



Ana Sofia Santos Almeida

TRANSPORTE AÉREO E DESENVOLVIMENTO REGIONAL

AIR TRANSPORT AND REGIONAL DEVELOPMENT

Dissertação de Mestrado Integrado em Engenharia Civil na área de Especialização em Urbanismo, Transportes e Vias de Comunicação,
orientada pelo Professor Doutor António José Pais Antunes

Coimbra, 4 de Maio de 2018



UNIVERSIDADE DE COIMBRA



FCTUC DEPARTAMENTO DE ENGENHARIA CIVIL
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Ana Sofia Santos Almeida

TRANSPORTE AÉREO E DESENVOLVIMENTO REGIONAL

AIR TRANSPORT AND REGIONAL DEVELOPMENT

Dissertação de Mestrado Integrado em Engenharia Civil na área de Especialização em Urbanismo, Transportes e Vias de Comunicação,
orientada pelo Professor Doutor António José Pais Antunes

Esta Dissertação é da exclusiva responsabilidade do seu autor.
O Departamento de Engenharia Civil da FCTUC declina qualquer
responsabilidade, legal ou outra, em relação a erros ou omissões
que possa conter.

Coimbra, 4 de Maio de 2018

AGRADECIMENTOS

A presente dissertação pretende assinalar o fim do meu percurso académico. Durante este percurso, muitas foram as pessoas que me marcaram, incentivaram e ajudaram a crescer a nível pessoal e profissional, e a quem não posso deixar de mostrar o meu profundo agradecimento.

Agradeço ao Professor Doutor António José Pais Antunes, pela orientação, apoio e disponibilidade ao longo dos últimos meses, mesmo quando o tempo era escasso.

Aos meus pais, por todo o apoio durante este percurso e ao longo da vida. Um obrigada especial à minha irmã que, apesar de ser um pouco chata, como os irmãos devem ser, me acompanha desde sempre e a quem desejo a maior sorte ao longo da vida.

Aos meus avós, pela ajuda incondicional ao longo de todo este percurso, e em particular à minha avó, por ter sempre uma palavra amiga capaz de me reconfortar quando preciso.

Agradeço também à minha amiga Ana Paula Lisboa, pelos conselhos e amizade demonstrada ao longo destes anos, por todo o apoio nos bons e maus momentos, e ainda em especial por me acompanhar durante todo o processo de realização desta dissertação. Sem dúvida que sem ela não seria a mesma coisa.

Aos meus amigos, por todos os momentos partilhados ao longo destes anos.

A todos, muito obrigada.

RESUMO

O transporte aéreo, principalmente os aeroportos, assumem um papel essencial na sociedade atual, sendo alvo de um crescente interesse, graças aos grandes investimentos realizados tanto na sua construção, como na sua eventual expansão. Um dos fatores que contribui de forma significativa para o crescimento do transporte aéreo é a sua capacidade de, quando comparado com outros meios de transporte, percorrer longas distâncias em tempo reduzido.

O principal objetivo deste estudo consiste em determinar quais as regiões do Sudoeste Europeu – nomeadamente as regiões NUTS 2 de Espanha, França, Itália e Portugal – que mais eficientemente estão a utilizar o transporte aéreo para o seu desenvolvimento, e quais os fatores que determinam essa eficiência.

Deste modo, aplicaram-se os métodos 2SLS e OLS para estimar os coeficientes de regressão dos modelos de regressão linear múltipla, utilizados para explicar a relação entre o transporte aéreo e o desenvolvimento regional. Com isto, pretendia-se ainda evitar o denominado problema da endogeneidade – é o transporte aéreo que causa o desenvolvimento regional ou o desenvolvimento regional que fomenta o transporte aéreo?

Com o modelo de regressão escolhido, verifica-se que existe uma relação positiva entre o transporte aéreo e o desenvolvimento regional, embora seja pouco acentuada, podendo ser explicada pelo facto de o aumento de 10% no movimento de passageiros resultar em aproximadamente um acréscimo de 1% no Valor Acrescentado Bruto (VAB).

Os resultados obtidos mostraram essencialmente que, a aplicação de 2SLS como método de estimação dos parâmetros de regressão não se torna necessária, isto é, não há razões para se pensar que existe endogeneidade entre as variáveis. Relativamente às regiões que utilizam o transporte aéreo como forma de favorecer o seu desenvolvimento, concluiu-se que a influência do transporte aéreo é mais acentuada nas regiões com grande atração turística, fator esse que contribui para um maior volume de fluxo de passageiros nestas regiões, sendo estas consideradas as que apresentam um desenvolvimento superior ao expectável. Por fim, verifica-se ainda que em Espanha, Itália e Portugal, cerca de 50% das regiões de cada país se

encontram mais desenvolvidas do que o previsto pelo modelo de regressão, contrariamente ao que se verifica em França, onde aproximadamente 70% das regiões se encontra com um desenvolvimento inferior ao esperado.

Palavras-chave: Transporte Aéreo; Desenvolvimento Regional; Sudoeste Europeu; Endogeneidade; 2SLS.

ABSTRACT

Air transport, especially airports, plays an essential role in today's society, and is receiving increasing interest thanks to the great investments made both in its construction and in its eventual expansion. One of the factors that contributes significantly to the growth of air transport is its ability to travel long distances in reduced time when compared to other means of transport.

The main objective of this study is to determine which regions of European Southwest - specifically the NUTS 2 regions of Spain, France, Italy and Portugal - are most efficiently using air transport for their development, and which factors determine this efficiency.

Thus, the 2SLS and OLS methods were used to estimate the regression coefficients of the multiple linear regression models used to explain the relationship between air transport and regional development. With this, it was also intended to avoid the supposed endogeneity problem - is air transport causing regional development or is regional development that encourages air transport?

With the chosen regression model, there is a positive relationship between air transport and regional development, although it is not very pronounced, and can be explained by the fact that a 10% increase in passenger movement results in approximately an increase of 1% in Gross Value Added.

The results obtained showed essentially that the application of 2SLS as a method of estimation of the regression parameters is not necessary, there is no reason to think that there is endogeneity between the variables. Regarding to the regions that use air transport as a way to promote their development, it has been concluded that the influence of air transport is more pronounced in regions with a large tourist attraction, which contributes to a greater flow of passengers in these regions, making those regions the ones with a development superior to the expected one. Finally, in Spain, Italy and Portugal, around 50% of the regions in each country are more developed than the one predicted by the regression model, contrary to what happens in France, where approximately 70% of regions is more underdeveloped than expected.

Keywords: Air transport; Regional Development; European Southwest; Endogeneity; Two Stage Least Squares (2SLS).

ÍNDICE

AGRADECIMENTOS.....	i
RESUMO.....	ii
ABSTRACT.....	iv
ÍNDICE.....	vi
ÍNDICE DE FIGURAS.....	viii
ÍNDICE DE QUADROS.....	x
ABREVIATURAS.....	xi
1 INTRODUÇÃO.....	1
1.1 Enquadramento.....	1
1.2 Objetivos.....	3
1.3 Estrutura da dissertação.....	3
2 METODOLOGIA DE ANÁLISE.....	6
2.1 Ordinary Least Squares (OLS).....	6
2.2 Two Stage Least Squares (2SLS).....	7
2.2.1 Motivação para a avaliação de variáveis instrumentais.....	8
2.2.2 Estimativa de variáveis instrumentais num modelo de regressão múltipla.....	9
2.3 Metodologia aplicada ao estudo.....	11
3 REVISÃO DA LITERATURA.....	14
4 ESTUDO DE CASO.....	21
4.1 Região em estudo.....	21
4.2 Definição de variáveis e recolha de dados.....	25
5 RESULTADOS DO ESTUDO.....	29
5.1 Especificação do modelo de regressão linear e log-log (Cobb-Douglas).....	29
5.1.1 Modelo de regressão linear com seis variáveis explicativas.....	29

5.1.2 Modelo log-log com seis variáveis explicativas.....	32
5.1.3 Modelo log-log com quatro variáveis explicativas.....	36
5.1.4 Modelo log-log com três variáveis explicativas.....	37
5.2 Análise de resíduos.....	39
5.3 Análise dos resultados.....	41
6 CONCLUSÃO.....	45
REFERÊNCIAS BIBLIOGRÁFICAS.....	47
ANEXO A.....	A-1

ÍNDICE DE FIGURAS

Figura 1.1 – Crescimento do tráfego internacional de passageiros (expressos em RPK) em 2017 (ICAO, 2018)	2
Figura 3.1 – Definição das variáveis (Brueckner, 2003)	18
Figura 4.1 – Regiões NUTS 2 de: (a) Espanha, (b) França, (c) Itália, (d) Portugal (NUTS 2013/EU-28, 2015)	23
Figura 5.1 – Estimação do modelo linear com seis variáveis explicativas e com MAER como variável explicativa endógena utilizando o <i>software R</i> e o método 2SLS	30
Figura 5.2 – Estimação do modelo linear com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o <i>software R</i> e o método 2SLS	31
Figura 5.3 – Estimação do modelo linear com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o <i>software R</i> e o método OLS	32
Figura 5.4 – Estimação do modelo log-log com seis variáveis explicativas e com MAER como variável explicativa endógena utilizando o <i>software R</i> e o método 2SLS.....	33
Figura 5.5 – Estimação do modelo log-log com seis variáveis explicativas e com MAER como variável explicativa endógena utilizando o <i>software R</i> e o método OLS.....	33
Figura 5.6 – Estimação do modelo log-log com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o <i>software R</i> e o método 2SLS.....	34
Figura 5.7 – Estimação do modelo log-log com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o <i>software R</i> e o método OLS.....	35

Figura 5.8 – Estimação do modelo log-log com quatro variáveis explicativas e com MAER como variável explicativa endógena utilizando o <i>software R</i> e o método OLS.....	36
Figura 5.9 – Estimação do modelo log-log com três variáveis explicativas e com MPAX como variável explicativa endógena utilizando o <i>software R</i> e o método 2SLS.....	37
Figura 5.10 – Estimação do modelo log-log com três variáveis explicativas e com MPAX como variável explicativa endógena utilizando o <i>software R</i> e o método OLS.....	38
Figura 5.11 – (a) Resíduos <i>versus</i> Valores modelados de VAB, (b) Gráfico de Probabilidade Normal dos Resíduos	40
Figura 5.12 – VAB real <i>versus</i> VAB modelado	41

ÍNDICE DE QUADROS

Quadro 3.1 – Quadro com os principais autores e resultados dos estudos (adaptado de: Van De Vijver <i>et al</i> , 2016)	20
Quadro 4.1 – Regiões NUTS 2 consideradas no estudo	25
Quadro 4.2 – Definição das variáveis	26
Quadro 4.3 – Áreas de educação definidas em ISCED (2011)	27
Quadro 4.4 – Atividades económicas segundo a NACE	27
Quadro 5.1 – Valores dos resíduos	42
Quadro A.1 – Valores das variáveis	A-1

ABREVIATURAS

2SLS – Two Stage Least Squares

EUA – Estados Unidos da América

ICAO – International Civil Aviation Organization

ISCED – International Standard Classification of Education

NACE – Nomenclature statistique des Activités économiques dans la Communauté
Européenne

NUTS – Nomenclature of Territorial Units for Statistics

OLS – Ordinary Least Squares

RPK – Revenue Passenger Kilometers

UE – União Europeia

VAB – Valor Acrescentado Bruto

1 INTRODUÇÃO

1.1 Enquadramento

O transporte aéreo, principalmente os aeroportos, assumem um papel essencial na sociedade atual, sendo alvo de um crescente interesse, graças aos grandes investimentos realizados tanto na sua construção, como na sua eventual expansão.

Para que os aeroportos sejam capazes de atingir todo o seu potencial, é necessário um desenvolvimento constante que vise o cumprimento dos requisitos internacionais e a implementação de infraestruturas capazes de gerir não só as mais modernas aeronaves, mas também o inerente aumento do número de passageiros (Esteves, 2014). Assim sendo, se os aeroportos possuírem características favorecedoras do seu desenvolvimento, isto contribuirá, por sua vez, de forma positiva para o desenvolvimento da região.

Ao nível do transporte aéreo, é de referir que um dos seus principais componentes, além das companhias aéreas (agrupadas principalmente em *full carriers vs* companhias *low-cost*), são os “utilizadores” (passageiros e/ou mercadorias), que podem originar importantes níveis de procura nos aeroportos, graças ao tráfego que originam (Postorino, 2010). É ainda de realçar que a maior parte do tráfego de mercadorias de alto valor é realizado por via aérea, o que leva a que muitas das indústrias se instalem cada vez mais nas proximidades dos aeroportos internacionais, o que contribui para o desenvolvimento destes e, conseqüentemente, das regiões onde estão inseridos.

Um dos fatores que também contribui de forma significativa para o crescimento do transporte aéreo é a sua capacidade de, quando comparado com outros meios de transporte, percorrer longas distâncias em tempo reduzido. Além disso, é também o único meio com capacidade de alcançar áreas remotas ou inacessíveis, ou regiões servidas com uma fraca rede de sistemas de transporte terrestre (rodoviário e ferroviário) (Postorino, 2010).

De acordo com os dados disponibilizados pela ICAO (International Civil Aviation Organization) em Janeiro de 2018, atingiu-se em 2017 um novo recorde, com 4.1 biliões de

passageiros a serem transportados pelas companhias aéreas, o que corresponde a um aumento de 7.1% em relação a 2016. O tráfego mundial de passageiros, expresso em RPK (Revenue Passenger Kilometers – número total de quilómetros percorridos pelos passageiros transportados), também sofreu uma forte melhoria em relação a 2016, registando um aumento de 7.6% (ICAO, 2018).

A procura pelos transportes aéreos tem tido um amplo crescimento, nomeadamente graças à contínua melhoria das condições económicas.

Relativamente ao tráfego internacional de passageiros, expresso em RPK, dados publicados pela ICAO mostram que cresceu 8.0% em 2017, um aumento quando comparado com os 7.8% registados em 2016. A nível da Europa, esta permaneceu como o maior mercado internacional, com 37% de participação no tráfego mundial de passageiros e cresceu 8.1%, como se apresenta na Figura 1.1, sustentada pelas melhores condições económicas na região.

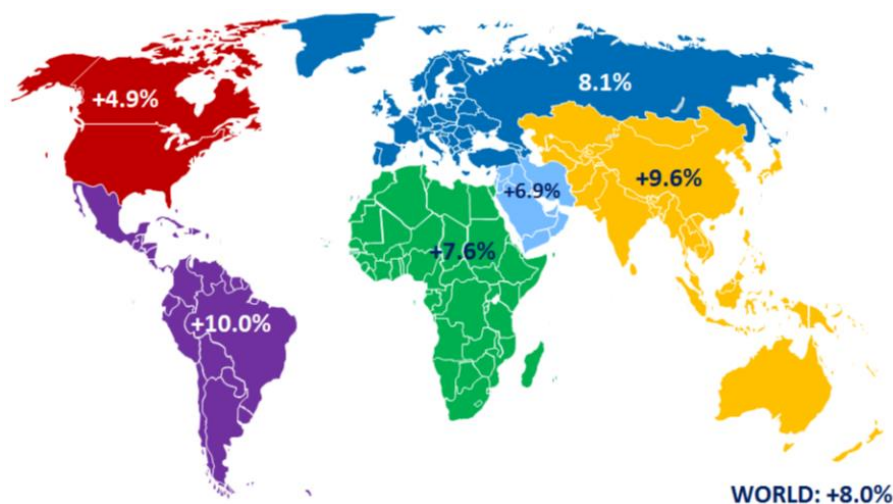


Figura 1.1 - Crescimento do tráfego internacional de passageiros (expressos em RPK) em 2017 (ICAO, 2018)

Todas as regiões registaram um crescimento mais forte do que no ano anterior, com exceção do Médio Oriente, que sofreu um decréscimo significativo, passando dos 11.8% observados em 2016 para 6.9% em 2017. Ainda pela análise da Figura 1.1, a Ásia/Pacífico cresceu 9.6%, sendo o segundo maior crescimento entre todas as regiões, e a América do Norte apresenta um crescimento na ordem dos 4.9%, uma melhoria em relação ao ano de 2016, no entanto, continua a ser a região de crescimento mais lento. Tanto a América Latina e Caribe como a África

também cresceram em 2017 no que respeita ao tráfego aéreo internacional, obtendo os valores de 10% e 7.6% respetivamente.

Por fim, segundo Postorino (2010), a importância do transporte aéreo está amplamente estabelecida não só como um meio de percorrer rapidamente longas distâncias, mas também como um impulsionador do crescimento económico regional.

1.2 Objetivos

O principal objetivo deste estudo consiste em determinar quais as regiões NUTS 2 (*Nomenclature of Territorial Units for Statistics*) do Sudoeste Europeu que mais eficientemente estão a utilizar o transporte aéreo para o seu desenvolvimento, e quais os fatores que determinam essa eficiência. Para tal, complementarmente, constituem objetivos da presente dissertação:

- revisão bibliográfica dos trabalhos disponíveis na literatura sobre o assunto, com o principal propósito de efetuar o levantamento das técnicas que têm sido empregues para estimar o impacto do transporte aéreo no desenvolvimento regional (evitando o denominado problema da endogeneidade – é o transporte aéreo que causa o desenvolvimento regional ou o desenvolvimento regional que fomenta o transporte aéreo?);
- análise e comparação de modelos de regressão linear múltipla que expliquem a relação entre o transporte aéreo e o desenvolvimento regional, com a finalidade de determinar o que melhor se ajusta tendo em conta o objetivo principal;
- análise de resíduos para verificação do modelo de regressão escolhido, com o intuito de confirmar se, efetivamente, de entre todos os modelos estudados, é o mais adequado.

1.3 Estrutura da dissertação

A presente dissertação encontra-se organizada em seis capítulos conforme o que se apresenta de seguida.

No primeiro capítulo é feito o enquadramento do tema onde se abordam aspetos gerais relativos às infraestruturas aeroportuárias, alguns fatores que promovem o aumento do transporte aéreo, que por sua vez possibilitam o desenvolvimento regional, e são ainda apresentados alguns dados

relativos à evolução e crescimento do tráfego aéreo internacional neste último ano, disponibilizados pela ICAO. Os principais objetivos da dissertação são também apresentados neste primeiro capítulo.

No segundo capítulo começam por se apresentar os tipos de métodos de análise de regressão utilizados na presente dissertação para estimar os valores dos coeficientes do modelo de regressão que explica a relação entre o transporte aéreo e o desenvolvimento regional. São também definidos os pressupostos necessários a adotar para a aplicação das metodologias, nomeadamente *ordinary least squares* (OLS) e *two stage least squares* (2SLS). Por fim, descreve-se a metodologia aplicada no estudo, onde se define de que maneira os resultados obtidos numa fase posterior serão analisados, indicando ainda o *software* utilizado para a obtenção dos resultados.

O terceiro capítulo é fundamental para uma melhor compreensão do trabalho desenvolvido. É dedicado à revisão da literatura, e nele são abordados diversos estudos e as metodologias mais utilizadas no estudo do impacto do transporte aéreo no desenvolvimento regional, indicando também as principais variáveis utilizadas pelos autores para caracterizarem o transporte aéreo e o desenvolvimento regional. Apresenta-se ainda um quadro resumo com os principais autores que já desenvolveram este tema, bem como as principais conclusões a que chegaram.

No quarto capítulo dedicado ao estudo de caso, define-se a área de estudo, que corresponde ao Sudoeste Europeu – as regiões NUTS 2 de Espanha, França, Itália e Portugal. São ainda apresentadas e definidas as variáveis escolhidas para caracterizar o transporte aéreo e o desenvolvimento regional, ou seja, as variáveis explicativas e a variável dependente, respetivamente.

O quinto capítulo consiste na especificação dos tipos de modelos de regressão linear múltipla utilizados, nomeadamente do modelo linear e do modelo log-log. Após a análise dos resultados obtidos para cada modelo, escolheu-se o mais adequado para explicar a influência do transporte aéreo no desenvolvimento regional. De seguida, efetua-se uma análise de resíduos que consiste na verificação dos três pressupostos relativos aos resíduos, isto é, as suposições de que os resíduos seguem uma distribuição normal com média zero, são independentes e apresentam variância constante, com o intuito de confirmar se o modelo escolhido é efetivamente adequado. Depois de analisados os resultados obtidos, estes comparam-se com os obtidos em estudos anteriores. No fim, determinam-se quais as regiões que mais eficientemente estão a utilizar o transporte aéreo para o seu desenvolvimento.

No último capítulo são apresentadas as principais conclusões retiradas acerca da influência do transporte aéreo no desenvolvimento regional, e propõem-se alguns estudos que poderão ser desenvolvidos no futuro.

2 METODOLOGIA DE ANÁLISE

Neste capítulo são apresentados os dois métodos utilizados para estimar os coeficientes do modelo de regressão escolhido para estudar a influência do transporte aéreo no desenvolvimento regional. A escolha de um modelo de regressão adequado é fundamental para ser possível cumprir os objetivos definidos na secção 1.2.

2.1 *Ordinary Least Squares (OLS)*

De modo a expressar a relação entre duas ou mais variáveis, é importante estabelecer um modelo matemático. Este tipo de modelo é chamado de regressão, e ajuda a entender como determinadas variáveis influenciam outra variável, ou seja, verifica como o comportamento de uma (ou mais) variável (ou variáveis) pode mudar o comportamento de outra.

Os modelos aplicados na presente dissertação são modelos de regressão linear múltipla, uma vez que se admite que o valor da variável dependente é função linear de duas ou mais variáveis explicativas, cujos parâmetros são estimados por OLS e 2SLS.

Considere-se o seguinte modelo linear de regressão múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u \quad (1)$$

onde β_0 é o termo independente, β_1 é o parâmetro associado à variável x_1 , β_2 é o parâmetro associado à variável x_2 , e assim por diante, para todas as k variáveis independentes (ou explicativas). As variáveis $x_1, x_2, x_3, \dots, x_k$ são as variáveis explicativas, e a variável u denomina-se o termo de erro.

Utilizando a notação matricial, supondo n observações ($n > p$) da variável dependente e das p variáveis independentes, o modelo linear de regressão múltipla pode ser escrito da seguinte forma:

$$[Y] = [X][\beta] + [u] \quad (2)$$

$$[\mathbf{Y}] = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, [\mathbf{X}] = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, [\boldsymbol{\beta}] = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_p \end{bmatrix} \text{ e } [\mathbf{u}] = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} \quad (3)$$

sendo:

$[\mathbf{Y}]$ – Vector $n \times 1$ cujos componentes correspondem às n variáveis dependentes;

$[\mathbf{X}]$ – Matriz de dimensão $n \times (p + 1)$ das variáveis independentes, também denominada matriz do modelo;

$[\boldsymbol{\beta}]$ – Vector $(p + 1) \times 1$ cujos elementos são os coeficientes de regressão;

$[\mathbf{u}]$ – Vector $n \times 1$ cujos elementos são os erros aleatórios (não observáveis).

Ao estabelecer o modelo de regressão linear múltipla, pressupõe-se que:

- i. A variável dependente (y) é função linear das variáveis explicativas (x_1, \dots, x_k);
- ii. Os valores das variáveis explicativas são fixos;
- iii. Dado qualquer valor para as variáveis explicativas, o termo de erro tem um valor esperado de zero, isto é, $E(u | x_1, \dots, x_k) = 0$;
- iv. Os erros são homocedásticos, o que significa que o termo de erro tem variância constante para qualquer valor de x_k , isto é, $Var(u | x_1, \dots, x_k) = \sigma^2$;
- v. Os erros não estão correlacionados entre si;
- vi. Os erros apresentam uma distribuição normal.

Relativamente às pressuposições descritas acima, as três primeiras são necessárias para demonstrar que os estimadores OLS são não-enviesados, e o conjunto dos cinco primeiros pressupostos definem o teorema de Gauss-Markov, que demonstra que os estimadores OLS são lineares não-enviesados de variância mínima. O último pressuposto é necessário para se realizar testes de hipótese (como por exemplo o teste t ou teste F) e para construir intervalos de confiança para os parâmetros.

Esta análise de regressão irá ser aplicada ao estudo de caso (capítulo 4), apenas se se verificar que a metodologia 2SLS é dispensável.

2.2 Two Stage Least Squares (2SLS)

A estimativa pelo método OLS da equação (1) geralmente torna inconsistentes e viesados todos os estimadores β_j , se $Cov(x_k, u) \neq 0$, isto é, se a covariância entre as variáveis explicativas e o termo de erro for diferente de zero, significa que as duas variáveis estão

correlacionadas, o que leva a dizer que a variável explicativa x_k é endógena pois se encontra correlacionada com u (Murteira e Castro, 2018). Isto define o princípio da endogeneidade.

O método apresentado nesta secção é um método de estimação que fornece uma solução geral do problema quando se tem uma ou mais variáveis explicativas endógenas, ou seja, permite estimar de forma consistente os parâmetros de um modelo linear.

2.2.1 Motivação para a avaliação de variáveis instrumentais

Considere-se, novamente, o modelo linear de regressão múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u \quad (4)$$

$$E(u) = 0 \quad (5)$$

$$Cov(x_j, u) = 0, \quad j = 1, 2, \dots, k - 1 \quad (6)$$

Onde x_k pode estar correlacionado com u , isto é, as variáveis explicativas x_1, x_2, \dots, x_{k-1} são exógenas mas x_k é potencialmente endógena na equação (2). Por outras palavras, admite-se que u contém uma variável omitida que não está correlacionada com todas as variáveis explicativas, excepto x_k (Wooldridge, 2012).

Na equação (4), $\beta_0, \beta_1, \dots, \beta_k$ são os chamados coeficientes de regressão e a variável u é o termo de erro, como já descrito para OLS.

Para aplicação deste método, é necessário introduzir, pelo menos, uma nova variável observada, z , que satisfaça duas condições (Wooldridge, 2010):

- i. z não deve estar correlacionada com o termo u :

$$Cov(z, u) = 0 \quad (7)$$

Isto é, tal como x_1, \dots, x_{k-1} são exógenas na equação (4), também a variável z é.

- ii. z está correlacionada com x_k :

$$Cov(z, x_k) \neq 0 \quad (8)$$

Quando z satisfaz as duas condições, diz-se que z é uma variável instrumental para x_k , ou simplesmente que z é um instrumento para x_k .

Wooldridge (2012) descreveu que o requisito de que o instrumento z satisfaça a condição imposta pela equação (7) é resumido dizendo que “ z é exógeno na equação (4)” e, por esse motivo, a equação (7) muitas vezes é definida como a exogeneidade do instrumento. Neste contexto, significa que z não deve ter um efeito parcial em y (após x_k e as variáveis omitidas terem sido controladas), e z não deve estar correlacionada com as variáveis omitidas. Wooldridge (2012) refere ainda que a equação (8) é definida como a relevância do instrumento (“ z é relevante para explicar a variação em x_k ”), o que significa que z deve estar correlacionada, de forma positiva ou negativa, com a variável explicativa endógena, x_k .

Ao estimar a variável instrumental, vários erros podem surgir, principalmente se z e x_1 forem fracamente correlacionados, ainda que a variável instrumental seja consistente quando z e u não estiverem correlacionados e z e x_1 tenham algum tipo de correlação. A fraca correlação entre z e x_1 pode levar ainda a que o estimador da variável instrumental se torne enviesado, mesmo que z e u estejam apenas moderadamente correlacionados. Quando z e x_1 não têm qualquer tipo de correlação, a estimativa não será boa, esteja z correlacionada ou não com u .

2.2.2 Estimativa de variáveis instrumentais num modelo de regressão múltipla

Considerando agora um modelo linear com duas variáveis explicativas, mas onde apenas uma dessas variáveis se encontra correlacionada com o erro:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1 \quad (9)$$

A equação acima é a chamada equação estrutural, cujo interesse se foca no cálculo dos parâmetros β_j .

Nesta secção é aplicada uma nova notação de modo a distinguir as variáveis endógenas das exógenas. Assim:

- y_1 é a variável dependente e é endógena, já que está correlacionada com o termo u_1 ;
- y_2 e z_1 são as variáveis explicativas;
- u_1 é o termo de erro.

Como anteriormente (e análogo à equação (5)), assume-se que o valor esperado do erro é zero, e por isso: $E(u_1) = 0$. Utiliza-se z_1 para indicar que esta variável é exógena na equação (9),

significando que z_1 não está correlacionada com u_1 ; e y_2 para indicar que se suspeita que esta variável está correlacionada com u_1 , assumindo-se que u_1 contém uma variável omitida correlacionada com y_2 .

Sabe-se que se a equação (9) for estimada por OLS, todos os estimadores serão enviesados e inconsistentes (Wooldridge, 2012).

Uma vez que z_1 aparece como variável explicativa na equação (9), não serve de variável instrumental para y_2 , é necessário acrescentar outra variável exógena – z_2 – que não faça parte da equação (9).

Então, os pressupostos a adotar são que z_1 e z_2 não estão correlacionados com u_1 (pois são variáveis exógenas), bem como que o valor esperado para u_1 é zero, o que se pode definir como:

$$E(u_1) = 0, Cov(z_1, u_1) = 0 \text{ e } Cov(z_2, u_1) = 0 \quad (10)$$

Geralmente, quando num modelo de regressão temos mais do que uma variável explicativa endógena, a identificação pode falhar de várias formas possíveis, e por isso Wooldridge (2012), indica uma condição necessária para a identificação de uma equação. Este problema da identificação refere-se ao facto das estimativas dos parâmetros β_j da equação estrutural (equação (9)) poderem ou não ser obtidos a partir dos coeficientes estimados na forma reduzida (isto é, quando a variável endógena é expressa através das variáveis explicativas exógenas, incluindo a variável instrumental, e do termo de erro). Se os parâmetros β_j forem obtidos através dos coeficientes estimados na forma reduzida, a equação é identificada. Esta condição de identificação de uma equação denomina-se *order condition* e refere que são necessárias, pelo menos, tantas variáveis exógenas não presentes na equação quanto as variáveis explicativas endógenas incluídas na equação estrutural. Esta condição é fácil de verificar pois envolve apenas a contagem de variáveis endógenas e exógenas. A condição suficiente para a identificação é chamada de *rank condition*.

De acordo com Wooldridge (2012), apresentam-se de seguida os pressupostos sob os quais se deve aplicar 2SLS:

- i. O modelo é linear nos parâmetros β_j . O modelo é o mesmo que o utilizado para estimar os parâmetros β_j pelo OLS. Na verdade, podem aplicar-se diferentes métodos de estimação ao mesmo modelo, e os métodos OLS e 2SLS são exemplo disso;
- ii. Tem-se uma amostra aleatória em y , x_j e z_j ;

- iii. Verifica-se a *Rank condition*. Com apenas uma variável explicativa endógena, como na equação estrutural

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \dots + \beta_k z_{k-1} + u_1 \quad (11)$$

a *rank condition* é facilmente definida. Sendo z_1, \dots, z_m as variáveis exógenas, onde z_k, \dots, z_m não aparecem no modelo estrutural definido pela equação (11). A forma reduzida para y_2 é:

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \dots + \pi_{k-1} z_{k-1} + \pi_k z_k + \dots + \pi_m z_m + v_2 \quad (12)$$

Depois, é necessário que pelo menos uma variável das π_k, \dots, π_m seja diferente de zero. Isto exige pelo menos uma variável exógena que não apareça em (11), pela *order condition*;

- iv. O termo de erro u tem média zero e cada variável instrumental não está correlacionada com u . Lembra-se que qualquer x_j que não esteja correlacionada com u também actua como variável instrumental;
- v. Seja Z o conjunto de todas as variáveis instrumentais, então $E(u^2 | Z) = \sigma^2$, o que equivale dizer que $Var(u | Z) = \sigma^2$ – homocedasticidade.

Segundo os quatro primeiros pressupostos, o estimador 2SLS considera-se assintoticamente normalmente distribuído e pode afirmar-se que é consistente (Wooldridge, 2012).

De acordo com os cinco pressupostos dados, admite-se que o estimador 2SLS é o melhor estimador de variável instrumental, isto porque é consistente e não-enviesado. Se o pressuposto da homocedasticidade não se mantiver, o estimador 2SLS continua a ser assintoticamente normal mas, em geral, deixa de ser um estimador eficiente de variável instrumental (Wooldridge, 2012).

2.3 Metodologia aplicada ao estudo

Relativamente à metodologia aplicada, esta engloba os dois métodos de análise de regressão especificados nas secções 2.1 e 2.2. Para isso, recorre-se ao *R Statistical Software*, o qual permite realizar estas duas regressões, com o intuito de obter coeficientes e alguns parâmetros, úteis para a compreensão e análise dos resultados obtidos.

Em primeiro lugar aplica-se 2SLS, no qual se irá verificar se realmente é necessário a utilização deste método de análise. Com os resultados da regressão 2SLS obtidos pelo *software R*, é possível quantificar a influência que cada variável explicativa assume, relativamente à variável dependente. Esta influência é quantificada pelo valor referente aos coeficientes de regressão β_j , quando se trata de um modelo linear, ou interpretada em termos de elasticidade, quando se trata

de um modelo de Cobb-Douglas. A utilização deste tipo de modelos irá ser explicada no capítulo 5. Outros parâmetros importantes que se obtêm dos resultados da regressão são o valor da estatística p e da estatística t . Estes serão fundamentais para, em conjunto com os coeficientes β_j (ou em termos de elasticidade), explicar a contribuição de cada variável explicativa na variável dependente. Relativamente a estes parâmetros, quanto maior for a estatística t menor será a estatística p . Adotando um nível de significância de 5% ($\alpha = 0.05$), o que se pretende é que a estatística p tome valores inferiores a 0.05, o que por sua vez equivale dizer que se deseja que a estatística t apresente valores superiores a 2. Assim, nas variáveis em que $|estatística\ t| \geq 2$, isto significa que essas variáveis irão contribuir de forma mais acentuada no crescimento da variável dependente do que as restantes, sendo que a percentagem com que influenciam o VAB (Valor Acrescentado Bruto) se define pelo valor da estimativa dos coeficientes β_j (ou das elasticidades).

De seguida, analisa-se o valor do teste de endogeneidade de Wu-Hausman (ou Durbin-Wu-Hausman), um teste referente à variável explicativa endógena, que revela a presença de endogeneidade da variável. O estimador 2SLS torna-se menos eficiente do que o de OLS quando as variáveis explicativas são exógenas, daí ser útil a realização deste teste. É importante referir que este teste compara apenas o estimador da variável endógena. Se houver uma pequena diferença entre o estimador OLS e o estimador 2SLS, não há a necessidade de utilizar variáveis instrumentais, e pode-se concluir que a variável é exógena. Se, pelo contrário, existir uma grande diferença, é necessário o procedimento introduzindo variáveis instrumentais de modo a produzir coeficientes consistentes, pois a variável explicativa é endógena.

O teste de Wu-Hausman é feito automaticamente no *software R* quando se aplica a regressão 2SLS e, na prática, o que irá ser avaliado é a estatística p que resulta do teste. Assim, e admitindo um nível de significância de 5%, nos casos em que a estatística p for superior a 0.05, pode afirmar-se com 95% de confiança que o teste não é significativo estatisticamente, não havendo razões para se pensar que existe endogeneidade (provocada pela relação entre a variável explicativa que se suspeitava endógena e a variável dependente), sendo, portanto, dispensável a utilização de 2SLS, bastando apenas aplicar a análise de regressão OLS. Nos outros casos, quando o valor da estatística p for inferior a 0.05, confirma-se a existência de endogeneidade, sendo imprescindível a aplicação de 2SLS, a qual contempla a introdução de uma variável instrumental que serve de instrumento para a variável explicativa endógena.

De seguida efetua-se uma análise de resíduos que consiste na verificação, principalmente, das três suposições relativas aos resíduos, isto é, as suposições de que os resíduos seguem uma distribuição normal com média zero, são independentes e apresentam variância constante. Isto

é verificado por meio da representação gráfica dos resíduos. Se os resíduos, após esta análise, não forem contra os pressupostos admitidos, considera-se que o modelo escolhido é o mais adequado.

Por fim, de modo a determinar quais as regiões que mais eficientemente estão a utilizar o transporte aéreo para o seu desenvolvimento, comparam-se os valores reais de VAB com os valores modelados de VAB através do modelo de regressão escolhido. Com isso, e uma vez que se espera que o modelo escolhido se aproxime do que realmente acontece na realidade, poderá observar-se quais são as regiões que utilizam o transporte aéreo a favor do desenvolvimento regional, isto é, as regiões que estão a ter um desenvolvimento melhor do que o esperado, através do valor dos resíduos positivos, bem como as que se encontram abaixo do previsto.

3 REVISÃO DA LITERATURA

Este capítulo centra-se na recolha de informação sobre os principais métodos de estimação dos coeficientes de regressão aplicados por diversos autores, bem como na definição das variáveis utilizadas para avaliar o impacto do transporte aéreo no desenvolvimento regional.

Muitos estudos foram realizados sobre este tema para os Estados Unidos da América (EUA), mas o estudo sobre outros lugares é escasso. Segundo Dobruszkes *et al.* (2011), a principal razão para essa falta de estudos é a dificuldade em encontrar dados homogéneos para um contexto multinacional. Além disso, nos EUA, a maioria dos aeroportos serve apenas uma cidade, com opções limitadas de transporte terrestre, enquanto, por exemplo, os aeroportos europeus estão rodeados por várias cidades, criando assim várias áreas de captação sobrepostas.

No contexto europeu, os estudos limitam-se ao de Mukkala e Tervo (2013) que avaliam a causalidade entre o tráfego aéreo e o desenvolvimento regional, representado pelo crescimento do PIB e do emprego, em 86 regiões da Europa Ocidental, e o estudo de Percoco (2010) sobre o impacto do transporte aéreo de passageiros sobre o emprego local em 103 províncias italianas.

Até à data, a avaliação do impacto dos aeroportos tem sido efetuada essencialmente de três formas, recorrendo-se a:

- Técnicas de levantamento, inquéritos (avaliação quantitativa subjetiva);
- Análise através de multiplicadores;
- Modelos econométricos (OLS, 2SLS).

Neste capítulo, apenas se apresentam os estudos que utilizam OLS e 2SLS como método de estimação dos coeficientes de regressão.

Goetz (1992) foi o pioneiro no estudo deste tema, apresentando o primeiro trabalho sobre a forma como a relação entre o fluxo de passageiros *per capita* e o crescimento da população urbana e do emprego variou ao longo do tempo.

Para cumprir o objetivo proposto, avaliou o desenvolvimento regional com base no número de passageiros *per capita* e por população total empregada, e o transporte aéreo através do

crescimento da população e do emprego. Em seguida, aplicou a metodologia de análise de regressão OLS e identificou e analisou os *outliers* positivos e negativos, concluindo que existe uma relação positiva entre a utilização do transporte aéreo e o desenvolvimento regional.

Bennel e Prentice (1993) também se dedicaram à análise desta relação. Após o estudo de impacto de 38 aeroportos, caracterizaram o desenvolvimento regional utilizando o produto anual e a relação entre pessoas e anos de emprego; e, para o transporte aéreo, centraram-se em características do próprio aeroporto.

Relativamente ao método escolhido, foi aplicado ao modelo de regressão a metodologia OLS, o que permitiu apoiar a relação positiva apresentada por Goetz.

Além disso, verificou-se que, de entre as variáveis que classificavam o tráfego aéreo, as mais significativas diziam respeito ao número de passageiros que embarcam e desembarcam no aeroporto, a riqueza da região, os movimentos das aeronaves de grandes dimensões e, por fim, a existência de uma base de manutenção da transportadora aérea (Bennel e Prentice, 1993).

Para aprofundar o conhecimento da relação entre o transporte aéreo e o desenvolvimento regional, Button e Taylor (2000) testaram a hipótese da economia de uma região ser favorecida através da introdução de serviços internacionais.

Estes serviços são definidos sobretudo em termos de destinos europeus servidos por voos diretos, porém foram indicadores suplementares para a qualidade do serviço e os efeitos da rede. Existem vantagens em possuir uma vasta gama destes serviços adicionais, todavia Button e Taylor (2000) partiram do pressuposto de que esta vantagem é anulada quando o número de destinos a servir é elevado.

Como variável dependente, representativa do desenvolvimento regional, foi escolhido o emprego, definido como *new economy employment*. Este conceito não é simples, por isso, neste estudo, as indústrias que compõem os diversos sectores de atividade que caracterizam o emprego, foram definidas de acordo com o Armington *index* (Button and Taylor, 2000). Relativamente ao tráfego aéreo, optou-se pelas seguintes variáveis independentes: (i) população das áreas metropolitanas em 1996, (ii) número de passageiros europeus em 1994, (iii) número de aeroportos europeus servidos em 1994, (iv) despesas militares em 1996, (v) fuso horário, (vi) número total de embarques em 1994. A variável relacionada com as despesas militares em 1996 foi escolhida uma vez que fatores não relacionados diretamente com o transporte aéreo também podem influenciar a economia local. Neste caso, o facto de existir atividade militar

numa determinada região, vai contribuir para o desenvolvimento desta, através da geração de produto (Button e Taylor, 2000). Quanto à variável referente ao fuso horário, esta foi introduzida pelo facto de a influência do serviço aéreo europeu ser suscetível de diminuir na zona oeste de onde a cidade está localizada, por isso surge como uma tentativa de capturar este efeito de distância, incluindo um fuso horário variável, tomando o valor zero para a zona este. Para ser capaz de isolar o efeito do transporte aéreo são ainda incorporadas as restantes variáveis, pois já foram consideradas importantes noutros estudos de desenvolvimento económico.

O método escolhido foi aplicado a 41 aeroportos dos EUA, e consistiu numa regressão estimada por OLS, método já utilizado anteriormente por Bennel e Prentice, e também por Goetz.

No final deste estudo, concluiu-se que, nos EUA, para um elevado número de aeroportos, existe uma relação entre a economia das zonas circundantes e a disponibilidade dos serviços aéreos internacionais para a União Europeia (UE). Esta relação é evidenciada pelo facto de as áreas que usufruíram da introdução desses serviços terem atraído, mantido ou gerado mais emprego (*new economy employment*) do que as restantes. Destaca-se ainda que o mais importante para este crescimento é o número de destinos internacionais servidos e a qualidade dos mesmos (Button e Taylor, 2000).

Um dos autores mais referenciados dentro desta temática, Brueckner (2003), dedicou-se ao estudo em 91 regiões metropolitanas dos EUA. Neste trabalho, Brueckner tentou resolver o problema da endogeneidade associado ao estudo do tráfego aéreo, algo que, até à data, tinha limitado de uma forma significativa, mas tinha sido considerado, nos estudos efetuados.

O método adotado consiste numa equação estimada pelo método 2SLS. Este método é uma extensão do método OLS e aplica-se quando os termos de erro estão correlacionados com as variáveis independentes. Na primeira etapa, o método encontra as parcelas das variáveis dependentes e independentes que podem ser atribuídas às variáveis instrumentais. Esta etapa envolve estimar uma regressão de OLS para cada variável do modelo no conjunto das variáveis instrumentais. A segunda etapa é uma regressão da equação original, com todas as variáveis substituídas pelos valores ajustados das regressões da primeira etapa.

Segundo Brueckner, a equação estimada toma a forma:

$$E = f(T, X; \theta) + u \quad (13)$$

onde X é um vetor de variáveis endógenas, dependentes, que influenciam o emprego, θ é um vetor de parâmetros e u é o termo de erro.

A relação estabelecida na equação (13) pode ser caracterizada como uma relação de "forma quase reduzida" (*quasi reduced-form*) na medida em que, além do tráfego aéreo, todas as variáveis potencialmente endógenas que ajudam a determinar o emprego foram eliminadas por substituição adequada de outras equações estruturais. Por exemplo, embora o nível salarial possa ser visto como um fator determinante do emprego, o salário é endógeno, sendo determinado em conjunto com E , e, portanto, é excluído da equação (13). Somente as variáveis que podem razoavelmente ser vistas como determinantes exógenos do emprego aparecem em X .

Numa equação estrutural separada que determina o tráfego aéreo, T depende de E , de outras variáveis endógenas, como os rendimentos e, portanto, os salários, e de um conjunto de variáveis exógenas (Brueckner, 2003).

Para que a equação (13) seja identificada, permitindo que o efeito de T sobre o emprego seja medido, o conjunto de variáveis exógenas deve incluir pelo menos uma variável instrumental (ou instrumento) que ainda não aparece como parte de X . Grande parte do problema é encontrar essas variáveis, que devem ser altamente correlacionados com T , mas não correlacionados com o termo de erro u . Uma vez selecionados os instrumentos, a equação (13) pode finalmente ser estimada pelo método 2SLS.

O desenvolvimento regional, caracterizado pela variável dependente E , é medido por o emprego não agrícola (EMP) em cada uma das 91 áreas metropolitanas consideradas na análise (no ano de 1996). Noutras duas equações, utiliza as parcelas do emprego relacionadas com a produção, construção e exploração mineira (GDSEMP) e relacionadas com os serviços (SVCEMP), em substituição da variável EMP. O tráfego aéreo, como variável independente, foi caracterizado pelas variáveis instrumentais e exógenas especificadas na Figura 3.1:

Variable	Definition
<i>TRAFFIC</i>	Total 1996 passenger enplanements at metro area airport(s)
<i>EMP</i>	Total metro-area non-farm employment for 1996 (in 1000s)
<i>GDSEMP</i>	Total metro-area goods-related employment for 1996 (in 1000s)
<i>SVCEMP</i>	Total metro-area service-related employment for 1996 (in 1000s)
<i>POP</i>	Metro-area population for 1990
<i>YOUNG</i>	Percentage of 1996 metro-area population of age 14 or younger
<i>OLD</i>	Percentage of 1996 metro-area population of age 65 or older
<i>RTW</i>	Dummy variable equal to one if metro area's state has right-to-work law
<i>HEATING</i>	Average heating degree days for metro area
<i>COLGGRAD</i>	Percentage of metro area's 1990 population over 25 with a college degree
<i>CORPTAX</i>	Maximum 1996 marginal corporate income tax rate for metro area's state
<i>PERSTAX</i>	Maximum 1996 marginal personal income tax rate for metro area's state
<i>HUB</i>	Enplanement share of any hub airports in metro area
<i>CENTRALITY</i>	Mileage from the metro area's largest airport to the US population centre of gravity
<i>LEISURE</i>	Dummy variable equal to one for Las Vegas and Orlando
<i>PROXIMITY</i>	Dummy variable equal to one for smaller metro areas within 150 miles of large airport
<i>SLOT</i>	Enplanement share of any slot-controlled airports in metro area

Figura 3.1 – Definição das variáveis (Brueckner, 2003)

Os resultados mostraram que o acréscimo de 10% no embarque de passageiros numa área metropolitana leva, aproximadamente, a um aumento de 1% na variável *SVCEMP*. No entanto, não há qualquer efeito na variável *GDSEMP* devido ao tráfego aéreo, concluindo assim que é menos importante e tem menos influência o transporte aéreo para as empresas de produção, construção e exploração mineira do que para as empresas referentes aos serviços (incluem o comércio grossista e retalhista, finanças, seguros, imobiliário, o governo, os transportes e os serviços públicos) (Brueckner, 2003).

Até então, poucos estudos tinham avaliado o impacto dos aeroportos e do tráfego aéreo sobre o desenvolvimento regional, sendo Brueckner a principal exceção. Após o ano de 2003, vários autores, como Green (2007) e Percoco (2010), usaram o trabalho de Brueckner como base e referência para estudar esse impacto.

Green (2007), para tentar resolver a questão da simultaneidade, baseia-se no modelo aplicado por Brueckner, fazendo uso das mesmas variáveis, mas mais completas. Realizou o estudo para 83 áreas metropolitanas dos EUA, englobando os 100 aeroportos de maior dimensão do país. Aplicou a metodologia 2SLS, sendo que a especificação do modelo definida em Green (2007) difere da de Brueckner (2003) em dois aspetos principais: Brueckner (2003) assume que o impacto do tráfego aéreo no desenvolvimento regional é contemporâneo, isto é, considera que o impacto gerado no desenvolvimento de uma região se dá no período para o qual realiza o estudo, enquanto que Green (2007) considera que esse impacto, que se sente a curto prazo,

(referido em Brueckner (2003)) sofre um atraso, ou seja, é fruto dos dados obtidos para anos anteriores.

Com isto, Green (2007), verificou que os embarques de passageiros *per capita* e as origens destes, tinham grande influência no desenvolvimento regional, medido pelo crescimento populacional e do emprego, afirmando que os resultados não são conclusivos, mas são consistentes com os obtidos em Brueckner (2003).

Percoco (2010) surge como o primeiro trabalho focado num país europeu, nomeadamente Itália, onde se concentra no impacto dos aeroportos em 103 províncias italianas. Opta por utilizar todas as províncias italianas, incluindo as que não têm aeroporto, contrariamente a Brueckner (2003) e Green (2007) que utilizaram um conjunto de dados que inclui apenas as áreas metropolitanas com um aeroporto em funcionamento. O facto de nem todas as 103 províncias italianas terem um aeroporto, exige especial cuidado a lidar com o *selection bias* (enviesamento de seleção) devido à escolha de localização dos aeroportos. As decisões de localização do aeroporto são tomadas principalmente a nível central, de acordo com a possível procura e o afastamento geográfico da área e, ainda mais importante, por razões políticas. Consequentemente, na abordagem adotada neste artigo, o *selection bias* é explicitamente considerado de modo que, em princípio, o autor possa obter melhores estimativas da elasticidade do emprego local (variável que caracteriza o desenvolvimento regional) para o tráfego aéreo. Esta abordagem podia afetar fortemente as estimativas, além de possibilitar a identificação de *spatial spillovers* (Percoco, 2010).

O trabalho mais próximo deste é o de Brueckner (2003), e neste estudo Percoco adota o mesmo modelo de regressão e as variáveis idênticas. Contudo, a abordagem difere um pouco, pela razão deste lidar também com o *selection bias*, como já dito anteriormente, pelo facto da equação estimada poder sofrer de um viés de endogeneidade causado tanto pelo facto de as decisões de localização do aeroporto serem muitas vezes tomadas com base no desenvolvimento da província (ou seja, o emprego pode influenciar o tráfego aéreo) e, também pelo facto de, após o aeroporto se encontrar realmente numa região, o tráfego aéreo ainda pode ser função do emprego, porque as áreas mais desenvolvidas tendem a interagir mais com o resto do mundo, provocando o aumento do tráfego aéreo (Percoco, 2010). Para superar esse problema, aplicou a metodologia 2SLS.

Por fim, observou que o valor das elasticidades estimadas para o emprego total e para o emprego no sector dos serviços foram aproximadamente metade quando comparadas com as obtidas em

Brueckner (2003), confirmando ainda uma contribuição significativa do tráfego aéreo de passageiros no emprego relativo ao sector dos serviços.

Apresentam-se de seguida, no Quadro 3.1, as principais conclusões retiradas pelos autores acima referidos.

Quadro 3.1 - Quadro com os principais autores e resultados dos estudos (adaptado de: Van De Vijver *et al*, 2016)

Literatura	Região	Resultados
Goetz (1992)	EUA, 1950-1987	Relação positiva entre passageiros <i>per capita</i> e emprego
Bennel e Prentice (1993)	Canadá, 1988	De entre as variáveis que classificam o tráfego aéreo, as mais significativas diziam respeito ao número de passageiros que embarcam e desembarcam no aeroporto, a riqueza da região, os movimentos das aeronaves de grandes dimensões e, por fim, a existência de uma base de manutenção da transportadora aérea
Button e Taylor (2000)	EUA, 1996 (e 1994)	Relação entre a quantidade e qualidade dos serviços aéreos para a EU e para as indústrias que compõem o conceito de emprego, definido por <i>new economy employment</i>
Brueckner (2003)	EUA, 1996 (e 1990)	O crescimento no embarque de passageiros leva ao aumento do emprego no setor de serviços, mas não no setor da indústria
Green (2007)	EUA, 1990-2000	Os embarques de passageiros <i>per capita</i> e as origens destes têm grande influência no desenvolvimento regional, medido pelo crescimento populacional e do emprego
Percoco (2010)	Províncias italianas, 2002	Influência significativa do transporte aéreo de passageiros no emprego no setor de serviços

Conclui-se, observando o Quadro 3.1 que, de acordo com a maioria dos autores mencionados, existe uma relação entre o transporte aéreo de passageiros e o emprego, isto é, o aumento no embarque de passageiros provoca o crescimento do emprego, essencialmente no setor dos serviços.

4 ESTUDO DE CASO

4.1 Região em estudo

De modo a estabelecer o modelo matemático que vai explicar a influência das variáveis explicativas, dando mais importância à variável relacionada com o transporte aéreo, na variável dependente, que caracteriza o desenvolvimento regional, é necessário definir a área de estudo, bem como todas as variáveis que fazem parte do modelo, para posteriormente ser aplicada a análise de regressão 2SLS.

A área de estudo definida para a aplicação da metodologia apresentada na secção 2.3, apresenta-se na Figura 4.1:



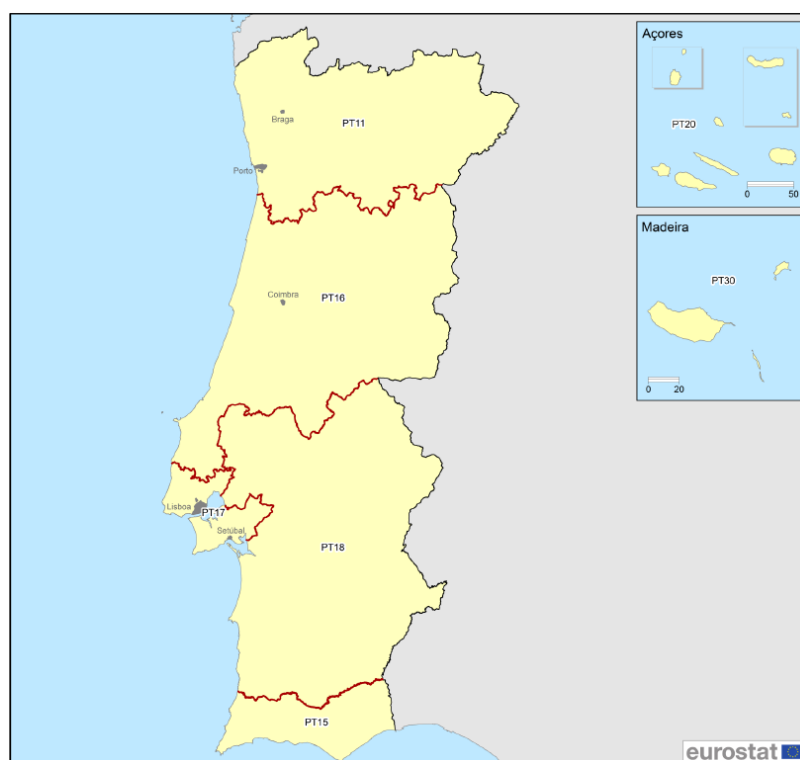
(a)



(b)



(c)



(d)

Figura 4.1 – Regiões NUTS 2 de: (a) Espanha, (b) França, (c) Itália, (d) Portugal (NUTS 2013/EU-28, 2015)

Engloba as regiões do Sudoeste Europeu, nomeadamente as NUTS 2 de Espanha, França, Itália e Portugal e apenas são consideradas as regiões para as quais todos os dados necessários estão disponíveis.

A NUTS (*Nomenclature of Territorial Units for Statistics*) foi elaborada pela Eurostat para fornecer uma discriminação do território económico da UE em unidades territoriais para a produção de estatísticas regionais e para orientar intervenções políticas a nível regional (NUTS 2013/EU-28). A nomenclatura NUTS subdivide o território económico dos Estados Membros em unidades territoriais e atribui a cada unidade territorial uma designação e um código específicos. A nomenclatura NUTS é hierárquica e a sua classificação compreende três níveis. Subdivide cada Estado Membro em unidades territoriais de nível NUTS 1, cada uma das quais é subdividida em unidades territoriais de nível NUTS 2, sendo estas, por sua vez, subdivididas em unidades territoriais de nível NUTS 3. As regiões NUTS 2, de acordo com a Eurostat, são definidas como as regiões de base para a aplicação das políticas regionais da UE relativas à

criação de emprego, à competitividade, ao crescimento económico, à melhoria da qualidade de vida e ao desenvolvimento sustentável (NUTS 2013/EU-28, 2015).

Na presente dissertação, das 66 regiões NUTS 2 do Sudoeste Europeu – Espanha, França, Itália e Portugal – apenas foram consideradas 52 regiões neste estudo.

Foram excluídas, pela falta de dados relativos ao transporte aéreo, nomeadamente o número total de movimentos de aeronaves e movimentos de passageiros, quatro regiões pertencentes a França e uma pertencente a Itália, ou seja, foram retiradas as seguintes regiões:

- Haute-Normandie (FR23);
- Basse-Normandie (FR25);
- Bourgogne (FR26);
- Franche-Comté (FR43);
- Provincia Autonoma di Bolzano/Bozen (ITH1).

Foram ainda retiradas as regiões que não originam tráfego aéreo nem fluxo de passageiros, pela ausência de aeroporto, nomeadamente:

- La Rioja (ES23);
- Castilla-la Mancha (ES42);
- Extremadura (ES43);
- Valle d'Aosta (ITC2);
- Molise (ITF2);
- Basilicata (ITF5);
- Provincia Autonoma di Trento (ITH2);
- Centro (PT) (PT16);
- Alentejo (PT18).

Consideraram-se neste estudo 14 regiões em Espanha, 17 regiões em França, 16 regiões em Itália e 5 regiões em Portugal, como se apresenta no Quadro 4.1.

Quadro 4.1 – Regiões NUTS 2 consideradas no estudo

ESPANHA		FRANÇA		ITÁLIA		PORTUGAL	
ES11	Galícia	FR10	Île de France	ITC1	Piemonte	PT11	Norte
ES12	Principado de Asturias	FR21	Champagne-Ardenne	ITC3	Liguria	PT15	Algarve
ES13	Cantabria	FR22	Picardie	ITC4	Lombardia	PT17	Área Metropolitana de Lisboa
ES21	País Vasco	FR24	Centre (FR)	ITF1	Abruzzo	PT20	Região Autónoma dos Açores
ES22	Comunidad Foral de Navarra	FR30	Nord - Pas-de-Calais	ITF3	Campania	PT30	Região Autónoma da Madeira
ES24	Aragón	FR41	Lorraine	ITF4	Puglia		
ES30	Comunidad de Madrid	FR42	Alsace	ITF6	Calabria		
ES41	Castilla y León	FR51	Pays de la Loire	ITG1	Sicilia		
ES51	Cataluña	FR52	Bretagne	ITG2	Sardegna		
ES52	Comunidad Valenciana	FR53	Poitou-Charentes	ITH3	Veneto		
ES53	Illes Balears	FR61	Aquitaine	ITH4	Friuli-Venezia Giulia		
ES61	Andalucía	FR62	Midi-Pyrénées	ITH5	Emilia-Romagna		
ES62	Región de Murcia	FR63	Limousin	ITI1	Toscana		
ES70	Canarias (ES)	FR71	Rhône-Alpes	ITI2	Umbria		
		FR72	Auvergne	ITI3	Marche		
		FR81	Languedoc-Roussillon	ITI4	Lazio		
		FR82	Provence-Alpes-Côte d'Azur				

4.2 Definição de variáveis e recolha de dados

Tendo por base o trabalho de Brueckner (2003) e Percoco (2010), as variáveis escolhidas para definir o transporte aéreo e o desenvolvimento regional aparecem descritas no Quadro 4.2.

Quadro 4.2 – Definição das variáveis

Variáveis	Definição	Unidades
VAB	Valor Acrescentado Bruto	10 ⁶ €
MAER	Nº total de movimentos (chegadas + partidas) de aeronaves	10 ³ voos
MPAX	Nº total de movimentos de passageiros	10 ⁶ hab
PACT	População economicamente ativa	10 ³ hab
PMED	População dos 25 aos 64 com o ensino secundário ou superior	10 ³ hab
PSEC	População empregada (25-64 anos) no sector secundário	10 ³ hab
PTER	População empregada (25-64 anos) no sector terciário	10 ³ hab
PDES	População desempregada	10 ³ hab
DORM	Nº de dormidas em hotéis e outros estabelecimentos turísticos	10 ³ noites

Todos os dados relativos a estas variáveis foram retirados da base de dados da Eurostat (<http://ec.europa.eu/eurostat/data/database>), referentes ao ano de 2015. Encontra-se no Anexo A um quadro com o valor das variáveis utilizadas e respetivas unidades.

Relativamente ao Quadro 4.2, e de um modo mais detalhado, a variável VAB é definida sendo o Valor Acrescentado Bruto em cada região, e é usualmente utilizado para medir a produção de todos os bens (produtos e serviços) de uma economia durante um determinado período, constituindo por isso um dos indicadores mais utilizados na atividade económica. Por essa razão, o VAB será a variável que caracteriza o desenvolvimento regional.

Quanto às variáveis MAER e MPAX, estas referem-se ao número total de movimento de aeronaves e movimento de passageiros, respetivamente. Para a determinação destas variáveis não se consideraram as regiões que não produzem fluxo de aeronaves nem de passageiros, como referido anteriormente, pelo facto dessas regiões não possuírem aeroporto.

A variável PACT, classificada como a população (dos 25 aos 64 anos) economicamente ativa, engloba o conjunto das pessoas empregadas e das pessoas que se encontram temporariamente desempregadas, ou seja, é a população que no período de referência, constituía a mão de obra disponível para a produção de bens e serviços (Instituto Nacional de Estatística). Contrariamente, a variável PDES define a população desempregada, e refere-se ao conjunto dos trabalhadores que não tem emprego. Inclui quem já trabalhou e procura um novo emprego e ainda quem procura o primeiro emprego (Eurostat).

Relativamente à variável caracterizadora do nível de escolaridade da população, encontra-se definida no Quadro 4.2 a variável PMED, que expressa a população dos 25 aos 64 anos com pelo menos o ensino médio. De acordo com o *International Standard Classification of*

Education (ISCED), considerou-se que PMED engloba as áreas de educação 3 a 8, definidas pelo ISCED 2011, e que se apresentam de seguida no Quadro 4.3:

Quadro 4.3 – Áreas de educação definidas em ISCED (2011)

Áreas de educação	
3	Ciências sociais, negócios e direito
4	Ciência
5	Engenharia, indústria e construção
6	Agricultura
7	Saúde e bem-estar
8	Serviços

No que se refere às variáveis que abrangem os sectores do emprego, visto que na base de dados utilizada os dados não estão divididos segundo os sectores de emprego como os conhecemos, mas sim segundo uma nomenclatura específica que apenas faz uma divisão de todas as atividades económicas existentes na UE, a NACE (*Nomenclature statistique des Activités économiques dans la Communauté Européenne*), houve a necessidade de aglomerar as atividades que pertencem ao sector secundário e ao terciário, de modo a poderem definir-se as variáveis PSEC e PTER, conforme o Quadro 4.2. Deste modo, para determinar a população empregada no sector secundário aglomeraram-se todas as atividades pertencentes à NACE que se relacionam com a indústria, e para definir a população empregada no sector terciário aglomeraram-se as que se relacionam com o comércio e os serviços, como se mostra no Quadro 4.4:

Quadro 4.4 – Atividades económicas segundo a NACE

Sector secundário	Sector terciário
Indústria transformadora	Comércio por grosso e a retalho, transportes, hotelaria e restauração
Construção	Informação e comunicação
	Atividades financeiras e de seguros
	Atividades imobiliárias
	Atividades profissionais, científicas e técnicas; atividades de serviços administrativos e de manutenção
	Administração pública, defesa, educação, saúde humana e ação social
	Artes, entretenimento e recreação; outras atividades de serviço; atividades de organizações e organismos domésticos e extraterritoriais

Resta apenas caracterizar a variável DORM que engloba o número de dormidas em hotéis e outros estabelecimentos turísticos, de forma contabilizar a atividade turística existente nas regiões pois esta também influencia o desenvolvimento regional.

Por último, de modo a estudar a influência do transporte aéreo no desenvolvimento das 52 regiões consideradas e para aplicação da metodologia apresentada no capítulo 2, é necessário definir quais são as variáveis que, de acordo com as definidas no Quadro 4.2, caracterizam o transporte aéreo e o desenvolvimento regional. A variável que representa o desenvolvimento de cada região, denominada a variável dependente, é o VAB. Esta será explicada por um conjunto de variáveis explicativas endógenas e exógenas, sendo que as variáveis explicativas endógenas são as que se encontram relacionadas com o transporte aéreo e as variáveis explicativas exógenas são um conjunto de variáveis, neste caso relativas ao emprego e à população, que também influenciam a variável dependente. Define-se ainda uma variável instrumental, DORM, relacionada com o turismo, que está correlacionada com as variáveis potencialmente endógenas do modelo (necessária para aplicação da metodologia 2SLS). Com isto, as variáveis utilizadas são as seguintes:

- Variável dependente: VAB;
- Variáveis explicativas endógenas: MAER e MPAX;
- Variáveis explicativas exógenas: PACT, PMED, PSEC, PTER e PDES;
- Variável instrumental: DORM.

5 RESULTADOS DO ESTUDO

5.1 Especificação do modelo de regressão linear e log-log (Cobb-Douglas)

Nesta secção apresentam-se definidos dois tipos de modelos com o intuito de explicar a relação entre a variável dependente e as variáveis explicativas, de modo a determinar qual a influência do tráfego aéreo no desenvolvimento regional. Para isso, define-se um modelo de regressão linear:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \quad (14)$$

e um modelo de Cobb-Douglas. Num modelo de Cobb-Douglas, na sua forma estocástica, a relação entre as variáveis não é linear, mas aplicando o logaritmo à equação (14), o modelo passa a ser linear nos parâmetros β_j , e, portanto, torna-se num modelo de regressão linear, podendo ser generalizado para um modelo log-log:

$$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x_1) + \dots + \beta_k \ln(x_k) + u \quad (15)$$

É este tipo de modelos lineares e log-log que são estimados por 2SLS (ou OLS nos casos em que 2SLS não se aplica) nas secções 5.1.1 até 5.1.4. Dar-se-á particular atenção aos modelos log-log, por serem diretamente comparáveis com os obtidos em outros trabalhos, tais como Brueckner (2003) e Percoco (2010), onde as variáveis estão expressas em logaritmos e não no seu valor real.

5.1.1 Modelo de regressão linear com seis variáveis explicativas

Neste modelo, como primeira tentativa de explicar a relação entre o transporte aéreo e o desenvolvimento regional, são utilizadas todas as variáveis explicativas presentes no Quadro 4.2. De início foi aplicado o método 2SLS ao problema, como já descrito anteriormente, verificando de seguida se seria necessário a aplicação desta metodologia com a introdução da variável instrumental, ou se esta seria dispensável. Para tal, introduziram-se todos os dados referentes às variáveis no *software R*, definiu-se o modelo e aplicou-se a metodologia, sendo 2SLS neste *software* definida pela função *ivreg()*. Em primeiro lugar utilizou-se a variável

MAER como variável explicativa endógena, de modo a caracterizar a influência do transporte aéreo, obtendo-se os resultados apresentados na Figura 5.1:

```
Call:
lmreg(formula = VAB ~ MAER + PACT + PMED + PSEC + PTER + PDES |
      PACT + PMED + PSEC + PTER + PDES + DORM)

Residuals:
    Min       1Q   Median       3Q      Max
-142705 -10896  12186  18945  73859

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -13130.39   8817.66  -1.489   0.142
MAER          846.94    917.04   0.924   0.360
PACT         694.31    617.07   1.125   0.265
PMED          77.64     88.60   0.876   0.385
PSEC        -572.20    473.86  -1.208   0.232
PTER        -908.62    993.40  -0.915   0.364
PDES        -671.97    464.54  -1.447   0.154

Diagnostic tests:
            df1 df2 statistic p-value
Weak instruments  1  54   1.051  0.3099
Wu-Hausman      1  53   4.586  0.0368 *
Sargan          0  NA      NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40780 on 54 degrees of freedom
Multiple R-Squared:  0.8137,    Adjusted R-squared:  0.793
Wald test: 46.69 on 6 and 54 DF, p-value: < 2.2e-16
```

Figura 5.1 – Estimação do modelo linear com seis variáveis explicativas e com MAER como variável explicativa endógena utilizando o *software R* e o método 2SLS

Na Figura 5.1, verifica-se que as variáveis PSEC, PTER e PDES contribuem de forma negativa para o crescimento do VAB das regiões (de notar que a contribuição negativa das variáveis PSEC e PTER não é realista), enquanto que todas as restantes contribuem de maneira positiva, uma vez que os coeficientes estimados β_j referentes a essas variáveis têm todos sinais positivos. Constata-se ainda que a variável com menor influência é PMED.

Com a realização do teste de endogeneidade Wu-Hausman, verifica-se de imediato que, sendo a estatística p do teste igual a $0.0368 < 0.05$, pode afirmar-se que o teste realizado tem significância estatística, podendo afirmar-se a existência de endogeneidade. De modo a contornar esse problema é indispensável a utilização de variáveis instrumentais e a aplicação da metodologia 2SLS, a qual contempla a introdução da variável instrumental DORM, variável esta que serve de instrumento para a variável explicativa endógena relacionada com o transporte aéreo, MAER.

Aplicando o mesmo procedimento, obtém-se os resultados presentes na Figura 5.2, utilizando desta vez a variável MPAX para definir o transporte aéreo:

```

Call:
ivreg(formula = VAB ~ MPAX + PACT + PMED + PSEC + PTER + PDES |
      PACT + PMED + PSEC + PTER + PDES + DORM)

Residuals:
    Min       1Q   Median       3Q      Max
-140190 -12776   7534   18329  50662

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -10822.86   8442.71  -1.282   0.205
MPAX         6580.38   6289.83   1.046   0.300
PACT          711.75   559.80   1.271   0.209
PMED           93.96    93.49   1.005   0.319
PSEC        -571.39   417.68  -1.368   0.177
PTER        -976.61   939.99  -1.039   0.303
PDES        -675.60   412.84  -1.636   0.108

Diagnostic tests:
              df1 df2 statistic p-value
Weak instruments  1  54   1.163  0.2855
Wu-Hausman       1  53   3.873  0.0543 .
Sargan           0  NA      NA      NA

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36000 on 54 degrees of freedom
Multiple R-Squared:  0.8548,    Adjusted R-squared:  0.8387
Wald test: 59.91 on 6 and 54 DF, p-value: < 2.2e-16

```

Figura 5.2 – Estimação do modelo linear com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o *software R* e o método 2SLS

Para este caso, observando a Figura 5.2, verifica-se de uma maneira geral que, comparativamente aos resultados obtidos na Figura 5.1, as variáveis têm uma maior influência no VAB quando se utiliza MPAX como variável que define o tráfego aéreo, em vez de MAER. De referir ainda que as variáveis PSEC, PTER e PDES, também neste modelo de regressão linear, continuam a relacionar-se de forma negativa com o VAB, isto é, à medida que estas três variáveis vão aumentando, isto irá provocar um decréscimo no VAB de cada região.

Relativamente à estatística p do teste de endogeneidade Wu-Hausman, conclui-se que este teste não é significativo estatisticamente (estatística $p = 0.0543 > 0.05$), não havendo razões para se pensar que existe endogeneidade entre as variáveis, e, portanto, a utilização do método 2SLS não é necessária, bastando apenas fazer a metodologia OLS. A metodologia OLS em tudo é semelhante à 2SLS no que se refere à definição do modelo no *software R*, com a diferença que para OLS, esta é definida no *software* pela função `lm()`. Os resultados para OLS apresentam-se na Figura 5.3:

```

Call:
lm(formula = VAB ~ MPAX + PACT + PMED + PSEC + PTER + PDES)

Residuals:
    Min       1Q   Median       3Q      Max
-49850  -5710   2354   10248  60788

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -14102.426   3769.200   -3.741 0.000445 ***
MPAX          729.795    447.828    1.630 0.109000
PACT          239.652    121.294    1.976 0.053297 .
PMED           8.459     10.625    0.796 0.429411
PSEC        -253.568    120.251   -2.109 0.039628 *
PTER         -126.584    126.040   -1.004 0.319704
PDES         -369.804    124.642   -2.967 0.004474 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17650 on 54 degrees of freedom
Multiple R-squared:  0.9651,    Adjusted R-squared:  0.9612
F-statistic: 249 on 6 and 54 DF,  p-value: < 2.2e-16

```

Figura 5.3 – Estimação do modelo linear com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o *software R* e o método OLS

Pela análise da Figura 5.3 observa-se que na sua maioria, as variáveis aumentaram relativamente ao valor da estatística t , sendo que a variável PMED é a que apresenta estatística t mais baixa, traduzindo-se, portanto, na variável que menos contribui para o crescimento do VAB, o que já sucedia para os modelos cujos resultados expostos se encontram nas Figuras 5.1 e 5.2. Quanto às que contribuem de forma mais acentuada aparecem as variáveis PSEC e PDES, com estatísticas t mais altas de entre todas, produzindo um efeito negativo na variável dependente, uma vez que um acréscimo nestas variáveis irá originar um decréscimo no VAB. Ao contrário destas, com a exceção também da variável PTER, todas as outras provocam um efeito positivo no VAB. Conclui-se ainda que nenhum destes modelos lineares é significativo, uma vez que, os valores representativos das estatísticas t nunca são superiores a 2, para o nível de significância admitido de 0.05.

5.1.2 Modelo log-log com seis variáveis explicativas

Com o mesmo intuito do especificado na secção 5.1.1, apresenta-se de seguida o modelo log-log com seis variáveis explicativas. Em primeiro lugar, tal como para o modelo de regressão linear, aplicou-se o método 2SLS, cujos resultados obtidos no caso em que se utiliza MAER como variável explicativa endógena são os expostos na Figura 5.4:

```

Call:
ivreg(formula = VAB ~ MAER + PACT + PMED + PSEC + PTER + PDES |
      PACT + PMED + PSEC + PTER + PDES + DORM)

Residuals:
    Min       1Q   Median       3Q      Max
-0.31185 -0.12416 -0.02665  0.06542  0.50731

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.56725    2.00522   2.278  0.0275 *
MAER         0.17522    0.23777   0.737  0.4650
PACT        -0.29476    1.06704  -0.276  0.7836
PMED         1.04986    0.80202   1.309  0.1972
PSEC         0.33929    0.36651   0.926  0.3595
PTER        -0.24965    1.27792  -0.195  0.8460
PDES        -0.01882    0.27837  -0.068  0.9464

Diagnostic tests:
              df1 df2 statistic p-value
Weak instruments  1  45   0.901  0.348
Wu-Hausman      1  44   1.744  0.193
Sargan          0  NA      NA      NA

Residual standard error: 0.1959 on 45 degrees of freedom
Multiple R-Squared: 0.9665, Adjusted R-squared: 0.9621
Wald test: 221.6 on 6 and 45 DF, p-value: < 2.2e-16

```

Figura 5.4 – Estimação do modelo log-log com seis variáveis explicativas e com MAER como variável explicativa endógena utilizando o *software R* e o método 2SLS

Pela Figura 5.4 pode observar-se que a utilização do método 2SLS é desnecessária, uma vez que pela estatística p do teste de Wu-Hausman se pode afirmar com 95% de confiança que o teste não tem significância estatística, pelo que não há razão para se pensar que existe endogeneidade entre as variáveis. Por esse motivo, recorre-se à estimativa das elasticidades por meio de OLS, obtendo-se os resultados presentes na Figura 5.5:

```

Call:
lm(formula = VAB ~ MAER + PACT + PMED + PSEC + PTER + PDES)

Residuals:
    Min       1Q   Median       3Q      Max
-0.33482 -0.06115  0.00705  0.04868  0.37818

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.220286    0.571607   5.634 1.09e-06 ***
MAER        -0.005707    0.019543  -0.292 0.771632
PACT         0.149700    0.526157   0.285 0.777321
PMED         0.449312    0.105950   4.241 0.000109 ***
PSEC         0.097949    0.110882   0.883 0.381739
PTER         0.591241    0.387393   1.526 0.133957
PDES        -0.194783    0.092901  -2.097 0.041674 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1149 on 45 degrees of freedom
Multiple R-squared: 0.9885, Adjusted R-squared: 0.9869
F-statistic: 643.3 on 6 and 45 DF, p-value: < 2.2e-16

```

Figura 5.5 – Estimação do modelo log-log com seis variáveis explicativas e com MAER como variável explicativa endógena utilizando o *software R* e o método OLS

Analisando os resultados na Figura 5.5, verifica-se que a maioria das variáveis tomam estatísticas t muito inferiores a 2, com a exceção da variável PMED e PDES, sendo a variável PMED aquela que apresenta maior valor, podendo admitir-se que será esta que gera um impacto maior no VAB. Observa-se ainda que a variável MAER, que caracteriza o tráfego aéreo, toma valor negativo e apresenta o valor mais baixo de estatística t , o que não se pretende neste caso, visto que o objetivo deste estudo é avaliar a influência do tráfego aéreo. Este acaba por não se considerar um modelo significativo para explicar a relação entre o transporte aéreo e o desenvolvimento regional, pelo facto de não ter todas as variáveis a atingir estatísticas t superiores a 2, bem como por apresentar fatores de correlação elevados entre todas as variáveis explicativas.

De forma análoga apresentam-se na Figura 5.6 os resultados para a situação em que se utiliza o mesmo modelo log-log com seis variáveis explicativas, mas desta vez o transporte aéreo é definido pela variável MPAX:

```
Call:
ivreg(formula = VAB ~ MPAX + PACT + PMED + PSEC + PTER + PDES |
      PACT + PMED + PSEC + PTER + PDES + DORM)

Residuals:
      Min       1Q   Median       3Q      Max
-0.30064 -0.10861 -0.01843  0.06877  0.45754

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.42289    1.57033   2.817  0.00718 **
MPAX         0.12699    0.14750   0.861  0.39382
PACT        -0.09123    0.80852  -0.113  0.91066
PMED         0.94636    0.56860   1.664  0.10299
PSEC         0.25668    0.23564   1.089  0.28183
PTER        -0.12664    0.97295  -0.130  0.89702
PDES        -0.10405    0.16549  -0.629  0.53268

Diagnostic tests:
              df1 df2 statistic p-value
Weak instruments  1  45   1.474  0.231
Wu-Hausman      1  44   1.682  0.201
Sargan          0  NA      NA      NA

Residual standard error: 0.1677 on 45 degrees of freedom
Multiple R-Squared: 0.9755, Adjusted R-squared: 0.9722
Wald test: 302.4 on 6 and 45 DF, p-value: < 2.2e-16
```

Figura 5.6 – Estimação do modelo log-log com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o *software R* e o método 2SLS

Observando a Figura 5.6, do mesmo modo que no modelo log-log anterior a utilização de 2SLS era dispensável para estimar as elasticidades, também neste caso a metodologia 2SLS não é necessária, pelo facto da estatística p do teste de Wu-Hausman ser superior a 0.05. Deste modo, aplica-se o método OLS e obtêm-se os resultados presentes na Figura 5.7:

```

Call:
lm(formula = VAB ~ MPAX + PACT + PMED + PSEC + PTER + PDES)

Residuals:
    Min       1Q   Median       3Q      Max
-0.33441 -0.06235  0.00585  0.04786  0.38155

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.25190    0.57725   5.633 1.09e-06 ***
MPAX         -0.00119    0.01802  -0.066  0.9477
PACT          0.13781    0.52542   0.262  0.7943
PMED          0.46377    0.10786   4.300 9.06e-05 ***
PSEC          0.10415    0.10998   0.947  0.3487
PTER          0.57120    0.38949   1.467  0.1495
PDES         -0.19003    0.09182  -2.070  0.0443 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.115 on 45 degrees of freedom
Multiple R-squared:  0.9885,    Adjusted R-squared:  0.9869
F-statistic: 642.1 on 6 and 45 DF,  p-value: < 2.2e-16

```

Figura 5.7 – Estimação do modelo log-log com seis variáveis explicativas e com MPAX como variável explicativa endógena utilizando o *software R* e o método OLS

Pela estimação de OLS, à semelhança do que sucede para o modelo anterior, também aqui se obtém um valor para a estatística t muito inferior ao esperado, relativamente à variável MPAX. A variável PMED continua a ser a que apresenta o valor mais elevado, chegando a ultrapassar o valor 2.

Note-se que não é de estranhar que a variável PDES apareça com sinal negativo (ou, se acontecer, com sinal positivo mas um valor baixo de elasticidade quando comparado com o das restantes variáveis), em qualquer dos modelos, pois isto significa que quanto mais população desempregada existir, menor será o VAB, o que em certa medida é verdade pois o desenvolvimento das regiões não cresce com a população desempregada, porque essa população não desempenha qualquer tipo de função de modo a contribuir para o aumento da economia de um país.

No entanto, atendendo a que apenas duas variáveis atingem valores de estatísticas t superiores a 2, que é o pretendido para o nível de significância de 0.05 admitido, torna-se necessário fazer alguns ajustes nos modelos de seis variáveis, como por exemplo, a eliminação de algumas variáveis explicativas com o objetivo de aumentar a estatística t , para que todas as variáveis obtenham estatísticas $t > 2$ para assim tornar o modelo significativo. Por esse motivo, testam-se de seguida modelos com quatro variáveis, nos quais se retiraram as variáveis PACT e PSEC, por apresentarem estatísticas t baixas. Lembra-se que os valores associados às variáveis endógenas, MAER e MPAX, apesar de serem as que tomam o valor mais baixo de entre todas as variáveis, não se vão retirar, pelo facto de serem estas as necessárias ao estudo para avaliar a influência do transporte aéreo.

5.1.3 Modelo log-log com quatro variáveis explicativas

Uma vez que os modelos definidos com seis variáveis não são significativos, e de acordo com o referido na secção 5.1.2, retiraram-se duas variáveis a cada modelo, formando assim dois modelos de quatro variáveis, sendo que o primeiro utiliza MAER e o segundo utiliza MPAX para caracterizar o tráfego aéreo.

No modelo que utiliza MAER, após a aplicação do método 2SLS, verificou-se que a estatística p associada ao teste de endogeneidade Wu-Hausman, obtido pelo *software R*, é superior a 0.05, pelo que se conclui que esta metodologia não é a indicada, por não se suspeitar da presença de endogeneidade entre variáveis. Por isso, recorre-se à aplicação da metodologia OLS, para o qual os resultados se apresentam na Figura 5.8:

```
Call:
lm(formula = VAB ~ MAER + PMED + PTER + PDES)

Residuals:
    Min       1Q   Median       3Q      Max
-0.23178 -0.07710 -0.02143  0.09418  0.29952

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.10383    0.15118  20.531 < 2e-16 ***
MAER        -0.02283    0.01914  -1.193  0.239007
PMED         0.46176    0.10986   4.203  0.000117 ***
PTER         0.84348    0.15814   5.334  2.70e-06 ***
PDES        -0.19446    0.04457  -4.363  6.97e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1206 on 47 degrees of freedom
Multiple R-squared:  0.9867, Adjusted R-squared:  0.9856
F-statistic: 874.2 on 4 and 47 DF, p-value: < 2.2e-16
```

Figura 5.8 – Estimação do modelo log-log com quatro variáveis explicativas e com MAER como variável explicativa endógena utilizando o *software R* e o método OLS

De uma maneira geral, observa-se que a eliminação de PACT e PSEC resultou num acréscimo em todos os valores de estatísticas t associados às variáveis, mas a variável explicativa endógena, MAER, continua a não tomar um valor significativo. Dessa forma, como todas as outras variáveis já atingem estatísticas t superiores a 2, não faz sentido retirar esta variável para verificar se assim o modelo já se torna significativo, como já foi referido no fim da secção 5.1.2. O mesmo se verificou para o modelo em que o desenvolvimento regional, definido pela variável VAB, é explicado pelas variáveis MPAX, PMED, PTER e PDES.

Portanto, nesta fase, para se obter um modelo que seja o mais adequado para explicar a contribuição do tráfego aéreo no desenvolvimento regional, atendendo ao modelo de seis

variáveis, optou-se por avaliar as correlações entre variáveis explicativas e retiraram-se as variáveis mais correlacionadas entre si.

5.1.4 Modelo log-log com três variáveis explicativas

Na definição do modelo de três variáveis, apenas se vai considerar o tráfego aéreo caracterizado pela variável MPAX, uma vez que se verificou anteriormente que produz resultados muito semelhantes aos que se obtêm quando se utiliza MAER. Outra das razões é também pelo facto de, em trabalhos anteriores, nomeadamente os que vêm especificados no Quadro 3.1, as conclusões retiradas sobre a influência do transporte aéreo ser sempre referida em termos de movimento de passageiros, tornando assim mais fácil a comparação com esses resultados.

Assim, para a caracterização do modelo log-log com três variáveis, foram excluídas as variáveis PACT, PTER e PDES, por apresentarem correlações fortes entre si (coeficiente de correlação de Pearson > 0.8). A variável dependente, VAB, fica assim explicada pelas variáveis MPAX, PMED e PSEC. Dos resultados obtidos para 2SLS, pela Figura 5.9, verifica-se que os valores da coluna relativa às estatísticas t aumentaram, como se pretendia, tomando todos valores de estatísticas $t > 2$. Ainda assim, pelo teste de endogeneidade de Wu-Hausman (estatística $p = 0.357 > 0.05$), conclui-se que este método não é aplicável na estimação dos valores de elasticidade, podendo afirmar-se que não existe endogeneidade entre as variáveis.

```
Call:
ivreg(formula = VAB ~ MPAX + PMED + PSEC | PMED + PSEC + DORM)

Residuals:
    Min       1Q   Median       3Q      Max
-0.349088 -0.100533 -0.004801  0.076674  0.501971

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.05831    0.23786   17.062 < 2e-16 ***
MPAX         0.06659    0.02722    2.446 0.018148 *
PMED         0.77536    0.06508   11.915 6.04e-16 ***
PSEC         0.20577    0.05609    3.668 0.000611 ***

Diagnostic tests:
            df1 df2 statistic  p-value
Weak instruments  1  48   13.300 0.000652 ***
Wu-Hausman      1  47    0.867 0.356640
Sargan          0  NA         NA         NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1396 on 48 degrees of freedom
Multiple R-Squared:  0.9818,    Adjusted R-squared:  0.9807
Wald test: 864.3 on 3 and 48 DF,  p-value: < 2.2e-16
```

Figura 5.9 – Estimação do modelo log-log com três variáveis explicativas e com MPAX como variável explicativa endógena utilizando o *software R* e o método 2SLS

Procede-se assim à estimação pelo método OLS, cujos resultados se apresentam na Figura 5.10, onde se verifica que a variável que mais contribui para o crescimento da variável dependente é PMED (considerando a estatística t obtida), que influencia a variável dependente de forma positiva, sendo que o aumento no valor desta variável provoca também um acréscimo no VAB. Esta influência no VAB traduz-se pelo valor dos parâmetros estimados, que para este tipo de modelos (log-log), se denominam elasticidades, e esta mede a variação percentual de y correspondente a uma dada variação percentual em x_k . Por exemplo, relativamente ao coeficiente estimado para a variável MPAX, este diz-se que mede a elasticidade do VAB em relação a MPAX.

```
Call:
lm(formula = VAB ~ MPAX + PMED + PSEC)

Residuals:
    Min       1Q   Median       3Q      Max
-0.31806 -0.10307 -0.00870  0.07104  0.53341

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.92586    0.18144   21.637 < 2e-16 ***
MPAX         0.04483    0.01228    3.650 0.000647 ***
PMED         0.80390    0.05512   14.584 < 2e-16 ***
PSEC         0.19615    0.05336    3.676 0.000597 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1353 on 48 degrees of freedom
Multiple R-squared:  0.983,    Adjusted R-squared:  0.9819
F-statistic: 923.1 on 3 and 48 DF,  p-value: < 2.2e-16
```

Figura 5.10 – Estimação do modelo log-log com três variáveis explicativas e com MPAX como variável explicativa endógena utilizando o *software R* e o método OLS

Analisando os resultados da Figura 5.10 observa-se ainda que PSEC e MPAX contribuem de forma muito semelhante para o crescimento de VAB, pelo facto das estatísticas t associadas a cada uma delas serem muito próximas. Todas as variáveis explicativas afetam a variável de forma positiva, uma vez que os valores associados ao valor da estatística t e elasticidade apresentam sinal positivo, o que equivale dizer que quando se aumenta o valor de qualquer uma destas variáveis isto gera igualmente um acréscimo no valor da variável dependente.

Relativamente à influência do tráfego aéreo no desenvolvimento regional, esta é quantificada através do valor aproximado de 0.045 estimado para a elasticidade de MPAX, podendo concluir-se que existe uma relação positiva entre o tráfego aéreo, caracterizado pelo movimento de passageiros, e o desenvolvimento regional, medido pelo VAB de cada região, embora sendo pouco acentuada. O aumento de 10% no movimento de passageiros irá resultar em, aproximadamente, um acréscimo de 1% no VAB.

Este resultado, de uma forma geral, vai de encontro às conclusões retiradas em Brueckner (2003) e Percoco (2010), apesar de considerarem o emprego como variável que caracteriza o desenvolvimento regional, em vez do VAB, mas em ambos se conclui que o aumento no movimento de passageiros provoca um acréscimo no desenvolvimento regional.

5.2 Análise de resíduos

De entre os modelos estudados na secção 5.1, o que melhor se adequa é o modelo log-log definido por três variáveis explicativas, pois é o modelo de regressão onde todas as variáveis explicativas tomam valores de estatísticas $t > 2$, sendo considerado o melhor e mais significativo estatisticamente. A variável dependente VAB é explicada pelas variáveis MPAX, PMED e PSEC. A equação do modelo toma a seguinte forma:

$$\ln(VAB) = 3.92586 + 0.04483 \times \ln(MPAX) + 0.80390 \times \ln(PMED) + 0.19615 \times \ln(PSEC) \quad (16)$$

De seguida, procedendo-se a uma análise de resíduos, pretende-se verificar se o modelo de regressão linear múltipla definido pela equação (16), respeita as hipóteses consideradas relativamente aos resíduos, como descrito no capítulo 2. Principalmente o facto de os resíduos serem considerados independentes e $u_i \sim N(0, \sigma^2), i = 1, \dots, n$. Esta última suposição, em que se admite que os erros seguem uma distribuição normal, é exigida para a construção, por exemplo, dos testes F. Se o modelo escolhido for o mais adequado, então os resíduos devem evidenciar tendências que confirmem, ou pelo menos não contestem, os pressupostos admitidos.

Os resíduos de um modelo de regressão representam as diferenças entre o que foi realmente observado e o que foi estimado através da equação de regressão, ou seja, a quantidade que a equação de regressão não foi capaz de explicar, isto é:

$$u_i = y_i - \hat{y}_i, \quad i = 1, \dots, n \quad (17)$$

Onde y_i é uma determinada observação da amostra e \hat{y}_i o seu correspondente valor estimado através do modelo de regressão múltipla. A análise dos resíduos é particularmente útil para se verificar se estes erros têm, aproximadamente, uma distribuição normal com média zero e variância constante.

Uma das maneiras mais simples de analisar os resíduos consiste na sua representação gráfica, que pode ser feita de diferentes modos, sendo neste caso efetuada a representação gráfica dos resíduos *versus* valores modelados de VAB. Fazendo a aplicação ao modelo da equação (16), obtém-se pelo *software R*, os gráficos necessários para a análise de resíduos, que se apresentam na Figura 5.11:

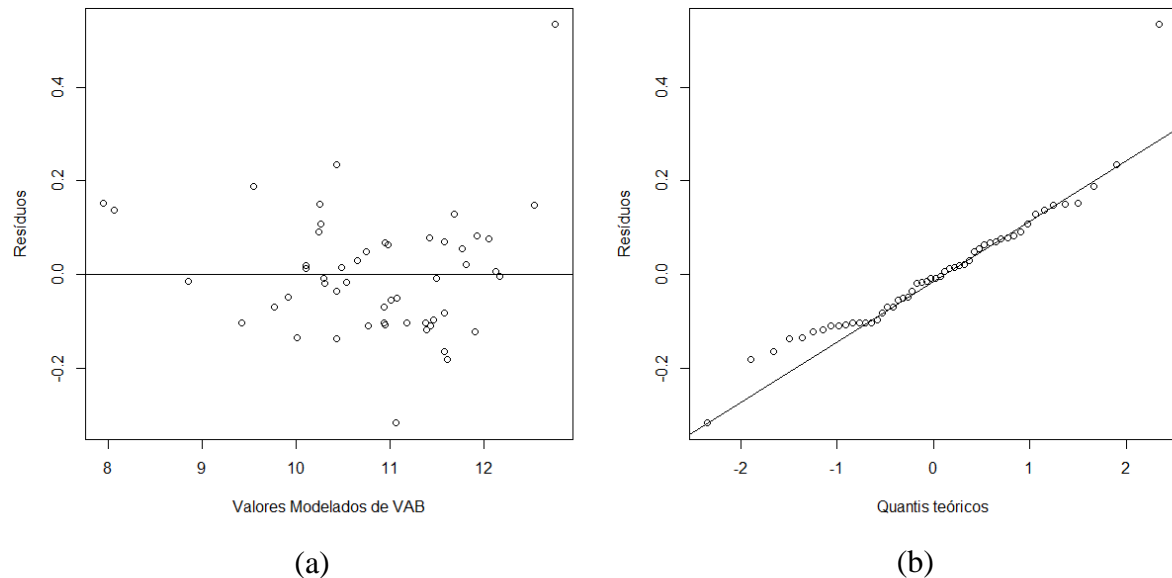


Figura 5.11 – (a) Resíduos *versus* Valores modelados de VAB, (b) Gráfico de Probabilidade Normal dos Resíduos

O gráfico representado na Figura 5.11 (a) é utilizado para confirmar que os pressupostos admitidos sobre os erros apresentarem variância constante e igual a σ^2 , e serem independentes, se verificam. Assim, uma vez que os resíduos se apresentam simétricos em relação à linha $h = 0$, isto é, resíduos homogêneos em torno de zero, com a exceção daquele valor extremo presente no gráfico, não há razões para duvidar da homocedasticidade (homogeneidade das variâncias) dos erros, podendo dizer-se que os erros possuem variância constante e não estão correlacionados entre si.

Apresenta-se ainda um gráfico com o objetivo de verificar o pressuposto de que os erros seguem uma distribuição normal. Observa-se pela Figura 5.11 (b) que há evidências de que os erros não sigam a distribuição normal. Este acontecimento pode estar associado devido à presença de um “outlier”, definido como um valor muito distante das restantes observações, isto é, aquele valor excessivamente grande em relação aos outros, que se pode observar na Figura 5.11 (a) e (b). Embora um valor discrepante possa resultar de um erro de medida, também pode ser um

resultado verdadeiro, indicando um comportamento extremo da variável, que merece ser estudado, e não removido.

Normalmente espera-se que as observações de amostras sigam uma distribuição normal ou aproximadamente normal. De facto, a suposição de normalidade é uma das mais comuns nos procedimentos estatísticos, mas é, frequentemente, a que mais falha. É mais comum que as observações tenham uma distribuição apenas aproximadamente normal. Felizmente, a maior parte da análise de variância pode ser desenvolvida sem tal suposição, que é necessária somente para justificar o uso de certos testes de significância formalmente precisos e certas fórmulas de estimação (Johnson e Leone, 1964).

5.3 Análise dos resultados

De modo a determinar quais as regiões que mais eficientemente estão a utilizar o transporte aéreo para o seu desenvolvimento, e segundo o modelo de regressão escolhido na secção 5.1.4, calcularam-se os valores modelados de VAB e compararam-se com os valores de VAB reais, isto é, os que foram retirados da base de dados da Eurostat. De seguida calcularam-se os resíduos, que consistem na diferença entre os valores reais observados e os valores modelados pela equação de regressão, e que se encontram expostos no Quadro 5.1. Para uma melhor compreensão e análise dos resultados, estes apresentam-se em forma de gráfico, na Figura 5.12, que relaciona os valores de VAB reais com os valores de VAB modelados.

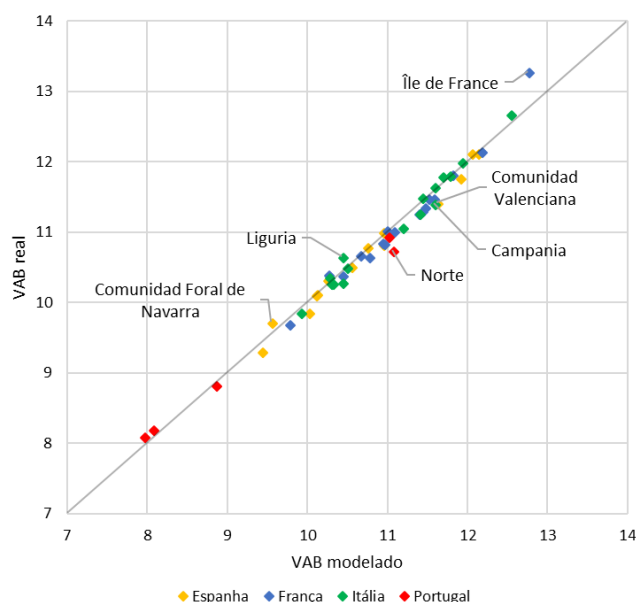


Figura 5.12 – VAB real *versus* VAB modelado

Quadro 5.1 – Valores dos resíduos

	Regiões NUTS 2	VAB real	VAB modelado	Resíduos
ES	Galícia	10.832	10.936	-0.104
	Principado de Asturias	9.868	10.003	-0.135
	Cantabria	9.314	9.417	-0.103
	País Vasco	11.008	10.942	0.067
	Comunidad Foral de Navarra	9.727	9.540	0.187
	Aragón	10.328	10.237	0.090
	Comunidad de Madrid	12.123	12.047	0.077
	Castilla y León	10.791	10.744	0.048
	Cataluña	12.129	12.123	0.006
	Comunidad Valenciana	11.429	11.612	-0.183
	Illes Balears	10.115	10.102	0.013
	Andalucía	11.779	11.901	-0.122
	Región de Murcia	10.126	10.106	0.020
	Canarias (ES)	10.521	10.539	-0.018
FR	Île de France	13.288	12.754	0.533
	Champagne-Ardenne	10.404	10.255	0.149
	Picardie	10.654	10.764	-0.110
	Centre (FR)	11.041	10.977	0.064
	Nord - Pas-de-Calais	11.488	11.496	-0.008
	Lorraine	10.862	10.931	-0.070
	Alsace	10.842	10.949	-0.108
	Pays de la Loire	11.490	11.573	-0.083
	Bretagne	11.314	11.424	-0.111
	Poitou-Charentes	10.681	10.651	0.030
	Aquitaine	11.364	11.462	-0.098
	Midi-Pyrénées	11.276	11.381	-0.104
	Limousin	9.697	9.767	-0.071
	Rhône-Alpes	12.160	12.165	-0.005
Auvergne	10.393	10.429	-0.036	
Languedoc-Roussillon	11.017	11.068	-0.050	
Provence-Alpes-Côte d'Azur	11.825	11.803	0.021	
IT	Piemonte	11.645	11.576	0.069
	Liguria	10.663	10.429	0.233
	Lombardia	12.677	12.530	0.147
	Abruzzo	10.283	10.303	-0.020
	Campania	11.409	11.574	-0.165
	Puglia	11.077	11.180	-0.103
	Calabria	10.289	10.425	-0.137
	Sicilia	11.269	11.388	-0.119
	Sardegna	10.279	10.288	-0.009
	Veneto	11.820	11.765	0.055
	Friuli-Venezia Giulia	10.373	10.266	0.107
	Emilia-Romagna	11.806	11.678	0.128
	Toscana	11.502	11.423	0.079
	Umbria	9.864	9.912	-0.048
Marche	10.502	10.487	0.015	
Lazio	12.006	11.923	0.083	
PT	Norte	10.740	11.058	-0.318
	Algarve	8.832	8.848	-0.016
	Área Metropolitana de Lisboa	10.951	11.006	-0.055
	Região Autónoma dos Açores	8.102	7.951	0.151
	Região Autónoma da Madeira	8.202	8.065	0.137

Analisando a Figura 5.12 e complementando com a informação exposta no Quadro 5.1, verifica-se que as regiões com um comportamento melhor que o expectável são aquelas que se encontram acima da diagonal, correspondendo a valores de resíduos positivos, enquanto que as que se encontram abaixo desta mostra que o valor real não atinge o modelado, o que se traduz pelo valor negativo dos resíduos. As regiões com valores de resíduos muito próximos de zero mostram que o modelo de regressão escolhido expressa quase exatamente o que se passa na realidade, visto que quando o resíduo toma o valor zero indica que o VAB modelado é igual ao VAB real.

Relativamente à Figura 5.12 destaca-se a região Île de France, que apresenta o maior valor positivo para os resíduos, significando que esta é a região que, de acordo com o modelo estudado, faz uso do transporte aéreo de forma mais eficaz para favorecer o seu desenvolvimento. A principal razão para que esta seja a região com um desenvolvimento melhor que o esperado poderá ser devido ao facto de esta ser uma região com grande atração turística, no qual se encontra a *Disneyland Paris*, considerada uma das atrações mais visitada em toda a França e Europa. Isto gera grande movimento turístico na região, recebendo milhões de visitantes anualmente, contribuindo para um grande volume anual de movimentos de aeronaves e de passageiros, aproximadamente 671.015 milhares de voos e 95361 milhões de habitantes no ano de 2015, fatores que caracterizam o tráfego aéreo e contribuem para o crescimento da região.

A outra região com maior valor positivo para os resíduos é Liguria, situada no Noroeste de Itália, na fronteira com França. Também esta é uma região com grande apelo turístico, dotada de belezas naturais, com altas montanhas e colinas, entre as quais as famosas *Cinque Terre*, constituídas por cinco cidades marítimas. Por esse motivo, o grande movimento gerado nessa zona leva a um impacto muito positivo no seu desenvolvimento, razão pela qual apresenta os valores mais elevados de resíduos, em conjunto com a região Île de France.

Destaca-se ainda, na Figura 5.12, a região espanhola denominada Comunidad Foral de Navarra, por apresentar também um valor de resíduos elevado quando comparado com os restantes. Pode dizer-se que utiliza o transporte aéreo em conjunto com as variáveis PMED e PSEC, em benefício do desenvolvimento.

Contrariamente às três regiões acima mencionadas, salientam-se as regiões Norte de Portugal, Comunidad Valenciana em Espanha e Campania em Itália. Estas são as três regiões que apresentam os maiores valores negativos para os resíduos, situando-se, portanto, abaixo da

diagonal apresentada no gráfico da Figura 5.12, o que se traduz num desenvolvimento inferior ao esperado pelo modelo de regressão aplicado.

A região Norte de Portugal é a que apresenta o valor de resíduos mais negativo (-0.318), o que não significa necessariamente que é a região com o menor valor de VAB, caracterizando o desenvolvimento regional. O que sucede é que apresenta a maior diferença negativa entre o valor real do VAB da região e o VAB esperado, ou seja, relativamente ao desenvolvimento regional este ficou aquém do previsto.

Por último, observa-se ainda que a Comunidad Valenciana e a região de Campania apresentam valores de resíduos próximos, -0.183 e -0.165 , respetivamente, cerca de metade do valor obtido para a região Norte de Portugal.

6 CONCLUSÃO

A presente dissertação tinha como objetivo avaliar a influência do transporte aéreo no desenvolvimento das regiões do Sudoeste Europeu, particularmente as NUTS 2 de Espanha, França, Itália e Portugal. Pretendia-se ainda verificar quais as regiões que estão a utilizar mais eficientemente o transporte aéreo para o seu desenvolvimento, bem como os principais fatores que para isso contribuem.

Para isso, definiu-se a variável que caracteriza o desenvolvimento regional, a variável que define o transporte aéreo e ainda um conjunto de variáveis que também têm influência no desenvolvimento regional. Com as variáveis escolhidas, analisaram-se vários modelos de regressão, de modo a determinar o mais adequado para expressar a relação entre o transporte aéreo e o desenvolvimento regional. De seguida, procedeu-se à análise de resíduos, verificando se estes seguem uma distribuição normal com média zero, são independentes e apresentam variância constante. Por fim, fazendo a comparação entre os valores de VAB reais e os valores de VAB modelados pelo modelo de regressão escolhido, determinaram-se as regiões com um desenvolvimento superior ao esperado, correspondendo aos valores de resíduos positivos, e também as que se encontram abaixo do previsto, que corresponde às regiões que apresentam valores negativos de resíduos.

Posto isto, as principais conclusões a retirar são:

- Não há razões para se pensar no problema de endogeneidade neste caso, visto que, a aplicação de 2SLS como método de estimação dos parâmetros de regressão, não se torna necessária;
- O modelo escolhido, com VAB a ser explicado pelas variáveis MPAX, PMED e PSEC, apesar de falhar no pressuposto da normalidade dos erros, é o modelo mais adequado para expressar a relação entre o transporte aéreo e o desenvolvimento regional;

- A influência do transporte aéreo é mais acentuada nas regiões Île de France e Liguria, em virtude de serem duas regiões com grande atração turística, fator esse que contribui para um maior volume de fluxo de passageiros nestas regiões. Este acréscimo no fluxo de passageiros que caracteriza o tráfego aéreo, gera um aumento no VAB regional, favorecendo o seu desenvolvimento, levando a que sejam as duas regiões com maior valor positivo dos resíduos, sendo estas consideradas as que apresentam um desenvolvimento maior que o expectável;
- Observa-se que em Espanha, Itália e Portugal, cerca de 50% dos resíduos de cada país são positivos, contrariamente ao que se verifica em França, que apresenta mais resíduos negativos do que positivos, isto é, aproximadamente 70% das regiões encontram-se com um desenvolvimento inferior ao esperado.

O estudo desenvolvido na presente dissertação serve de referência para futuros trabalhos que tenham como objetivo avaliar o efeito do transporte aéreo no desenvolvimento regional nas regiões do Sudoeste Europeu. Assim, com o intuito de aprofundar e desenvolver este tema, apresentam-se de seguida algumas sugestões para trabalhos futuros:

- Aplicar outros métodos de análise de regressão, como por exemplo a metodologia baseada em modelos de efeitos aleatórios (*random-effect model*) e modelos de efeitos fixos (*fixed-effect model*), através da utilização de multiplicadores, aplicada por Button *et al.* (2010) ou a metodologia de não-causalidade de Granger (*Granger noncausality methodology*) utilizada por Makkala e Tervo (2013);
- Escolha de diferentes variáveis instrumentais, que estejam mais correlacionadas com a variável explicativa endógena;
- Recolha de dados para um período mais alargado em vez de dados anuais;
- Estender a análise ao resto da Europa.

REFERÊNCIAS BIBLIOGRÁFICAS

- Benell, D. e Prentice, B. (1993), “A regression model for predicting the impacts of Canadian airports”, *Logistics and Transportation Review*, Vol. 29, No. 2, 139-158.
- Brueckner, J. (2003), “Airline traffic and urban economic development”, *Urban Studies*, Vol. 40, No. 8, 1455-1469.
- Button, K. e Taylor, S. (2000), “International air transportation and economic development”, *Journal of Air Transport Management*, 6, 209-222.
- Button, K., Doh, S. e Yuan, J. (2010), “The role of small airports in economic development”, *Journal of Airport Management*, 4, 125-136.
- Dobruszkes, F., Lennert, M. e Van Hamme, G. (2011), “An analysis of the determinants of air traffic volume for European metropolitan areas”, *Journal of Transport Geography*, Vol. 19, No. 4, 755-762.
- Esteves, A. (2014). “O impacto do Aeroporto Francisco Sá Carneiro no desenvolvimento da região norte”. Dissertação de Mestrado, Departamento de Aeronáutica e Transportes da Universidade Lusófona de Humanidades e Tecnologias, Lisboa.
- Goetz, A. (1992), “Air passenger transportation and growth in the U.S. urban system 1950 – 1987”, *Growth and Change*, 23, 218-242.
- Green, R. (2007), “Airports and economic development”, *Real Estate Economics*, 35, 91-112.
- ICAO@ (2018), <https://www.icao.int/Newsroom/Pages/Continued-passenger-traffic-growth-and-robust-air-cargo-demand-in-2017.aspx>.
- ISCED@ (2011), [http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_\(ISCED\)](http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED)).
-

- Johnson, L. e Leone, C. (1964) “Statistics and experimental design in engineering and the physical sciences”. Wiley, Nova Iorque.
- Mukkala, K. e Tervo, H. (2013), “Air transportation and regional growth: which way does the causality run?”, *Environment and Planning A*, Vol. 45, 1508-1520.
- Murteira, M. e Castro, V. (2018). “Introdução à Econometria” (2ª edição). Almedina, Coimbra.
- Percoco, M. (2010), “Airport activity and local development: evidence from Italy”, *Urban Studies*, Vol. 47, No. 11, 2427-2443.
- Postorino, M. (2010) “Development of Regional Airports - Theoretical Analyses and Case Studies”. WIT Press, Southampton.
- Regions in the European Union, Nomenclature of territorial units for statistics, NUTS 2013/EU-28 – Luxemburgo: Office for Official Publications of the European Communities, 2015.
- Van De Vijver, E., Derudder, B. e Witlox, F. (2016), “Air Passenger Transport and Regional Development: Cause and Effect in Europe”, *Promet - Traffic & Transportation*, Vol. 28, No. 2, 143-154.
- Wooldridge, J. (2012). “Introductory Econometrics – A modern approach” (5ª edição). South-Western, Cengage Learning, EUA.

ANEXO A

Quadro A.1 – Valores das variáveis

REGIÕES NUTS 2		Variável Dependente	Variáveis Explicativas							Variável Instrumental
CÓDIGO NUTS 2	Designação	VAB [10 ⁶ €]	Endógenas			Exógenas				DORM [10 ³ noites]
			MAER [10 ³ voos]	MPAX [10 ⁶ hab]	PACT [10 ³ hab]	PMED [10 ³ hab]	PSEC [10 ³ hab]	PTER [10 ³ hab]	PDES [10 ³ hab]	
ES	ES11 Galicia	50 639.200	24.687	4.022	1 186.200	1 512.265	224.900	685.000	238.200	5 868.298
	ES12 Principado de Asturias	19 303.500	9.512	1.118	441.900	664.571	75.300	274.100	87.400	2 867.346
	ES13 Cantabria	11 091.100	8.206	0.874	262.900	365.264	46.700	165.200	48.000	2 060.268
	ES21 País Vasco	60 375.600	39.361	4.517	973.700	1 513.405	224.500	606.200	149.000	3 209.970
	ES22 Comunidad Foral de Navarra	16 769.000	2.871	0.144	288.100	414.934	78.200	164.900	40.900	1 287.895
	ES24 Aragón	30 561.800	3.068	0.424	602.500	809.106	138.400	346.400	101.500	3 721.370
	ES30 Comunidad de Madrid	184 125.200	348.550	46.297	3 139.800	4 565.488	399.400	2 255.600	555.600	10 651.266
	ES41 Castilla y León	48 591.200	1.886	0.212	1 078.000	1 424.895	210.600	625.700	203.000	5 907.550
	ES51 Cataluña	185 235.700	290.367	41.877	3 480.400	4 349.431	736.700	2 112.900	668.900	15 702.545
	ES52 Comunidad Valenciana	91 959.700	115.932	15.600	2 237.700	2 781.036	425.400	1 281.400	529.400	15 190.004
	ES53 Illes Balears	24 701.200	248.645	32.354	565.300	614.235	79.600	393.000	100.600	4 772.064
	ES61 Andalucía	130 493.900	148.177	20.807	3 707.700	4 082.214	361.800	2 053.300	1 236.600	23 483.467
	ES62 Región de Murcia	24 973.000	7.089	1.066	642.400	723.104	90.900	344.000	166.300	2 076.669
	ES70 Canarias (ES)	37 085.800	283.731	32.538	1 024.600	1 077.955	73.700	648.500	315.600	7 803.215
FR	FR10 Île de France	589 843.270	671.015	95.361	5 262.400	9 586.335	597.000	4 114.700	531.200	28 479.371
	FR21 Champagne-Ardenne	32 994.470	0.728	0.081	506.000	956.871	111.000	312.200	69.500	2 077.817
	FR22 Picardie	42 341.690	27.658	4.329	735.600	1 357.401	143.100	496.800	85.400	2 032.544
	FR24 Centre (FR)	62 349.970	1.419	0.187	999.600	1 927.444	206.100	667.500	111.400	4 362.315
	FR30 Nord - Pas-de-Calais	97 546.340	16.795	1.520	1 513.300	3 004.342	293.300	1 003.300	227.300	4 608.123
	FR41 Lorraine	52 131.780	5.127	0.245	908.800	1 826.895	191.600	607.800	117.800	2 927.181
	FR42 Alsace	51 105.590	28.165	8.203	796.000	1 515.266	202.700	524.800	77.600	3 978.599
	FR51 Pays de la Loire	97 703.670	48.806	4.322	1 498.800	3 005.364	340.100	986.300	139.700	5 899.030
	FR52 Bretagne	81 920.350	26.360	1.665	1 320.100	2 771.217	276.700	879.200	106.700	6 335.836
	FR53 Poitou-Charentes	43 514.690	2.732	0.216	688.700	1 400.178	140.400	459.600	70.000	4 255.935
	FR61 Aquitaine	86 191.190	71.236	7.234	1 355.400	2 758.699	244.600	921.000	139.300	7 522.535
	FR62 Midi-Pyrénées	78 932.250	80.277	8.020	1 221.600	2 496.717	237.300	830.600	106.900	5 751.580
	FR63 Limousin	16 261.000	4.453	0.292	283.600	591.217	49.700	195.000	26.600	1 136.137
	FR71 Rhône-Alpes	190 956.660	106.779	9.141	2 668.400	5 294.354	575.500	1 800.500	257.200	13 867.397
FR72 Auvergne	32 615.490	8.647	0.393	523.400	1 084.345	112.600	339.500	45.900	3 099.870	
FR81 Languedoc-Roussillon	60 924.890	25.402	2.718	1 026.800	2 016.986	147.800	713.400	138.200	5 944.500	
FR82 Provence-Alpes-Côte d'Azur	136 573.920	228.699	20.709	1 964.300	3 839.453	282.400	1 435.400	226.300	13 135.893	
IT	ITC1 Piemonte	114 175.970	36.107	3.779	1 845.800	2 716.623	540.700	1 093.300	194.000	5 085.447
	ITC3 Liguria	42 727.450	13.720	1.354	625.800	1 013.288	112.400	452.400	60.100	5 551.721
	ITC4 Lombardia	320 210.370	306.544	38.480	4 267.900	6 311.650	1 295.900	2 627.800	344.600	12 420.427
	ITF1 Abruzzo	29 216.850	5.788	0.610	502.800	844.218	149.300	276.300	66.300	3 526.375
	ITF3 Campania	90 132.100	51.989	6.132	1 785.500	3 006.964	315.700	1 098.200	370.100	8 001.210
	ITF4 Puglia	64 665.350	45.990	6.201	1 326.000	1 979.611	234.100	782.500	272.900	6 386.064
	ITF6 Calabria	29 398.690	20.728	2.823	610.300	1 077.264	72.400	362.200	148.300	5 098.088
	ITG1 Sicilia	78 334.050	105.522	14.038	1 555.500	2 535.856	203.000	967.300	351.500	6 001.083
	ITG2 Sardegna	29 117.530	57.842	7.602	627.600	824.990	85.600	409.400	115.800	4 770.084
	ITH3 Veneto	135 931.820	110.257	13.678	2 039.100	3 030.472	674.300	1 186.600	152.000	10 508.040
	ITH4 Friuli-Venezia Giulia	31 975.240	8.303	0.738	499.900	803.765	145.000	308.100	41.400	1 850.825
	ITH5 Emilia-Romagna	134 040.930	59.399	7.200	1 913.800	2 910.632	589.600	1 141.000	153.300	21 446.689
	ITII Toscana	98 906.270	62.414	7.201	1 577.200	2 341.656	393.000	1 012.500	152.800	10 410.259
	ITII Umbria	19 218.420	3.278	0.280	369.800	608.438	93.000	234.400	40.200	2 168.600
ITII Marche	36 389.540	6.080	0.515	634.500	994.060	203.100	363.500	66.300	4 434.165	
ITII Lazio	163 729.020	346.409	46.034	2 436.700	4 101.128	331.200	1 816.500	299.200	8 088.003	
PT	PT11 Norte	46 161.580	65.675	8.089	1 600.100	1 394.387	497.400	834.100	236.100	3 027.593
	PT15 Algarve	6 852.060	42.852	6.439	195.800	198.661	19.700	136.200	26.100	3 982.011
	PT17 Área Metropolitana de Lisboa	56 991.720	160.778	20.113	1 253.500	1 651.791	154.700	939.500	171.800	3 107.961
	PT20 Região Autónoma dos Açores	3 301.320	18.575	1.657	107.700	74.645	15.300	64.900	14.600	538.621
	PT30 Região Autónoma da Madeira	3 647.770	25.600	2.706	116.100	90.799	11.000	77.600	18.400	693.820