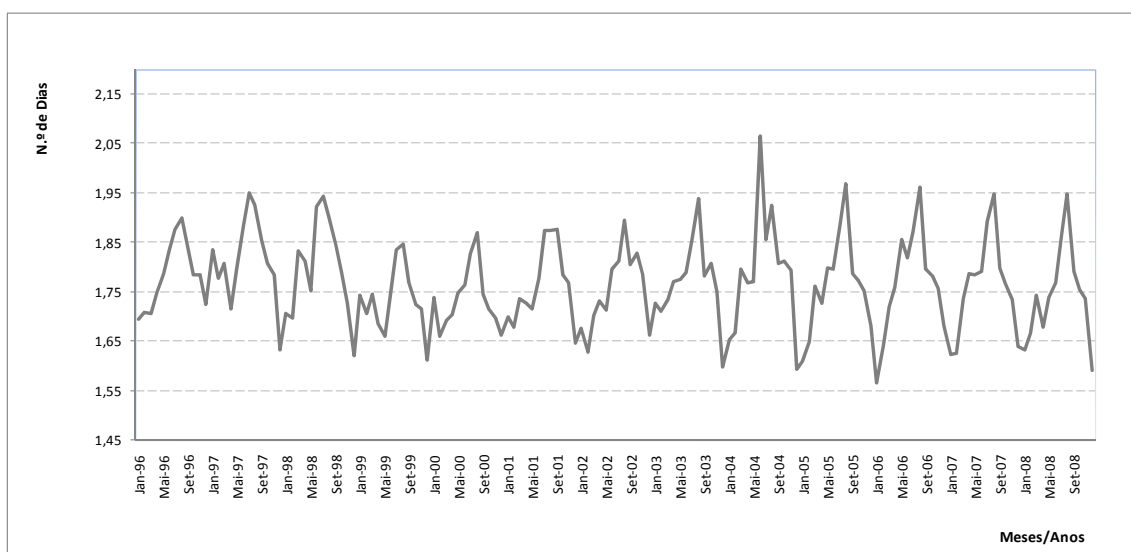


Figura 2: Permanência média [Jan-96:Dez-08].

O Índice de Preços no Consumidor (IPC) é um indicador que tem por finalidade medir a evolução no tempo dos preços de um conjunto de bens e serviços considerados representativos da estrutura de consumo da população residente nos principais mercados emissores (INE, 2010). Optou-se por considerar o Índice Harmonizado de Preços no Consumidor (IHPC) por ser o indicador de inflação mais apropriado para comparações entre os diferentes países, neste caso da União Europeia. Na Figura 3, apresenta-se o comportamento da série IHPC₀₅ para o mercado emissor de Espanha, onde se observa uma tendência crescente, para o período em análise, reflectindo um aumento nos preços do cabaz o que indicia uma diminuição do poder de compra. Os dados foram obtidos na base de dados da EUROSTAT.

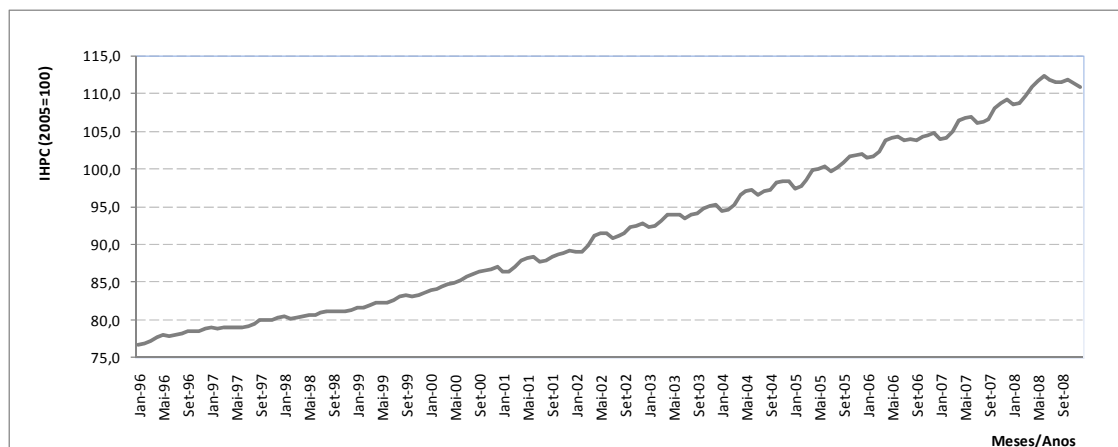


Figura 3: Índice Harmonizado de Preços no Consumidor, Espanha, [Jan-96:Dez-08].



16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

O número de desempregados é um indicador relevante da situação económica de um país, podendo ser um dos indicadores indirectos da propensão para o consumo e procura de bens e serviços não essenciais, como é o caso dos serviços prestados no sector turístico. Na figura seguinte pode observar-se uma diferença significativa quando se comparam as duas séries, pois a série de Espanha apresenta valores, em média, 7 vezes superiores à série de Portugal. Enquanto a série de Portugal apresenta uma flutuação quase constante, já a série de Espanha a flutuação torna-se mais evidente, embora ambas apresentem o mesmo comportamento durante o período de análise. Os dados foram obtidos na base de dados da EUROSTAT.

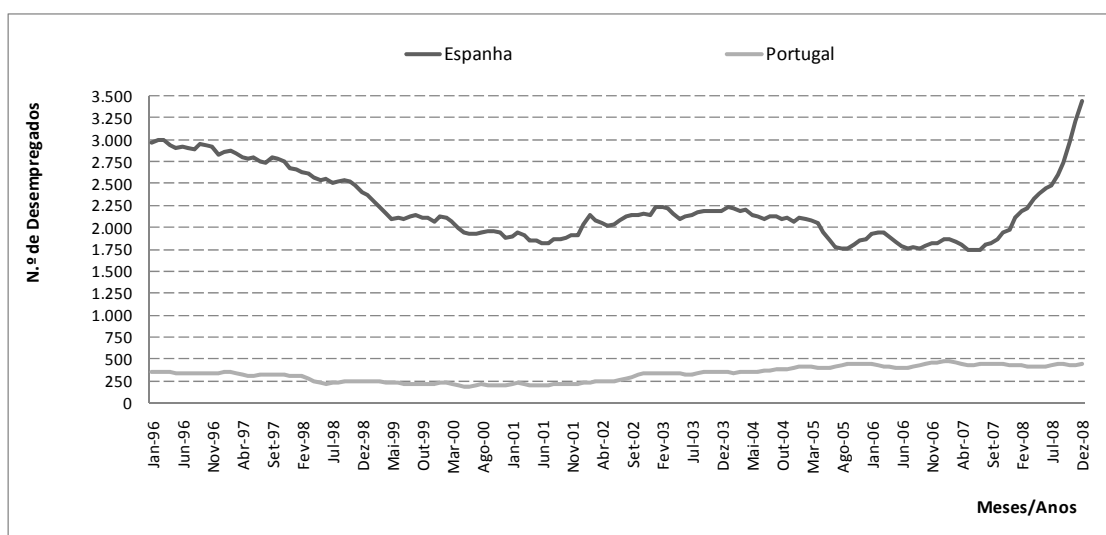


Figura 4: N.º de Desempregados de Portugal e Espanha, [Jan-96:Dez-08].

4. Aplicação da Metodologia

4.1. Modelo Estático

Como já foi referido, em secções anteriores, as variáveis que serviram de base à construção do modelo foram a Permanência Média [PM], Índice Harmonizado de Preços no Consumidor-Espanha [IHPC_SP], N.º de Desempregados de Espanha [ND_SP] e N.º de Desempregados de Portugal [ND_PT]. Assim, o modelo matemático pode escrever-se da seguinte forma:

$$Dormidas_t = a + b_0 PM + b_1 IHPC_SP + b_2 ND_SP + b_3 ND_PT + u_t \quad (6)$$

Seguidamente apresentam-se os resultados obtidos para o modelo estimado pela aplicação de OLS (Tabela 1).



16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

O coeficiente de determinação é de $0,767$ e indica que as variáveis permanência média, índice de preço ao consumidor de Espanha, n.º de desempregados de Espanha e o n.º de desempregados de Portugal explicam cerca de 77% das variações que ocorrem nas dormidas na região Norte de Portugal. O coeficiente de determinação ajustado é de $0,76$ e indica que cerca de 76% das variações ocorridas nas dormidas da região Norte de Portugal foram explicadas pelas variações ocorridas nas variáveis independentes.

Tabela 1: Medidas de Desempenho do Modelo Estimado (OLS).

	<i>Coeficiente</i>	<i>Erro Padrão</i>	<i>rácio-t</i>	<i>valor p</i>	
const	-1500480	97702,6	-15,3577	<0,00001	***
ND_SP	-1,59526	12,2612	-0,1301	0,89666	
IHPC_SP	6525,58	658,767	9,9057	<0,00001	***
ND_PT	-349,106	71,6531	-4,8722	<0,00001	***
PM	732563	40087,6	18,2741	<0,00001	***
Média var. dependente	276382,4	D.P. var. dependente		91179,13	
Soma resíd. quadrados	3,00e+11	E.P. da regressão		44591,18	
R-quadrado	0,767002	R-quadrado ajustado		0,760830	
F (4,151)	124,2685	valor P(F)		1,01e-46	
Log. da verosimilhança	-1888,839	Critério de Akaike		3787,678	
Critério de Schwarz	3802,927	Critério Hannan-Quinn		3793,871	
rho	0,306298	Durbin-Watson		1,385916	

Se a variável permanência média variar um dia, a variável dormidas na região Norte de Portugal aumenta em cerca de 732.563 dormidas, existindo uma relação positiva entre estas duas variáveis. Esta variável é estatisticamente significativa.

Se a variável índice harmonizado de preço ao consumidor de Espanha variar em uma unidade, a variável dormidas na região Norte de Portugal aumenta em cerca de 6.526 dormidas, existindo uma relação positiva entre estas duas variáveis. Esta variável é estatisticamente significativa, a um nível de significância de 1%, ou seja, 99% do valor da variável índice de preço ao consumidor de Espanha é um valor correcto.

Variando a variável número de desempregados de Espanha em uma unidade, a variável dormidas na região Norte de Portugal diminuirá em cerca de 1.60 dormidas, existindo uma relação inversa entre estas variáveis. Esta variável não tem significância estatística.

Se o número de desempregados de Portugal aumentar em uma unidade, a variável dormidas na região Norte de Portugal diminuirá em cerca de 349 dormidas, existindo uma relação inversa entre estas variáveis. Esta variável é estatisticamente significativa, a um nível de significância de 1%, ou seja, 99% do valor da variável número de desempregados de Portugal é um valor correcto.



Relativamente à estatística do F ($4,151$) = $124,2685$, valor de prova = $1,01e-46$, inferior a 1% aceita-se a hipótese que existem variáveis que assumem valores diferentes de zero e como já foi referido, anteriormente, as variáveis incluídas no modelo em conjunto explicam de forma satisfatória as variações ocorridas na variável dormidas na região Norte de Portugal.

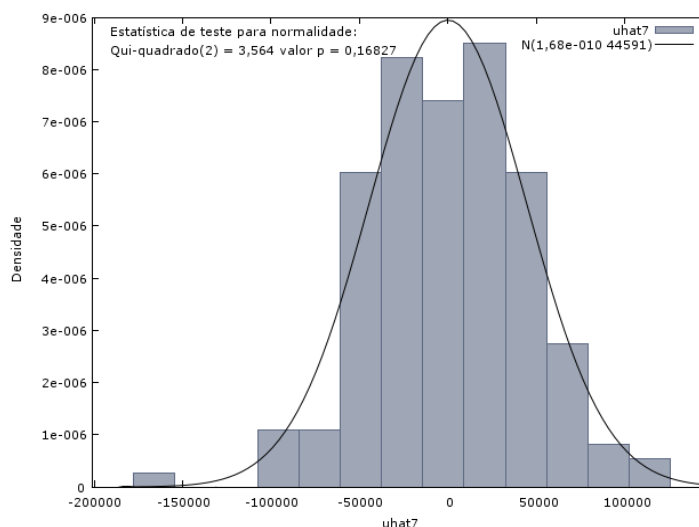
No que diz respeito à análise da infracção às hipóteses básicas do MLG deve salientar-se que:

- quanto à multicolinearidade e tendo por base os valores do VIF^{23} , verifica-se que não existe infracção à hipótese básica da multicolinearidade, uma vez que, os valores do VIF para as variáveis é inferior a 10 valores. Pode concluir-se que existe ausência de independência das variáveis explicativas, ou seja, as variáveis não tem qualquer relação entre elas. Importa referir que os estimadores permanecem BLUE;

Tabela 2: VIF do modelo estático.

Variáveis	VIF
ND_SP	1,704
IHPC_SP	3,765
ND_PT	3,014
PM	1,009

- o teste da normalidade do resíduo feito através da estatística de teste $\chi^2=3,564$, com valor de prova= $0,16827$, o que significa que este modelo segue uma distribuição normal a um nível de significância de 5%, logo esta hipótese não é violada;



²³ Do inglês *Variance inflation factors*.

Figura 5: Distribuição normal do modelo estático.

- Através da observação gráfica (Figura 5) verifica-se que a média é igual $\mu = 1,68e-010$. Este valor é aproximadamente zero então a hipótese da média zero também não é infringida $E(\mu) = 0$;

- Quanto à Homocedasticidade, variância constante do termo de erro, através do teste de *White* para a heterocedasticidade e da estatística de teste $TR^2=45,6439$ com valor de prova $(\chi^2(14) > 45,6439) = 0,0032$, como o valor de prova é inferior a 10% conclui-se que se rejeita a hipótese da homocedasticidade de acordo com os resultados obtidos, pelo que houve a necessidade de corrigir utilizando o teste de Breusch-Pagan (p-value=0,175491) obtendo-se um valor superior a 5%, pelo que se pode concluir que não existe infracção à homocedasticidade, isto é, a variância é constante de observação para observação. Não existe perda das características dos estimadores OLS, continuam a ser BLUE;

- Nas tabelas estatísticas para quatro variáveis independentes o dl é igual a 1,68, o du igual a 1,79, o 4-du igual a 2,20 e por fim o 4-dl é igual a 2,32. Obteve-se a seguinte estatística de *Durbin-Watson*=1,385. O valor da estatística de *Durbin-Watson* encontra-se na zona de autocorrelação positiva. Então pode concluir-se que existe infracção à independência do termo de erro e que este modelo sofre de autocorrelação dos erros. Para tentar ultrapassar este problema, ou seja, tentar corrigir a infracção à hipótese da independência dos erros aplicou-se o teste de *Cochrane - Orcutt*, pelo que através da estimação obteve-se a seguinte estatística de *Durbin-Watson* =1,8260 encontra-se na zona de independência dos erros e o modelo restrito não infringe a hipótese da independência dos termos de erro.

4.2. Modelo Dinâmico

O modelo dinâmico representa as relações de uma determinada variável num determinado momento relacionado com variáveis relacionadas nos momentos anteriores. Pelo no caso em estudo vem dado pela seguinte expressão:

$$\begin{aligned}
 Dormidas_t = & a + b_0 PM + b_1 IHPC_SP + b_2 ND_SP + b_3 ND_PT + b_{4t-1} PM + \\
 & + b_{5t-1} IHPC_SP + b_{6t-1} ND_SP + b_{8t-1} ND_PT + u_{t-1}
 \end{aligned}
 \tag{7}$$



16º Congresso da APDR

Universidade da Madeira, Funchal

Colégio dos Jesuítas, 8 a 10 Julho 2010

Seguidamente apresentam-se os resultados obtidos para o modelo dinâmico estimado pela aplicação de OLS (Tabela 3).

O coeficiente de determinação é de 0,818 e significa que as variáveis permanência média, índice de preço ao consumidor de Espanha, nº de desempregados de Espanha e o n.º de desempregados de Portugal no tempo actual e no tempo anterior explicam cerca de 82% das variações que ocorrem nas dormidas na região Norte de Portugal. O coeficiente de determinação ajustado é igual a 0,81 e significa que cerca de 81% das variações ocorridas nas dormidas da região Norte de Portugal foram explicadas pelas variações ocorridas nas variáveis independentes no tempo actual e no tempo anterior.

Tabela 3: Medidas de Desempenho do Modelo Estimado (OLS).

	<i>Coeficiente</i>	<i>Erro Padrão</i>	<i>rácio-t</i>	<i>valor p</i>	
const	-1694910	107106	-15,8246	<0,00001	***
ND_SP	81,2544	76,3415	1,0644	0,28893	
ND_SP_1	-98,1646	84,2617	-1,1650	0,24592	
IHPC_SP	13785,5	7698,84	1,7906	0,07543	*
IHPC_SP_1	-7659,95	7778,8	-0,9847	0,32639	
ND_PT	-1402,15	409,865	-3,4210	0,00081	***
ND_PT_1	1063,73	412,761	2,5771	0,01095	**
PM	581335	45602	12,7480	<0,00001	***
PM_1	298373	46791,9	6,3766	<0,00001	***
Média var. dependente	277346,7	D.P. var. dependente		90673,07	
Soma resid. quadrados	2,30e+11	E.P. da regressão		39700,50	
R-quadrado	0,818253	R-quadrado ajustado		0,808294	
F(8, 146)	82,16426	valor P(F)		3,39e-50	
Log. da verosimilhança	-1856,613	Critério de Akaike		3731,226	
Critério de Schwarz	3758,617	Critério Hannan-Quinn		3742,352	
rho	0,390048	Durbin-Watson		1,214135	

A componente autónoma indica que 1694.910 dormidas na região Norte de Portugal não são explicadas pelas restantes variáveis independentes. Esta variável é estatisticamente significativa a um nível de significância de 1%.

Se a variável permanência média variar um dia, a variável dormidas na região Norte de Portugal aumenta em cerca de 581.335 dormidas, existindo uma relação positiva e directa entre estas duas variáveis. Esta variável é estatisticamente significativa a um nível de significância de 1%. Em relação ao período anterior também existe uma relação positiva, pelo aumentando uma noite a permanência média também vem aumentada o número de dormidas em 298.373. A permanência média no período anterior é significativamente estatística a um nível de significância de 1%.



Variando o índice de harmonização do preço no consumidor de Espanha em uma unidade, a variável dormidas na região Norte de Portugal aumenta em cerca de 13786 dormidas, existindo uma relação positiva entre estas duas variáveis. Esta variável é estatisticamente significativa a 10%. O comportamento da variável no período de tempo anterior é o inverso, dado que apresenta um sinal negativo.

Se o número de desempregados de Espanha aumentar, a variável dormidas na região Norte de Portugal variará em sentido directo em cerca de 81 dormidas. Esta variável não tem significância estatística. A variável no período anterior apresenta uma relação inversa, se o número de desempregados de Espanha no período anterior aumentar, a variável dormidas na região Norte de Portugal diminuirá cerca de 98 dormidas. Esta variável também não tem significância estatística.

Relativamente à variável número de desempregados de Portugal, se esta aumentar, a variável dormidas na região Norte de Portugal diminui em cerca de 1400 dormidas, existindo uma relação inversa entre estas variáveis. Esta variável é estatisticamente significativa, a um nível de significância de 1%, ou seja, 99% do valor da variável número de desempregados de Portugal é um valor correcto. A variável no período anterior tem uma relação positiva directa com a variável dependente, se o número de desempregados de Portugal no período anterior aumentar a variável dormidas na região Norte de Portugal aumentará em cerca de 1063 dormidas. Esta variável é estatisticamente significativa, a um nível de significância de 5%, ou seja, 95% do valor da variável número de desempregados de Portugal no período anterior é um valor correcto.

Quanto à estatística $F(8,146) = 82,164$ (valor de prova = 0,00000), valor de prova inferior a 10%, aceita-se a hipótese que existem variáveis que assumem valores diferentes de zero. Apesar de existir alguma insignificância de algumas variáveis individuais, as variáveis no momento actual e no momento anterior incluídas no modelo, em conjunto explicam de forma satisfatória as variações ocorridas na variável dormidas na região Norte de Portugal.

No que diz respeito à análise da infracção às hipóteses básicas do MLG, dinâmico, deve salientar-se que:



- através da análise à tabela seguinte verifica-se que existe infracção à hipótese básica da multicolinearidade, uma vez que, os valores do VIF, para quase todas as variáveis, no momento actual e no momento anterior é superior a 10 no geral. Pode concluir-se que existe independência das variáveis explicativas, ou seja, a variável permanência média, índice de preço ao consumidor de Espanha, o número de desempregados de Espanha e o número de desempregados de Portugal com as mesmas variáveis no período de tempo anterior tem uma relação significativa entre elas;

Tabela 4: VIF do modelo dinâmico.

Variáveis	VIF	Variáveis	VIF
IHPC SP	639,874	IHPC SP 1	648,622
ND SP	81,227	ND SP 1	94,816
ND PT	124,313	ND PT 1	124,703
PM	1,640	PM 1	1,691

- a presença de multicolinearidade implica que as variáveis explicativas estão correlacionadas entre si, tendo como consequência a impossibilidade do cálculo dos estimadores mínimos quadrados, impossibilita a separação dos efeitos individuais das variáveis explicativas, impede o normal funcionamento dos testes de significância, uma vez que os desvios padrões aparecem com valores elevados, obtêm-se estimadores instáveis e sensíveis, contudo não afecta a propriedade estatísticas dos estimadores – os estimadores dos mínimos quadrados permanecem *BLUE*. Para correcção desta infracção podia-se estimar o modelo utilizando as primeiras diferenças, sugerindo-se esta recomendação para futuros trabalhos;

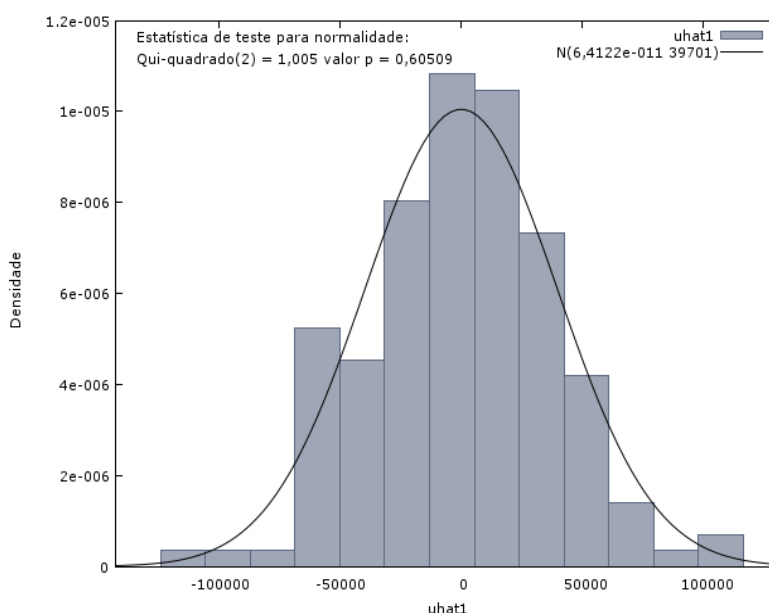


Figura 6: Distribuição normal do modelo dinâmico.

- teste da normalidade do resíduo feito através da estatística de teste $\chi^2(2) = 1,00477$, com valor de prova = 0,605086, significando que este modelo segue uma distribuição normal a um nível de significância de 10%, logo esta hipótese não é infringida;
- através da observação gráfica verifica-se que a média é igual $\mu = 6,41e-011$. Este valor é aproximadamente zero então a hipótese da média zero também não é infringida;
- quanto à análise da homocedasticidade e utilizando o teste de *White* para a heterocedasticidade através da estatística de teste $TR^2 = 62,104$ com valor $p = P(\chi^2(44) > 62,104) = 0,037225$, como o valor de prova é inferior a 10% conclui-se que se aceita a hipótese da heterocedasticidade, infringindo a hipótese básica do termo de erro. Para tentar ultrapassar este problema, ou seja, tentar corrigir a infracção à hipótese da homocedasticidade irá efectuar-se o teste *Breusch-Pagan*. Através da estimação obteve-se a seguinte estatística teste $TR^2 = 16,534$ com valor $p = P(\chi^2(8) > 16,534) = 0,035343$, como o valor de prova é inferior a 10% conclui-se que se aceita a hipótese da heterocedasticidade e apesar da utilização do teste *Breusch-Pagan* o modelo continua a infringir a hipótese da variância constante do termo de erro;
- de acordo com os resultados obtidos pode concluir-se que existe infracção à homocedasticidade, isto é, a variância não é constante de observação para observação, neste sentido existe perda das características dos estimadores dos mínimos quadrados, continuam a ser não enviesados e consistentes contudo deixam de ser eficientes, isto é, não são estimadores com a variância mínima. A heterocedasticidade afecta também a validade dos testes de hipótese e dos intervalos de confiança, uma vez que as variâncias dos estimadores estão alteradas e sobre-estimadas. Como solução para ultrapassar o problema da heterocedasticidade se a variância for conhecida pode utilizar-se o método dos mínimos quadrados ponderados (*Weighted Least Squares*), se a variância for desconhecida uma solução é utilizar-se a transformação da raiz quadrada, ficando como recomendação para futuros trabalhos;
- através das tabelas estatísticas para oito variáveis independentes - no momento actual e no momento anterior, o dl é igual a 1,63023, o du é igual a 1,8469, o 4-dl é igual a 2,3697 e por fim o 4-du é igual a 2,1531. Obteve-se a seguinte estatística de *Durbin-*



$Watson=1,214$. O valor da estatística de *Durbin-Watson* encontra-se na zona de autocorrelação positiva, então pode concluir-se que existe infracção à independência do termo de erro e o modelo dinâmico sofre de autocorrelação dos erros. Para tentar ultrapassar este problema, ou seja, tentar corrigir a infracção à hipótese da independência dos erros irá efectuar-se o teste *Cochrane – Orcutt*. Através da estimação obteve-se a seguinte estatística de *Durbin-Watson*=1,8817, encontra-se na zona de independência dos erros e o modelo dinâmico não infringe a hipótese da independência dos termos de erro. Neste sentido pode concluir-se que no modelo dinâmico os erros são independentes entre si, os estimadores dos mínimos quadrados são os estimadores com variância mínima, isto é, são eficientes, consistentes e não enviesados.

4.3. Escolha do melhor modelo

De modo geral, pela interpretação dos resultados apresentada anteriormente o melhor modelo seria o restrito. Contudo para comprovar esta hipótese necessita-se de se efectuar o teste de significância das restrições lineares, sendo elas:

H_0 : modelo restrito;

H_1 : modelo dinâmico.

Ou seja,

$H_0 : b_7 = b_8 = b_9 = b_{10} = b_{11} = b_{12} = 0$;

$H_1 : b_7, b_8, b_9, b_{10}, b_{11}, b_{12} \neq 0$.

Conjunto de restrições:

1: $b[\text{IHPC_SP_1}] = 0$

2: $b[\text{ND_SP_1}] = 0$

3: $b[\text{ND_PT_1}] = 0$

4: $b[\text{PM_1}] = 0$

Pela estatística do teste do F (4, 146) = 82,1643, com valor de prova=0,33936, significa que se aceita a hipótese do modelo restrito, uma vez que o valor de prova é superior a



10%. Neste sentido as variáveis no período actual contribuem para a explicação das dormidas na região Norte de Portugal.

5. Análise e Discussão dos resultados

Pela análise efectuada concluiu-se que o modelo restrito foi o que produziu melhores resultados, ou seja, tem validade para o conjunto de dados que lhe serviu de suporte e apresentou qualidades estatísticas e de ajustamento aceitáveis evidenciando ser adequado para explicar o comportamento da série Dormidas registadas na região Norte de Portugal.

A introdução das variáveis explicativas no período de tempo anterior não é relevante para a explicação da variação ocorrida nas dormidas na região Norte de Portugal no período actual.

Os coeficientes de determinação do modelo dinâmico melhoraram em relação ao do modelo restrito. De um coeficiente de determinação de cerca de 77% no modelo restrito passou-se para um coeficiente de determinação de cerca de 82%, de facto houve um aumento significativo. Este aumento do coeficiente de determinação poderia ser pelo facto de se introduzirem as variáveis desfasadas no modelo, no entanto o coeficiente de determinação ajustado que é corrigido pelos graus de liberdade e não considera o elevado número de variáveis explicativas, o seu valor passou de cerca de 76% no modelo restrito para cerca de 80 % no modelo dinâmico.

Em ambos os modelos as hipóteses da normalidade do termo, média zero, independência dos erros foram garantidas, contudo no modelo dinâmico a hipótese da multicolinearidade, homocedasticidade foram violadas.

Através do teste de *Fisher* verifica-se que as variáveis incluídas no modelo em conjunto explicam de forma satisfatória as variações ocorridas nas dormidas na região Norte de Portugal.

Com o teste de significância das restrições lineares concluiu-se que se deve aceitar a hipótese do modelo restrito, pois é o modelo que garante as hipóteses básicas do modelo linear geral e os estimadores permanecem BLUE, ou seja, foi o que produziu melhores resultados, tem validade para o conjunto de dados que lhe serviu de suporte e



apresentou qualidades estatísticas e de ajustamento aceitáveis evidenciando ser adequado para explicar o comportamento da série Dormidas registadas na região Norte de Portugal.

6. Conclusão e linhas futuras de investigação

O modelo OLS, na forma restrita, foi o modelo que produziu melhores resultados e se adequou à série original Dormidas na Região Norte de Portugal, comparando com o modelo dinâmico analisado.

O modelo restrito não violou as hipóteses básicas, apresentando um coeficiente de determinação e coeficiente de determinação ajustado de aproximadamente 77% e 76%, respectivamente, pelo que se entendeu ser um bom modelo, gerando estimadores BLUE (*Best Linear Unbiased Estimators*).

O modelo OLS dinâmico (inclui as variáveis desfasadas), as variáveis explicativas estão correlacionadas entre si, tendo como consequência a impossibilidade do cálculo dos estimadores mínimos quadrados, impossibilita a separação dos efeitos individuais das variáveis explicativas.

O modelo OLS dinâmico precisa de ser corrigida a heterocedasticidade, com testes mais avançados, uma vez que heterocedasticidade apresentada e corrigida com os teste *White* e *Breusch-Pagan* não foi suficiente para eliminar tal fenómeno, ou seja, a variância não é constante de observação para observação, neste sentido existe perda das características dos estimadores dos mínimos quadrados - deixam de ser eficientes, isto é, não são estimadores com a variância mínima.

Estas infracções afectam também a validade dos testes de hipótese e dos intervalos de confiança.

Deste modo concluiu-se, que se deveria aceitar o modelo restrito para prosseguir no estudo e assim ir de encontro ao objectivo deste estudo, ou seja, foi o que apresentou qualidades estatísticas mais satisfatórias e o que melhor explicou o comportamento da variável Dormidas na Região Norte de Portugal.

Ao longo do tratamento da parte empírica e uma vez que o modelo dinâmico violou as hipóteses básicas do termo de erro (hipótese da variância constante do termo de erro),



bem como a hipótese básica da multicolinearidade, sugere-se em futuras investigações a utilização de testes mais avançados que permitam corrigir estas infracções.

Sugere-se também a introdução no modelo de mais variáveis explicativas, como por exemplo: a temperatura, motivos das viagens, taxas de câmbio, gastos médios no destino turístico, entre outras.

Toda esta investigação e metodologia, referida anteriormente e uma vez que foi testada para a região Norte de Portugal, poderá ser alargada a um estudo para Portugal.

Como limitação tem a apontar-se a inexistência de dados estatísticos para os anos de anteriores a 1996, para a variável índice de harmonização de preços no consumidor quer para Espanha quer para Portugal, o que balizou assim o estudo das autoras.

Referencias Bibliográficas

Chaves, Cristina; Maciel, Elisabete; Guimarães, Paulo e Ribeiro, José Carlos; (2000); “*Instrumentos estatísticos de apoio à economia: conceitos básicos*”; Lisboa: McGraw-Hill. ISBN 972-773-034-5.

Cunha, Licínio; (2003); “*Introdução ao turismo*”; Editorial Verbo, Lisboa/São Paulo.

EUROSTAT; [http:// epp.eurostat.ec.europa.eu](http://epp.eurostat.ec.europa.eu). (acedido em 20 de Fevereiro de 2010).

Fernandes, P. Odete; Teixeira, J. Paulo; Ferreira, João M. and Azevedo, Susana G.; (2008); “Modelling Tourism Demand: A Comparative Study between Artificial Neural Networks and the Box-Jenkins Methodology”; *Romanian Journal of Economic Forecasting*, n.º 5(3), pp. 30/50. ISBN:978-84-92453-69-6.

Frechtling, C; (2009); “The Tourism Satellite account – a primer”; *Annals of Tourism Research*, Vol. 37, n.º 1, pp. 136/153.

Goh, Carey and Law, Rob; (2002); “Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention”; *Tourism Management*; n.º 23, pp.499/510.

Guizzardi, A. and Mazzocchi, M. (2010); “Tourism demand for Italy and the business cycle”; *Tourism Management*, 31, pp. 367/377.

Gujarati, Damodar N.; (1995); “*Basic Econometrics*” (3ª edição); McGraw-Hill; ISBN 0-07-113964-8.



- INE; (1996-2010); “*Estatísticas do Turismo*”; Lisboa, Instituto Nacional de Estatística.
- Johnston, J. e Dinardo, J.; (2000); “*Métodos Econométricos*”; 4ª Edição, McGraw-Hill.
- Kairat, T; (2010); “Asymptotic distribution of the OLS estimator for a mixed spatial model”; *Journal of Multivariate Analysis*, 101, pp. 733/748.
- Liam, C.; (1997); “An econometric classification and review of international tourism demand models”; *Tourism Economics*, n.º 3, pp. 69/81.
- Makridakis, Spyros and Hibon, Michèle; (1997); “ARMA Models and the Box-Jenkins Methodology”; *Journal of Forecasting*; Vol. 16; pp.147/163.
- Moutinho, Luiz; Huarng, K.H.; Yu, T.H.K. and Chen, C.Y; (2008); “Modeling and forecasting tourism demand: the case of flows from Mainland China to Taiwan”; *Service Business*; Vol. 2; n.º 3, pp. 219/232(14).
- Preez, Johann and Witt, Stephen F.; (2003), “Univariate versus multivariate time series forecasting: an application to international tourism demand”; *International Journal of Forecasting*; n.º 19; pp. 435/451.
- Thawornwong, S. and Enke, D.; (2004); “The adaptive selection of financial and economic variables for use with artificial neural networks”; *Neurocomputing*; n.º 6; pp.205/232.
- Thomakos, Dimitrios D. and Guerard, John B.; (2004); “Naive, ARIMA, nonparametric, transfer function and VAR models: A comparison of forecasting performance”; *International Journal of Forecasting*; n.º 20; pp. 53/67.
- Witt, Stephen F. and Witt, Christine A.; (1995); “Forecasting tourism demand: a review of empirical research”; *International Journal of Forecasting*; n.º 11; pp. 447/475.
- Yu, Gongmei and Schwartz, Zvi; (2006); “Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models”; *Journal of Travel Research*; n.º 45, pp. 194/203.
- Zhihua, S. and Qihua, W. (2009); “Checking the adequacy of a general linear model with responses missing at random”; *Journal of Statistical Planning and Inference*, n.º 139, pp. 3588/3604.