

Modelação da procura turística em Portugal

NATÁLIA SOFIA DOS SANTOS * [natália@ipb.pt]

PAULA ODETE FERNANDES ** [pof@ipb.pt]

Palavras-chave | Procura Turística, Séries Temporais, Modelo Linear Geral, Modelação Econométrica.

Objetivos | Dado o crescimento substancial do setor do turismo em Portugal, será de todo proveitoso desenvolver modelos suscetíveis de serem utilizados para modelar e efetuar previsões fiáveis da procura turística, dado que esta assume um papel relevante no processo de planeamento e de tomada de decisões tanto no contexto do setor público como do privado (Fernandes, 2005). Desta forma, o principal objetivo do presente trabalho assentou em modelar a procura turística de Portugal utilizando modelos econométricos, tendo por base o modelo linear geral. Pretendeu-se assim, construir um instrumento que permita obter uma visão antecipada da evolução da procura turística para Portugal. Para tal e como objeto de estudo utilizou-se a série temporal de turismo “dormidas mensais registadas em Portugal”, compreendida entre o período de janeiro de 1996 a dezembro de 2009, dados recolhidos junto da EUROSTAT (2012).

Metodologia | Para modelar a procura turística em Portugal utilizou-se o Modelo Linear Geral (MLG). Este modelo especifica uma relação económica numa forma matemática cujos parâmetros vão ser estimados e confirmados estatisticamente, segundo diferentes autores (e.g., Oliveira *et al.*, 1997; Chaves, 2000; Johnston & Dinardo, 2000; Maroco, 2003; Pestana & Gageiro, 2008; Zhihua & Qihua, 2009; Kairat, 2010). É um modelo linear porque a forma funcional que assume é linear, geral porque se refere a uma regressão múltipla e por fim é estocástico porque inclui um termo de erro. Assim, pretende-se encontrar um modelo econométrico que dê resposta ao principal objetivo deste trabalho, pelo que inicia-se com a construção do modelo estático que 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 (Preez & Witt, 2003; Machado *et al.*, 2010; Santos & Fernandes, 2011): $Y_t = f(X_t)$.

Deste modo, utilizou-se a série temporal do turismo “dormidas mensais registadas em Portugal” [dormidas], como sendo a variável dependente e para a sua explicação utilizaram-se como variáveis explicativas as seguintes: índice harmonizado de preços no consumidor [IHPC], número de desempregados [ND] e Produto Interno Bruto [PIB]. Ainda, optou-se por considerar as variáveis explicativas dos países emissores com peso significativo na procura turística de Portugal em 2009, ou seja, os países com maior quota de mercado, sendo eles: Portugal [PT], 41%; Espanha [SP], 10%; Alemanha [AL], 10%; Reino Unido [RU], 17%, e França [FR], 5%). Para a construção de cada uma das séries, mensais, teve-se por base os dados disponíveis na EUROSTAT (1996-2009). Assim, o modelo matemático (OLS estático) utilizado foi o seguinte:

$$Dormidas_t = \alpha + \beta_0 IHPC_{PT} + \beta_1 IHPC_{SP} + \beta_2 IHPC_{AL} + \beta_3 IHPC_{FR} + \beta_4 IHPC_{RU} + \beta_5 ND_{PT} + \beta_6 ND_{SP} + \beta_7 ND_{AL} + \beta_8 ND_{FR} + \beta_9 ND_{RU} + \beta_{10} PIB_{PT} + \beta_{11} PIB_{SP} + \beta_{12} PIB_{AL} + \beta_{13} PIB_{FR} + \beta_{14} PIB_{RU} + u_t$$

* **Mestre em Gestão das Organizações** pelo Instituto Politécnico de Bragança, **Técnica de Acompanhamento** no Gabinete de Relações Internacionais do Instituto Politécnico de Bragança.

** **Doutorada em Economia Aplicada e Análise Regional** pela Universidade de Valladolid, **Professora Adjunta** do Instituto Politécnico de Bragança.

Através da leitura dos resultados obtidos pôde verificar-se que o coeficiente de determinação ajustado é de 0,114 e significa que as variáveis índice de preço harmonizado no consumidor, n.º de desempregados e produto interno bruto no tempo atual explicam cerca de 11,4% das variações que ocorrem na variável dependente. Apenas as variáveis IHPC para Portugal, Espanha, França e Alemanha, o ND de França e o PIB para Alemanha e França, apresentam significância estatística. Ainda, quanto à multicolinearidade e tendo por base os valores do VIF (*Variance Inflation Factors*), verificou-se que existe infração à hipótese básica da multicolinearidade, uma vez que, os valores do VIF para as variáveis são superiores a 10 valores, podendo-se concluir-se que existe correlação/dependência das variáveis explicativas, ou seja, as variáveis estão relacionadas entre si. O teste da normalidade do resíduo feito através da estatística de teste $\chi^2 = 45,4187$, com valor de prova = $1,37229e-010$, o que significa que este modelo não segue uma distribuição normal, logo esta hipótese é violada. Verificou-se que a média é igual $\mu = 1,3492e-008$. Este valor é aproximadamente zero então a hipótese da média zero também não é infringida. Quanto à homocedasticidade, variância constante do termo de erro, através do teste de White para a heterocedasticidade e da estatística de $TR^2 = 133,012$, com $(\chi^2(135) > 133,012) = 0,532262$, como o valor de prova é superior a 10%, pelo que se pode concluir que não existe infração à homocedasticidade, isto é, a variância é constante de observação para observação. Não existe perda das características dos estimadores OLS, continuam a ser blue (best linear unbiased estimators). O valor da estatística de Durbin Watson (0,782675) encontra-se na zona de autocorrelação positiva. Então pode concluir-se que existe infração 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 infração à hipótese da independência dos erros aplicou-se o teste de Cochran-Orcutt, pelo que através da estimação obteve-se a seguinte estatística de Durbin Watson = 1,661003, continua a encontrar-se na zona de teste inconclusivo. Neste sentido, aplicou-se o teste de Hildreth-Lu, pelo que através da estimação obteve-se a seguinte estatística de Durbin Watson = 1,661149, continua a encontrar-se na zona de teste inconclusivo. Por último, aplicou-se o teste de Prais-Winsten, pelo que através da estimação obteve-se a seguinte estatística de Durbin Watson = 1,665356, encontrando-se na zona de teste inconclusivo. Este modelo sofre de autocorrelação dos erros, ou seja, os erros não são independentes entre si tendo como consequência que os estimadores dos mínimos quadrados não são os estimadores com variância mínima, isto é, não são eficientes embora permaneçam não enviesados.

Assim, dado se terem violado as hipóteses básicas do modelo foi necessário proceder a uma transformação, pelo que se estimou um outro modelo aplicando as primeiras diferenças. Este modelo estático aplicando as primeiras diferenças representa as relações de uma variável num determinado momento relacionado com variáveis relacionadas nos momentos anteriores (Santos & Fernandes, 2011). Assim, o novo modelo pode escrever-se da seguinte forma:

$$\Delta Dormidas_t = \alpha + \beta_0 \Delta IHPC_PT + \beta_1 \Delta IHPC_SP + \beta_2 \Delta IHPC_AL + \beta_3 \Delta IHPC_FR + \beta_4 \Delta IHPC_RU + \beta_5 \Delta ND_PT + \beta_6 \Delta ND_SP + \beta_7 \Delta ND_AL + \beta_8 \Delta ND_FR + \beta_9 \Delta ND_RU + \beta_{10} \Delta PIB_PT + \beta_{11} \Delta PIB_SP + \beta_{12} \Delta PIB_AL + \beta_{13} \Delta PIB_FR + \beta_{14} \Delta PIB_RU + u_t$$

Os resultados obtidos para este último modelo (considerando o tempo atual e o tempo anterior) apresentou um coeficiente de determinação ajustado baixo (0,01%) e apenas a variável explicativa IHPC da Alemanha apresenta significância estatística, a um nível de significância de 1%. No entanto, este modelo não violou as hipóteses básicas do MLG, pelo que deve salientar-se que a multicolinearidade e tendo por base os valores do VIF, verificou-se que não existe infração à hipótese básica da multicolinearidade, uma vez que, os valores do VIF para as variáveis no período de tempo anterior é inferior a 10 valores, podendo concluir-se que existe ausência de independência das variáveis explicativas, ou seja, as variáveis no período de tempo anterior não tem qualquer relação entre elas. Importa referir que os estimadores permanecem BLUE; o teste da normalidade do resíduo feito através da estatística de teste $\chi^2 = 4,1379$, com valor de prova = 0,126318, significando que este modelo segue uma distribuição normal a um nível de significância de 1%, logo esta hipótese não é infringida; observou-se que a média é aproximadamente zero então a hipótese da média zero também não é infringida; quanto à homocedasticidade, variância constante do termo de erro, através do teste de White para a heterocedasticidade e da estatística de teste $TR^2 = 139,773$, com valor de prova $(\chi^2(120) > 139,773) = 0,104782$, significando que este modelo é homocedástico a um nível de significância de 1%, logo aceita-se a hipótese da homocedasticidade, pelo que de acordo com os resultados obtidos pôde concluir-se que não existe infração à homocedasticidade, isto é, a variância é constante de observação para observação. Não existe perda das características dos estimadores OLS, continuam a ser BLUE; obteve-se a seguinte estatística de Durbin-Watson = 1,888882, encontrando-se na zona de teste inconclusivo. Para tentar ultrapassar este problema, ou seja, tentar corrigir a infração à hipótese da independência dos erros aplicou-se o teste de Hildreth-Lu,

pelo que através da estimação obteve-se a seguinte estatística de Durbin Watson =1,993369, este valor da estatística de Durbin-Watson encontra-se na zona de independência dos erros e o modelo das primeiras diferenças não infringe a hipótese da independência dos termos de erro

Principais resultados e contributos | Tendo por base os resultados obtidos, estes revelaram que o modelo das primeiras diferenças 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 em Portugal. Pois, verificou-se que as variáveis explicativas são independentes entre si, o modelo segue uma distribuição normal e apresenta média zero, a variância é constante de observação para observação, os termos de erros das variáveis explicativas também são independentes, pelo que desta forma não existe perda das características dos estimadores OLS. Assim, pretendeu-se contribuir para a criação de um instrumento que permitisse obter uma visão antecipada da evolução da procura turística de Portugal, uma vez que esta tem revelado, nos últimos anos, algumas potencialidades. Por outro lado, e uma vez que se conseguiram obter resultados satisfatórios, os mesmos são suscetíveis de poderem vir a ser utilizados por outros públicos, desde de profissionais do setor, investigadores, entre outros.

Limitações | A principal limitação do estudo consistiu na obtenção de informação necessária de dados estatísticos, para construir as séries temporais, ou seja, existe ausência e incoerência na contabilização de alguns dados estatísticos referentes a algumas variáveis, o que originou a uma reformulação e alteração do plano inicial (inexistência de dados estatísticos para os anos anteriores a 1996, para a variável índice de harmonização de preços no consumidor). Para superar este aspeto, decidiu-se só incluir, analisar e tratar as variáveis cuja construção obedecia a critérios coerentes e tentou-se recolher os dados disponíveis mais recentes e publicados pela EUROSTAT.

Conclusões | O modelo OLS estático estimado, aplicando as primeiras diferenças, foi o modelo que produziu melhores resultados e se adequou à série original dormidas mensais em Portugal, comparando com o modelo estático analisado. O modelo OLS estático necessitou que se corrigisse a autocorrelação dos erros, uma vez que não se conseguiu corrigir com os testes de Cochrane-Orcutt, Hildreth-Lu e Prais-Winsten, ou seja, este modelo apresentou dependências no termo de erro de observação para observação. Estas infrações afetam também a validade dos testes de hipótese e dos intervalos de confiança, como tal aplicou-se as primeiras diferenças para ultrapassar às infrações básicas do modelo linear estático. Neste sentido, concluiu-se, que se deveria aceitar o modelo estático estimado, aplicando as primeiras diferenças, para prosseguir no estudo e assim ir de encontro ao objetivo deste estudo, ou seja, foi o que apresentou qualidades estatísticas mais satisfatórias e o que melhor explicou o comportamento da variável dormidas mensais em Portugal.

Referências

- Chaves, C., 2000, *Instrumentos estatísticos de apoio à economia: conceitos básicos*. McGraw-Hill, Lisboa.
- EUROSTAT, 2012. [<http://epp.eurostat.ec.europa.eu>.] (acedido em 9 de janeiro de 2012).
- Fernandes, P., 2005. *Modelación, Predicción y Análisis del Comportamiento de la Demanda Turística en la Región Norte de Portugal España*: Dissertação de Doutoramento: Universidad de Valladolid.
- Johnston, J. & Dinardo, J., 2000, *Métodos Económicos*. 4ª Edição, McGraw-Hill.
- Kairat, T., 2010, Asymptotic distribution of the OLS estimator for a mixed spatial model, *Journal of Multivariate Analysis*, Vol. 101, pp. 733-748.
- Machado, T., Teixeira, J., & Fernandes, P., 2010, Modelação da procura turística em Portugal: regressão linear versus redes neuronais artificiais, *Revista Turismo & Desenvolvimento*, Vol. 1, pp. 435-445.
- Maroco, J., 2003, *Análise Estatística com utilização do SPSS*, Edições Sílabo, Lisboa.
- Oliveira, M., Aguiar, A., Carvalho, A., Martins, F., Mendes, V. & Portugal, P., 1997. *Econometria Exercícios*, McGraw-Hill, Lisboa.
- Pestana, M. & Gageiro, J., 2008. *Análise de Dados para Ciências Sociais A complementaridade do SPSS*, (5.ª ed.), Edições Sílabo, Lisboa.
- Preez, J. & Witt, S., 2003, Univariate versus multivariate time series forecasting: an application to international tourism demand, *International Journal of Forecasting*, Vol.19, pp. 435-451.
- Santos, N. & Fernandes, P., 2011. Modelação e Caracterização da Procura Turística: O Caso da Região Norte de Portugal. *TÉKHNE – Polytechnical Studies Review*, Vol. 16(9), pp. 119-137.