

Departamento de Estatística e Investigação Operacional da Faculdade de Ciências
Universidade de Lisboa



MODELAÇÃO E PROJEÇÃO ESTOCÁSTICA DA POPULAÇÃO PORTUGUESA PARA 2050

Ana Patrícia da Mota Afonso Pedrosa

Mestrado em Bioestatística

2011

Departamento de Estatística e Investigação Operacional da Faculdade de Ciências
Universidade de Lisboa



MODELAÇÃO E PROJEÇÃO ESTOCÁSTICA DA POPULAÇÃO PORTUGUESA PARA 2050

Ana Patrícia da Mota Afonso Pedrosa

Dissertação orientada pela Prof. Doutora Maria Lucília Salema Carvalho

Coorientador: Dr. Francisco de Freitas Vala Salvador

Mestrado em Bioestatística

2011

As projeções populacionais são cruciais para a tomada de decisões no campo social, económico e político possibilitando informação de grande relevância na prossecução de políticas públicas e na definição das estratégias nacionais e europeias.

Apresenta-se, pela primeira vez, uma projeção puramente estocástica da população portuguesa para 2050. Contrariamente às projeções convencionais baseadas em modelos determinísticos, as projeções estocásticas permitem a associação de distribuições de probabilidade à projeção das taxas demográficas assim como a construção de intervalos de confiança, quantificando a incerteza associada.

Utilizam-se metodologias que incorporam modelos demográficos com modelos de séries temporais para descrever os movimentos migratórios e os índices temporais caracterizadores da mortalidade e fertilidade. Estes modelos são utilizados no processo estocástico de renascimento da população por sexo e idades. Os resultados são comparados com as projeções oficiais do Instituto Nacional de Estatística, I. P. e do Eurostat.

Palavras Chave: Demografia, projeção estocástica, população, séries temporais.

Abstract

Population projections are crucial for many social, political and economical decisions providing information of great relevance in the pursuit of public policies and the circumscription of European and national strategies.

It is presented for the first time, a purely stochastic projection up to 2050 of the Portuguese population. Unlike conventional projections based on deterministic models, stochastic projections allow the association of probability distributions to the projection of demographic rates as well the construction of confidence intervals, quantifying the associated uncertainty.

Methodologies that incorporate demographic models with time series models are used to describe the migration and the time varying indices of mortality and fertility. These models are used in a stochastic population renewal process, by sex and age. The results are compared with official projections from INE Portugal and Eurostat.

Keywords: Demography, stochastic projection, population, time series.

| | |
|--------------------------------------------------------------------------|-------------|
| Lista de Figuras | ix |
| Lista de Tabelas | xiii |
| Lista de Acrónimos e Abreviaturas | xv |
| 1 Introdução | 1 |
| 1.1 Tema | 1 |
| 1.2 Organização | 2 |
| 2 Dinâmica demográfica recente em Portugal | 3 |
| 2.1 A evolução da população residente | 3 |
| 2.2 Contributo da componentes natural | 3 |
| 2.2.1 Mortalidade | 4 |
| 2.2.2 Fertilidade | 6 |
| 2.3 Contributo da componente migratória | 8 |
| 3 Principais metodologias de projeções populacionais | 9 |
| 3.1 Revisão de literatura | 9 |
| 3.2 Tipo de projeções | 10 |
| 3.3 Um discurso sobre a vantagem das projeções probabilísticas | 11 |
| 4 Metodologia para a realização de projeções estocásticas | 13 |

| | | |
|----------|------------------------------------------------------------------------------------------|-----------|
| 4.1 | Mortalidade | 13 |
| 4.1.1 | Modelo de Lee Carter | 13 |
| 4.1.2 | Projeção e construção da tábua de mortalidade | 16 |
| 4.2 | Fertilidade | 18 |
| 4.2.1 | Modelo de Lee Carter modificado | 18 |
| 4.2.2 | Projeção da fertilidade | 19 |
| 4.3 | População | 19 |
| 4.4 | Metodologia de Box-Jenkins | 20 |
| 4.4.1 | Identificação | 22 |
| 4.4.2 | Estimação | 23 |
| 4.4.3 | Avaliação do diagnóstico | 23 |
| 5 | Aplicação do modelo desenvolvido de projeções estocásticas à realidade portuguesa | 27 |
| 5.1 | Modelação da componente da mortalidade | 27 |
| 5.1.1 | O modelo de Lee Carter | 28 |
| 5.1.2 | Modelação do índice geral da mortalidade como uma série temporal | 30 |
| 5.1.3 | Projeção do índice γ_t | 36 |
| 5.2 | Modelação da componente da fertilidade | 38 |
| 5.2.1 | O modelo de Lee Carter modificado | 38 |
| 5.2.2 | Modelação do índice transformado da fertilidade como uma série temporal | 41 |
| 5.2.3 | Projeção da série g_t | 44 |
| 5.2.4 | Adequação do modelo ao caso português | 45 |
| 5.3 | Modelação da componente migratória | 47 |
| 5.4 | Projeção da população | 53 |
| 6 | Conclusão | 57 |
| | Referências Bibliográficas | 59 |
| | Apêndices | 63 |
| A | Código em R | 65 |
| A.1 | Mortalidade | 65 |
| A.2 | Fertilidade | 68 |

| | |
|-----------------------------|-----------|
| A.3 Migrações | 69 |
| B Gráficos e Quadros | 71 |

Lista de Figuras

| | | |
|------|------------------------------------------------------------------------------------------------------|----|
| 2.1 | Pirâmides etárias da população portuguesa, 1970 e 2009 | 4 |
| 2.2 | Saldo natural, 1970-2009 | 4 |
| 2.3 | Esperança de vida à nascença por sexo, Portugal, 1970-2009 | 5 |
| 2.4 | Índice de envelhecimento em Portugal, 1970-2009 | 5 |
| 2.5 | Índice de longevidade em Portugal, 1970-2009 | 6 |
| 2.6 | Índice Sintético de Fecundidade em Portugal, 1970-2009 | 6 |
| 2.7 | Idade média da mãe ao nascimento do primeiro filho em Portugal, 1970-2009 | 7 |
| 2.8 | Maternidade precoce e maternidade tardia | 7 |
| 2.9 | Imigração, emigração e saldo migratório em Portugal, 1993-2009 | 8 |
| 5.1 | Proporção de população com mais de 85 anos de idade na população total por sexo, 1970-2009 | 28 |
| 5.2 | Estimativas do parâmetro α_a por sexo e idade, 1970-2009 | 28 |
| 5.3 | Estimativas do parâmetro β_a por sexo e idade, 1970-2009 | 29 |
| 5.4 | Estimativas de γ_t e de γ_t reestimado, Homens, 1970-2009 | 29 |
| 5.5 | Estimativas de γ_t e de γ_t reestimado, Mulheres, 1970-2009 | 30 |
| 5.6 | FAC e FACP do índice γ_t | 30 |
| 5.7 | Primeiras diferenças de γ_t , 1971-2009 | 31 |
| 5.8 | FAC e FACP das primeiras diferenças do índice γ_t | 31 |
| 5.9 | Q-Q plot dos resíduos estandardizados | 34 |
| 5.10 | Resíduos dos modelos | 35 |

| | | |
|------|----------------------------------------------------------------------------------------------------|----|
| 5.11 | Projeção do índice de mortalidade γ_t , Homens, 2010-2050 | 36 |
| 5.12 | Projeção do índice de mortalidade γ_t , Mulheres, 2010-2050 | 36 |
| 5.13 | Projeção das taxas de mortalidade específicas por idade para 2010, 2030 e 2050 | 37 |
| 5.14 | Projeção da esperança média de vida à nascença, 2010-2050 | 37 |
| 5.15 | Estimativas do parâmetro θ_a por idade fértil da mulher, 1980-2009 | 38 |
| 5.16 | Estimativas do parâmetro δ_a por idade fértil da mulher, 1980-2009 | 39 |
| 5.17 | Estimativas do parâmetro κ_t , 1980-2009 | 39 |
| 5.18 | Taxas de fertilidade específicas dos 20 aos 23 anos, 1980-2009 | 40 |
| 5.19 | Estimativas do índice transformado da fertilidade g_t , 1980-2009 | 41 |
| 5.20 | Primeiras diferenças de g_t , 2010-2050 | 41 |
| 5.21 | Segundas diferenças de g_t , 2010-2050 | 42 |
| 5.22 | Resíduos do modelo | 44 |
| 5.23 | Projeções de 2010 a 2050 | 44 |
| 5.24 | Projeção do Índice Sintético de Fecundidade (ISF), 1980-2009 | 45 |
| 5.25 | Taxas de fertilidade específicas por mulher, 2009, 2010, 2030 e 2050 | 46 |
| 5.26 | Projeção da idade média da mulher ao nascimento do primeiro filho, 2010-2050 | 46 |
| 5.27 | Processos migratórios por sexo, 1993-2009 | 47 |
| 5.28 | Projeção do saldo migratório, ensaio 1, 2010-2050 | 48 |
| 5.29 | Projeção do saldo migratório, ensaio 2, 2010-2050 | 49 |
| 5.30 | Projeção do saldo migratório, ensaio 3, 2010-2050 | 49 |
| 5.31 | Projeção do saldo migratório, ensaio 4, 2010-2050 | 50 |
| 5.32 | Resíduos dos modelos respeitantes ao saldo migratório, ensaio 4 | 51 |
| 5.33 | Esquema do método utilizado para desagregação da projeção do saldo migratório por idades | 52 |
| 5.34 | Distribuição da proporção de entradas e saídas por sexo e idade, 2008/2009 | 52 |
| 5.35 | Projeção da população total portuguesa, 2010-2050 | 53 |
| 5.36 | Taxa de crescimento efetiva populacional, 2010-2050 | 54 |
| 5.37 | Saldo total, natural e migratório, 2010-2050 | 54 |
| 5.38 | Pirâmide etária da população portuguesa, 2009 e 2050 | 55 |
| 5.39 | Projeção do índice de dependência de idosos por sexo, 2010-2050 | 55 |
| 5.40 | Projeção do índice de dependência dos jovens por sexo, 2010-2050 | 56 |
| 5.41 | Projeção do índice de envelhecimento e longevidade, 2010-2050 | 56 |

| | | |
|-----|----------------------------------------------------------------------------------------------------------|----|
| B.1 | FAC e FACP dos resíduos dos modelos ajustados ao índice geral de mortalidade γ_t , por sexo . . . | 71 |
| B.2 | Projeção da esperança média de vida aos 65 anos de idade, 2010-2050 | 72 |
| B.3 | FAC e FACP dos resíduos dos modelos ajustados ao índice transformado da fertilidade, g_t | 72 |
| B.4 | Projeção da imigração e emigração dos homens, ensaio 1, 2010-2050 | 72 |
| B.5 | Projeção da imigração e emigração das mulheres, ensaio 1, 2010-2050 | 73 |
| B.6 | Projeção da imigração e emigração dos homens, ensaio 2, 2010-2050 | 73 |
| B.7 | Projeção da imigração e emigração das mulheres, ensaio 2, 2010-2050 | 73 |
| B.8 | FAC e FACP dos resíduos dos modelos ajustados ao saldo migratório, por sexo | 74 |
| B.9 | Q-Q plot dos resíduos estandardizados, dos modelos ajustados ao saldo migratório, por sexo . . . | 74 |

Lista de Tabelas

| | | |
|------|-------------------------------------------------------------------------------------------------------------------|----|
| 4.1 | Principais propriedades da FAC e da FACP empíricas das séries temporais estacionárias | 23 |
| 5.1 | Resultados (<i>valor p</i>) do teste estatístico ADF | 32 |
| 5.2 | Designação dos modelos e respetivos resultados, Homens | 33 |
| 5.3 | Designação dos modelos e respetivos resultados, Mulheres | 33 |
| 5.4 | Resultados da aplicação do teste de Ljung-Box (<i>valor p</i>) | 34 |
| 5.5 | Estimativa dos parâmetros para o modelo ARIMA(0,1,1) com <i>drift</i> , Homens | 35 |
| 5.6 | Estimativa dos parâmetros para o modelo ARIMA(2,1,0) com <i>drift</i> , Mulheres | 35 |
| 5.7 | Designação dos modelos e respetivos resultados | 43 |
| 5.8 | Resultados da aplicação do teste de Ljung-Box (<i>valor p</i>) aos resíduos dos modelos seleccionados | 43 |
| 5.9 | Estimativa dos parâmetros para o modelo ARIMA(1,2,0) | 43 |
| 5.10 | Estimativa dos parâmetros do modelo ARMA(2,1), Homens | 50 |
| 5.11 | Estimativa dos parâmetros do modelo ARMA(4), Mulheres | 51 |
| B.1 | Projeção dos principais indicadores demográficos | 75 |
| B.2 | Projeção dos principais indicadores demográficos, intervalo de confiança superior a 95% | 76 |
| B.3 | Projeção dos principais indicadores demográficos, intervalo de confiança superior a 80% | 77 |
| B.4 | Projeção dos principais indicadores demográficos, intervalo de confiança inferior a 95% | 78 |
| B.5 | Projeção dos principais indicadores demográficos, intervalo de confiança inferior a 80% | 79 |

Lista de Acrónimos e Abreviaturas

ADF Aumentado de Dickey-Fuller

AIC Critério de Informação Akaike

EPC Comité de Política Económica

ISF Índice Sintético de Fecundidade

ARMA Modelo Misto Autoregressivo e Médias Móveis

ARIMA Modelo Misto Integrado Autoregressivo e Médias Móveis

BIC Critério Bayesiano de Informação Akaike

FAC Função de Autocorrelação

FACP Função de Autocorrelação Parcial

HMD *Human Mortality Database*

IDD Índice de Dependência de Idosos

INE Instituto Nacional de Estatística, I.P.

PIB Produto Interno Bruto

SVD Decomposição de Valores Singulares

1.1 Tema

Embora as pessoas vivam cada vez mais e tendencialmente tenham menos filhos, as experiências demográficas repetem-se geracionalmente com padrões altamente regulares. Padrões esses que só lentamente se vão modificando ao longo do tempo. Por estas razões é possível projetar a população por longos períodos (Lee, 1998). Não obstante, o futuro demográfico de qualquer país é sempre marcado pela incerteza.

Nesta dissertação apresenta-se uma projeção puramente estocástica para a população portuguesa com o horizonte temporal de 2050. Contrariamente às projeções convencionais, baseadas em modelos determinísticos assentes em vários cenários para refletir a incerteza, as projeções estocásticas, permitem a associação de distribuição de probabilidades à projeção das taxas demográficas assim como a construção de intervalos de confiança quantificando a incerteza associada.

A importância das projeções estocásticas da população tem sido frequentemente sublinhada na literatura científica. No encontro científico, *How to deal with uncertainty in population forecasting*, chegou-se ao seguinte consenso entre os participantes - *We believe that the quantification of uncertainty will enhance the usefulness of population projections and make the work of forecasting agencies an even more valuable product for planners, policy-makers, scientists and the public around the world* (Scherbov et al., 2006).

Antes de avançar, importa distinguir dois conceitos, o da projeção e da previsão. Ambos assentam em pressupostos, numa previsão os pressupostos representam as expectativas do demógrafo em relação ao futuro, numa projeção os pressupostos em relação ao futuro não são os desejáveis mas sim os mais prováveis. Nesta dissertação os pressupostos fundamentam-se através do comportamento passado das séries demográficas analisadas, nomeadamente a mortalidade, a fertilidade e os movimentos migratórios.

Todavia, as projeções da população são baseadas em grandes pressupostos que muito raramente são explicitados. As projeções assumem que não vai haver nenhuma catástrofe no futuro, como por exemplo uma guerra nuclear ou uma colisão com um cometa. Também raramente incorporam covariáveis em relação a mudanças mais

previsíveis, como é o caso do aquecimento global e a falta de água potável. No geral as projeções assumem que não irá ocorrer nenhuma mudança estrutural profunda e, desta forma, extrapolam o passado e esperam que o futuro seja semelhante (Lee and Tuljapurkar, 2000). Todo o tipo de projeções a longo prazo está associado a um elevado grau de incerteza, as projeções demográficas não são exceção .

As projeções populacionais são cruciais para a tomada de decisões no campo social, económico e político tal como o financiamento dos sistemas de pensões e de saúde, o desenvolvimento do mercado de trabalho assim como o planeamento dos sistemas de educação (Härdle and Mysickova, 2009), possibilitando informação de grande relevância na prossecução de políticas públicas e na definição das estratégias nacionais e europeias.

Os objetivos propostos passam pela operacionalização de um modelo de projeções estocásticas para a população portuguesa para 2050, por sexo e idades, mais especificamente pela projeção estocástica das componentes da mortalidade, fertilidade e migração. Pela realização de uma revisão literária das projeções aplicadas ao caso português, entende-se que é a primeira vez que são aplicadas projeções puramente estocásticas.

1.2 Organização

No capítulo 1 desta dissertação é apresentada a introdução da mesma, sendo expressa a intenção de analisar e modelar estocasticamente as séries demográficas portuguesas com o objetivo de proporcionar uma projeção estocástica para a população portuguesa.

No capítulo 2 discutem-se as principais características e tendências da demografia portuguesa, iniciando a análise com a evolução da população portuguesa e prosseguindo-se com o contributo do saldo natural e do saldo migratório. Toda a informação estudada neste capítulo foi disponibilizada pelo Instituto Nacional de Estatística, I.P. (INE).

O capítulo 3 apresenta uma revisão histórica das projeções assim como o tipo de projeções existentes e as vantagens das projeções estocásticas sobre as projeções determinísticas.

O capítulo 4 expõe as metodologias utilizadas na projeção das taxas demográficas, nomeadamente a mortalidade e a fertilidade, sendo que relativamente à componente migratória não se apresenta nenhuma metodologia por se ter optado pela construção de projeções que derivam diretamente da série histórica. É também apresentado o processo estocástico de renascimento aplicado à população assim como a metodologia de Box-Jenkins.

Finalmente no capítulo 5 é aplicada a metodologia apresentada e desenvolvida anteriormente.

Ao longo do texto são feitas comparações com os resultados das projeções oficiais para 2050 realizadas pelo INE (INE, I.P., 2003) e também com os resultados de um exercício realizado pelo Eurostat - EUROPOP 2004 - que constrói projeções da população para os estados membros (Eurostat, 2004). Ambas as projeções se baseiam na construção de cenários, e para a comparação utiliza-se sempre o cenário base.

A informação foi tratada e validada no Microsoft Office Access, a análise e a implementação dos modelos foi feita no *software* estatístico R. Por fim, na realização dos gráficos, na construção das tábuas de mortalidade e no cálculo da projeção da população utilizou-se o Microsoft Office Excel. Para todos os testes estatísticos utilizados, definiu-se à partida, o nível de significância de 5%.

Naturalmente, no capítulo 6 são expostas as conclusões e outras considerações retiradas desta dissertação.

Dinâmica demográfica recente em Portugal

2.1 A evolução da população residente

A redução generalizada das taxas de fecundidade e a consequente diminuição da natalidade, por um lado, e a estabilização dos níveis de mortalidade nas faixas etárias mais baixas assim como a sua atenuação nos grupos etários mais avançados, prolongando a esperança de vida dos cidadãos, por outro lado, consubstanciam-se numa aceleração do ritmo de envelhecimento da população que pode ser agravado por movimentos imigratórios líquidos pouco significativos.

Analisando a evolução da população através das pirâmides etárias de 1970 e de 2009 (figura 2.1), observa-se que Portugal tem sofrido em decréscimo de população nas idades mais jovens e por outro lado, um aumento considerável de população idosa. De facto, em termos médios a percentagem de população entre os 0 e os 14 anos de idade decresceu cerca de 34% e a população com idades superiores a 65 anos aumentou mais do dobro.

Em termos gerais Portugal experienciou um processo de duplo envelhecimento no topo, caracterizado por uma maior longevidade nas idades mais avançadas, e na base, reflexo de uma descida das taxas de fecundidade. Este perfil coloca o país na mesma linha dos padrões demográficos que caracterizam as sociedades mais desenvolvidas, incluindo algumas congéneres europeias (Pappámikail *et al.*, 2010).

2.2 Contributo da componentes natural

Ao longo dos últimos 40 anos o saldo natural (figura 2.2) de Portugal tem, em termos globais, sofrido um decréscimo acentuado embora acompanhado de algumas oscilações, refletindo tanto uma diminuição do número de nados-vivos como também do número de óbitos. O saldo natural que em 1970 era de 120 mil indivíduos atingiu, pela primeira vez na história demográfica portuguesa recente, um saldo natural de valor negativo em 2007, consequência da conjugação de 102492 nados vivos filhos de mães e de 103512 óbitos de indivíduos residentes em

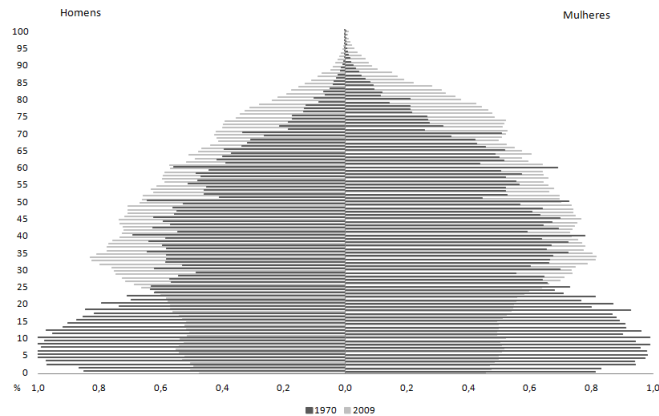


Figura 2.1: Pirâmides etárias da população portuguesa, 1970 e 2009

Portugal. Todavia, o saldo total manteve-se positivo pelos ganhos populacionais obtidos por intermédio do saldo migratório.

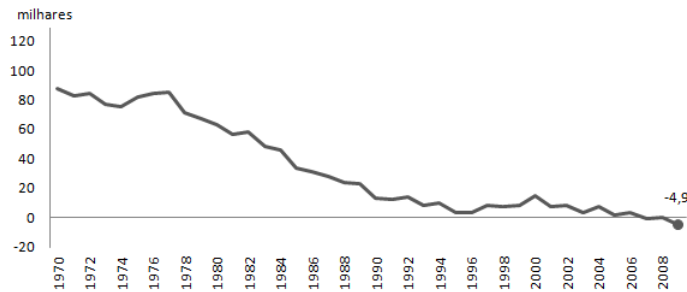


Figura 2.2: Saldo natural, 1970-2009

2.2.1 Mortalidade

No século XX assistiu-se a uma descida acentuada da mortalidade em todos os países desenvolvidos, Portugal não foi exceção. Os progressos na assistência médica e na qualidade de vida têm permitido, de um modo geral, a estabilização da mortalidade nas faixas etárias mais baixas e o aumento da esperança média de vida (INE, I. P., 2010). De facto, a esperança média de vida à nascença aumentou de 67 anos em 1970 para se encontrar no período de 2009 nos 79 anos. Simultaneamente ao aumento sem precedentes da longevidade da população, e à medida que a esperança média de vida aumentou, a desigualdade entre os sexos acentuou-se (figura 2.3).

Segundo Oliveira and Mendes (2010) esta desigualdade entre sexos na esperança média de vida à nascença está associada a múltiplos fatores de natureza biológica e social: o trabalho feminino tende a acontecer em profissões de menor risco para a saúde, o consumo de tabaco e álcool é mais moderado, a condução mais prudente e, por fim, a relação com a medicina é mais próxima.

Já em termos estruturais e tendo em conta a distribuição da população pelos vários grupos etários, ao aumento da esperança média de vida está associado uma tendência de aumento da população idosa (indivíduos com mais

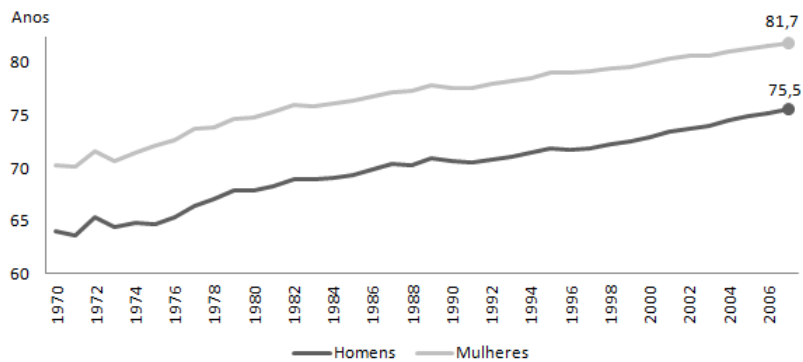


Figura 2.3: Esperança de vida à nascença por sexo, Portugal, 1970-2009

de 65 anos de idade) e paralelamente uma diminuição da população jovem (indivíduos com idades compreendidas entre os 0 e os 14 anos de idade) (INE, I. P., 2010). Em apenas 40 anos o índice de envelhecimento¹ cresceu de 33,9% em 1970 para 117,6% em 2009 (figura 2.4) anunciando um Portugal, com uma população cada vez mais envelhecida e dependente das gerações mais novas o que é revelador de dificuldades a exigências de desenvolvimento económico e social sustentável.

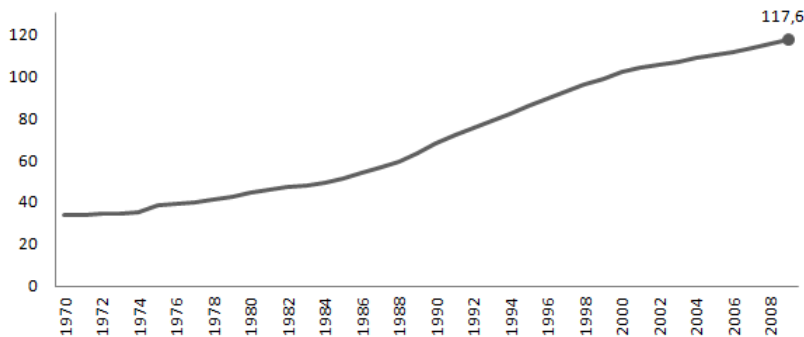


Figura 2.4: Índice de envelhecimento em Portugal, 1970-2009

À semelhança do índice de envelhecimento, também o índice de longevidade² registou um acréscimo sustentado ao longo dos últimos 40 anos. Com efeito, a proporção de indivíduos com 75 ou mais anos de idade no total de indivíduos com 65 ou mais anos que era de 32,7% em 1970, alcançando o valor de 47,4% em 2009 (figura 2.5).

A tendência demográfica de envelhecimento tem impacto em vários planos da vida social, como é o caso do sistema de proteção social e do mercado de trabalho. O crescente número de idosos exige um esforço adicional por parte da população ativa que, por sua vez, constitui um ativo fundamental para a manutenção de um sistema de segurança e proteção social que garanta um conjunto alargado de apoios aos idosos. Adicionalmente, o decréscimo no efetivo populacional mais jovem, apesar de, no curto prazo, poder ser lido como um fator amortizador da

¹Como índice de envelhecimento entende-se a relação entre a população idosa e a população jovem, definida como o quociente entre o número de pessoas com 65 ou mais anos de idade e o número de pessoas com idades compreendidas entre os 0 e os 14 anos de idade.

²Como índice de longevidade entende-se a relação entre a população mais idosa e a população idosa, definida como o quociente entre o número de pessoas com 75 ou mais anos de idade e o número de pessoas com 65 ou mais anos de idade.

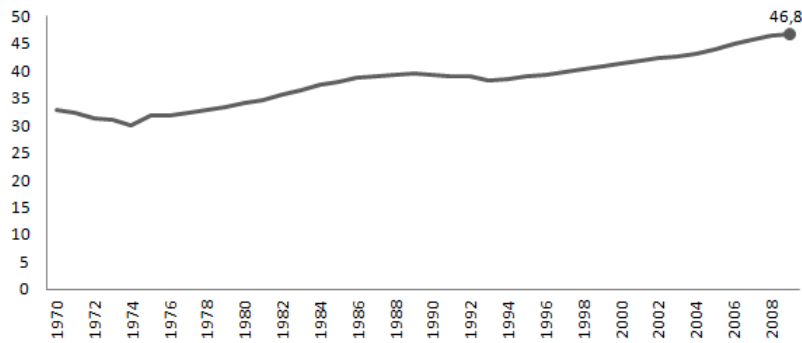


Figura 2.5: Índice de longevidade em Portugal, 1970-2009

pressão que recai sobre a população ativa (Pappámikail *et al.*, 2010), a médio e a longo prazo, não apresenta um impacto positivo no sentido em que a geração mais nova é aquela que possibilita a renovação, numa ótica de sustentabilidade, da população ativa.

2.2.2 Fertilidade

Ao longo das últimas décadas, observou-se uma queda acentuada da fecundidade (Almeida *et al.*, 2002). O primeiro declínio da fecundidade registado remonta à segunda metade dos anos 20, seguindo-se um período de relativa estabilidade para, nos anos 60 e 70 voltar a verificar-se uma nova tendência de decréscimo (Oliveira, 2008). Esta nova tendência, foi referida pela primeira vez pelos demógrafos Lesthaeghe and Van de Kaa (1986) como a segunda transição demográfica. Esta fase caracteriza-se pela manutenção da fecundidade abaixo do nível de substituição de gerações³, pelo adiamento do casamento e dos nascimentos, pelo aumento da coabitação e dos nascimentos fora do casamento, bem como pela intensificação do divórcio (Bandeira, 1996a; Oliveira, 2008).

Em Portugal, no espaço de aproximadamente 50 anos o ISF passou de 3,1 (em 1960) para 1,3 (em 2009), verificando-se uma redução em mais de metade do número médio de filhos nascidos por mulher em idade fértil (INE, I. P., 2010). Há mais de 25 anos que Portugal permanece abaixo do limiar do nível de substituição de gerações (figura 2.6), tendo sido em 1982 que tal deixou de ser assegurado e atingindo em 2009 o valor de 1,3.

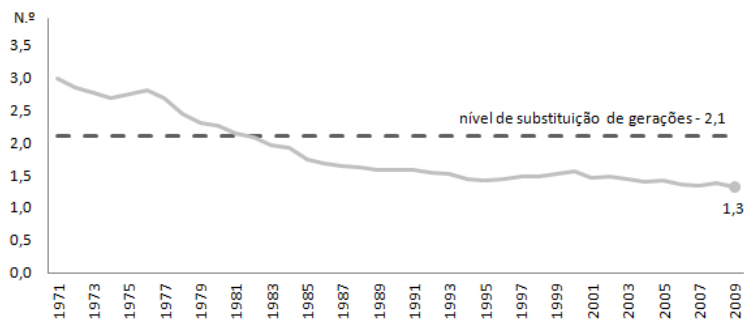


Figura 2.6: Índice Sintético de Fecundidade em Portugal, 1970-2009

Contudo as transformações não se esgotam com a queda acentuada do ISF, sendo acompanhadas por outra

³Decreta-se o nível de substituição de gerações em 2,1 crianças, pois um dos filhos reporiria a mãe, o outro o pai e a diferença de 0,1 seria para compensar a taxa de mortalidade infantil.

tendência, que se prende com o adiamento de calendários da fecundidade. A transição para a parentalidade tende a acontecer cada vez mais tarde no percurso de vida dos indivíduos (Ferreira and Nunes, 2010), e tal tem diretamente impacto no momento em que este evento acontece ao longo do período de fertilidade feminino. Este reflete-se, nomeadamente, num potencial aumento da maternidade tardia (INE, I. P., 2010). De facto, o comportamento no que respeita aos calendários de fecundidade apresenta-se num sentido homogéneo de aumento da idade média das mulheres aquando do nascimento do primeiro filho, sendo desde 1983 estritamente crescente (figura 2.7).

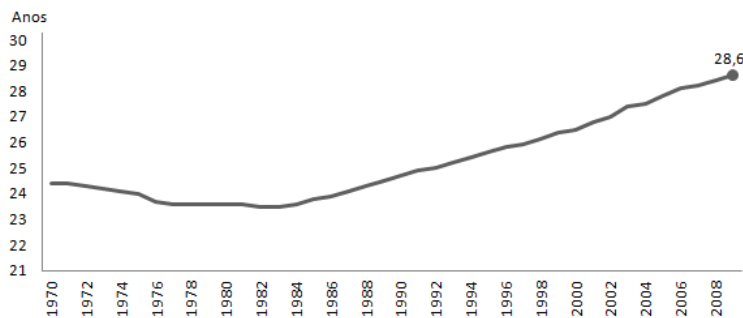


Figura 2.7: Idade média da mãe ao nascimento do primeiro filho em Portugal, 1970-2009

Considerando também os nascimentos por idade da mãe, é possível obter um indicador que coloca em evidência as parcelas de nascimentos que ocorrem nas idades mais tardias do período de fecundidade feminino. A maternidade tardia (mães com idade igual ou superior a 35 anos) tem vindo a registar um aumento desde o início da década de 90 (Almeida *et al.*, 2002; Cunha, 2007). Importa salientar que, a par das questões de adiamento dos calendários de fecundidade, a tendência de crescimento da maternidade tardia não é alheia à disponibilização de meios de diagnóstico mais eficazes e seguros, que permitem reduzir grandemente potenciais situações de risco em gestações tardias. Por comparação, a maternidade precoce (mães com idade igual ou inferior a 20 anos) apresenta-se em declínio (figura 2.8).

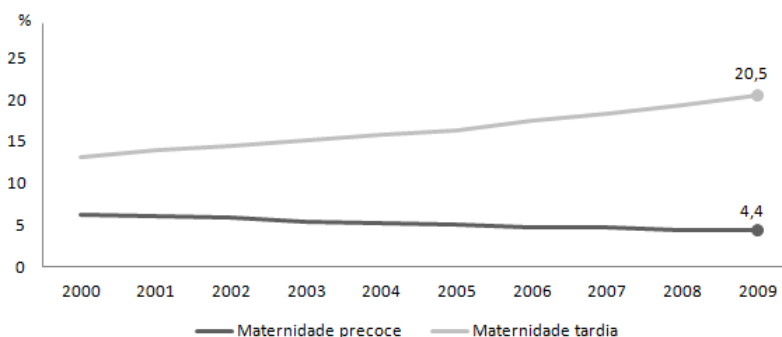


Figura 2.8: Maternidade precoce e maternidade tardia

2.3 Contributo da componente migratória

As migrações constituem um ativo importante para o desenvolvimento dos países, quer ao nível do seu potencial contributo demográfico (Coleman, 2006; Haug *et al.*, 2002), quer ao nível do reforço que podem significar em termos de renovação da força de trabalho, colmatando, em muitos contextos, não só o défice de nascimentos mas também as necessidades do mercado de trabalho (INE, I. P., 2010).

De facto, tendo como pano de fundo um dos principais fundamentos das teorias das migrações, nomeadamente no plano da nova perspectiva económica, os indivíduos procuram deslocar-se para contextos onde perspetivam mais oportunidades e, conseqüentemente, um retorno positivo face ao investimento e expectativas que são depositadas na migração (Castles and Miller, 2003; Massey *et al.*, 1993). Assim, se numa primeira fase as migrações são maioritariamente masculinas, pode seguir-se numa fase seguinte o reagrupamento familiar o que pode originar uma transferência de nascimentos do país de origem para o país de acolhimento (Carrilho and Lurdes, 2002).

Neste sentido, o impacto da imigração no plano demográfico em Portugal tem revelado contribuições, não só ao nível do seu peso no crescimento do efetivo populacional, mas também por via do reforço que representam nas idades mais jovens, com destaque ao nível das idades fecundas, o que se traduz em maiores índices sintéticos de fecundidade (Rosa *et al.*, 2004).

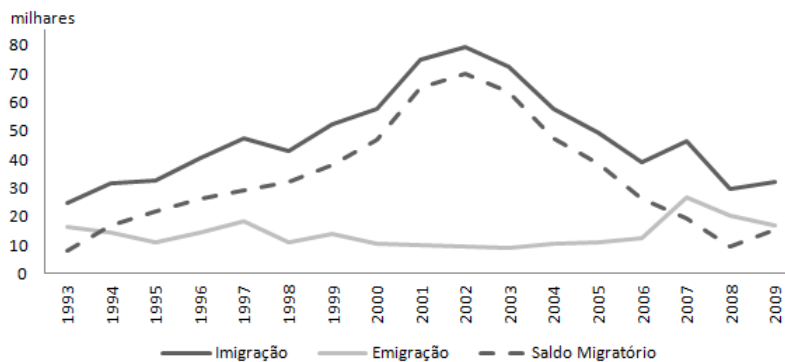


Figura 2.9: Imigração, emigração e saldo migratório em Portugal, 1993-2009

Portugal, que nos anos 60 era um país com uma forte emigração externa, tem assistido desde essa altura até ao início do presente século, à redução progressiva dos fluxos emigratórios devido ao registo de um movimento de regresso dos emigrantes (Conim, 1996). Nos anos mais recentes a emigração sofre ligeiras variações, encontrando-se o seu máximo em 2007 com cerca de 27 mil emigrantes. Por outro lado a imigração sofreu grandes oscilações no período entre 1993 e 2010 (figura 2.9). Sendo até 2004 constantemente crescente, sofreu a partir desse ano uma descida gradual, encontrando-se em 2009 cerca de 32 mil imigrantes. Caso esta tendência continue, o saldo migratório, que nos anos mais recentes compensou um saldo natural negativo, pode tornar-se ele próprio negativo.

Principais metodologias de projeções populacionais

3.1 Revisão de literatura

Em 1895 os censos começavam a fornecer informação atual e fiável da população estruturada por sexos e idade em muitos países europeus. Foi neste ano que, pela primeira vez, se realizaram projeções da população com base em informação censitária (Gans, 1999). O autor deste feito foi o economista inglês Edwin Cannan que projetou a população de Inglaterra e do País de Gales para o horizonte temporal de 1991.

Edwin Cannan formulou a conhecida e ainda hoje utilizada metodologia Coorte-Componente que parte de uma população de inicial desagregada por sexo e idade e calcula a população futura de cada geração (coorte) aplicando, de forma individual, os pressupostos sobre a fertilidade, mortalidade e migrações.

Já na década de 20 realizaram-se muitas outras projeções de carácter determinístico e quase sempre ao nível territorial do país. Segundo Alho and Spencer (2005) uma das razões para o interesse súbito nas projeções da população, nesta década, parece ser a declínio da fertilidade, com particular foco nas grandes cidades da Europa e dos Estados Unidos da América. Destaque-se Tarasov e Strumilin que em 1922 projetaram a população da União Soviética, Wielbols em 1925 a população da Holanda, Wicksell em 1926 a população da Suécia e em 1928 Sauvy projetou a população francesa, e Whelpton a população dos Estados Unidos da América de 1925 a 1975.

A dificuldade com o método utilizado assenta na volatilidade das componentes demográficas, por exemplo, antes da segunda guerra mundial acreditava-se que a fertilidade iria baixar de tal forma, que o baby boom tomou os demógrafos de surpresa. Whelpton reagiu à volatilidade das componentes demográficas com a criação de vários cenários para a projeção, usualmente cenário alto, base e baixo (Alho, 1999).

É de referir também a contribuição de Patrick Leslie que em 1945 formalizou a método em termos matemáticos, com a construção de matrizes representativas da população (O'Neill *et al.*, 2001) e conhecidas ainda hoje como matrizes de Leslie. Já em 1949 realizavam-se, pela primeira vez, as projeções mundiais da população sobre a supervisão de Notestein, fundador da teoria da transição demográfica Bandeira (1996b).

A primeira tentativa de explicação das projeções da população de um ponto de vista estocástico ocorreu com o estatístico Leo Törnqvist na produção da projeção da população finlandesa (Alho and Spencer, 2005). Segundo Wilson and Rees (2005) a quantificação da incerteza através da construção de projeções estocásticas da população pode ser considerado como o desenvolvimento mais importante na passada década.

De facto, no passado recente, foi muito o interesse dos investigadores nas projeções probabilísticas, registese as projeções para os Estados Unidos da América (Alho e Spencer em 1986, Plaufmer em 1988 e Lee and Tuljapurkar em 1994), para a Áustria por (Hanika em 1997 e Lutz e Scherbov em 1998), para a Finlândia (Alho em 1998), para a Alemanha (Lutz and Scherbos em 1998 e Härdle em 2009), para a Holanda (de Beer em 1997) e para a Noruega (Keilman em 2002).

3.2 Tipo de projeções

Historicamente, as projeções produzidas pelos institutos oficiais de estatística, assentam em modelos determinísticos utilizando diferentes cenários para o futuro de forma a descrever a incerteza associada.

Estas projeções partem da aceitação de um conjunto de hipóteses estabelecidas sobre a evolução da fecundidade, mortalidade e migrações que se conjugam em diferentes cenários. A validade de uma projeção depende da verificação ou não das hipóteses que são previamente estabelecidas e que se podem ou não tornar realidade (Carrilho, 2005). Este método permite uma análise prospetiva de cada uma das componentes através de cenários que incorporam não apenas os determinismos do passado mas também parâmetros possíveis de evolução.

Sendo cada um dos cenários um jogo de hipóteses coerente e dado que as hipóteses delineadas para cada componente são claramente explicitadas, este método possibilita a análise de evoluções alternativas para cada componente, assumindo determinadas tendências no que se refere à mortalidade, fecundidade e migrações, bem como a observação dos seus impactos, por forma a reduzir a incerteza naturalmente associada aos cálculos envolvidos. Os resultados são assim sempre condicionais, dependendo da especificação dos parâmetros estabelecidos para cada cenário (INE, I.P., 2003).

Uma alternativa aos cenários, como forma de caracterizar a incerteza associada às projeções, é considerar explicitamente a incerteza associada às tendências projetadas das componentes demográficas fertilidade, mortalidade e migração, e assim derivar as distribuições de probabilidade para a dimensão da população estruturada por sexo e grupo etário. Vão-se discriminar os três principais métodos que possibilitam a determinação das probabilidades associadas às principais taxas demográficas: parecer de um especialista, métodos estatísticos de séries temporais e análise *a posteriori* de erros de anteriores projeções da população (O'Neill *et al.*, 2001).

A análise *a posteriori* de erros em projeções passadas pode ser utilizada como uma base para gerar intervalos de incerteza em torno de uma projeção produzida por outros meios. Este método assenta no pressuposto de que as projeções atuais estão sujeitas a erros similares àqueles cometidos no passado. Stoto (1983) e Keifitz (1981) compararam projeções da população anteriores realizadas pelas Nações Unidas e pelo gabinete dos censos dos Estados Unidos da América ¹ com os movimentos observados da população, e a partir destas comparações desenvolveram erros padrão para a projeção da taxa de crescimento populacional. Estes erros padrão podem ser utilizados para definir intervalos probabilísticos para as projeções futuras da população, com o pressuposto de que o futuro vai ser tão fácil ou difícil de prever como foi o passado (Lee, 1998), ou seja, assume-se que a distribuição de probabilidades de uma dada projeção tem as características essenciais dos erros observados no passado, sendo que tem a desvantagem de limitar o horizonte temporal da projeção atual ao horizonte temporal da projeção

¹Nota de tradução: United States Bureau of the Census

passada (O'Neill *et al.*, 2001).

Os modelos de séries temporais têm sido recentemente utilizados na projeção da população, em particular ao nível geográfico do país. A ideia é desenvolver modelos baseados nas principais taxas demográficas mortalidade, fertilidade e migração, fazer a sua projeção individualmente e posteriormente aplicar as taxas obtidas a uma população de partida.

3.3 Um discurso sobre a vantagem das projeções probabilísticas

A quantificação da incerteza pode ser considerada como o desenvolvimento mais importante na área das projeções da população na década de 90 (Wilson and Rees, 2005), este facto é apontado por diversos autores como a característica de excelência das projeções probabilísticas. Esta metodologia fornece a população futura por sexo e por grupos etários não apenas como um número mas sim como uma série de distribuição de probabilidades (Keilman, 2008).

Já as projeções tradicionais lidam com a incerteza das projeções da população através da construção de cenários (Keilman *et al.*, 2002) que dependem de pressupostos de evolução das componentes demográficas. Com frequência obtêm-se um cenário base e outros que correspondem a desenvolvimentos da população mais otimistas ou pessimistas. Estamos portanto perante uma abordagem determinística. McNown *et al.* (1993) desenvolveu esta abordagem escolhendo cenários baseados em modelos estocásticos, contudo a estrutura lógica inerente das projeções não se altera. Segundo Lee and Tuljapurkar (1994) os graus de incerteza associados às principais taxas demográficas, números de eventos, dimensão dos grupos etários e à dimensão total da população são exatamente os mesmos.

As projeções probabilísticas oferecem uma boa alternativa face às projeções determinísticas. Ao invés das projeções determinísticas possibilitam a associação de probabilidades à projeção, tanto da população como das principais taxas demográficas. O futuro é com certeza incerto, mas alguns desenvolvimentos demográficos são mais prováveis que outros (Keilman, 2008). Já nas projeções determinísticas não existe nenhuma probabilidade associada aos cenários considerado e, do ponto de vista do utilizador este fica sem qualquer sugestão de qual o cenário mais provável (Bertino *et al.*, 2010), tomando-o tipicamente como o cenário base (Alho, 1999).

Por outro lado, a utilização de cenários mais otimistas ou pessimistas é irrealista e inconveniente do ponto de vista estatístico (Lee, 1998). Num cenário otimista a fertilidade é assumida como elevada em todo o horizonte temporal que se irá projetar, analogamente se estamos perante um cenário pessimista a fertilidade é sempre tida como baixa. Quando a projeção é acompanhada da mortalidade, os cenários ainda se tornam mais extremos. Neste caso, um cenário com elevado crescimento da população é obtido através da combinação de elevada fertilidade com baixa mortalidade (maior esperança de vida), o que significa que em todo o horizonte temporal da projeção se a fertilidade é alta, a esperança média de vida também o é (Keilman *et al.*, 2002). Como consequência as várias projeções assumem valores extremados, bastante distantes entre si. Em termos estatísticos assume-se uma perfeita mas irrealista correlação entre as componentes demográficas, mais, as projeções para a dimensão dos grupos etários, dimensão da população e rácios entre grupos etários não são consistentes entre si (Hårdle and Mysickova, 2009). Uma projeção probabilística não assume uma correlação perfeita entre as componentes demográficas fertilidade, mortalidade e migração.

Assim para encontrarmos um intervalo para a dimensão da população ou para a taxa de crescimento tem de se combinar a trajetória de elevada fertilidade e imigração com baixa mortalidade, e vice versa. Consequentemente o intervalo do Índice de Dependência de Idosos (IDD) irá ser bastante estreito, porque elevada fertilidade diminui

o IDD enquanto que baixa mortalidade o eleva (Lee, 1998).

Outro problema das projeções determinísticas é a probabilidade de cobertura ², que para os cenários baixo a elevado aumenta rapidamente com o horizonte temporal da projeção. Isto acontece porque este tipo de projeções assume implicitamente uma correlação perfeita ao longo do tempo. Ou seja, se no cenário elevado, a fertilidade é assumida como alta para determinado ano, então com toda a certeza que também vai ser alta no ano seguinte, e analogamente para o cenário baixo. Nas projeções probabilísticas isto não acontece (Keilman, 2008).

²A probabilidade de cobertura de um procedimento para encontrar intervalos de confiança define-se como a chance do procedimento produzir um intervalo que cobre a verdade.

Metodologia para a realização de projeções estocásticas

4.1 Mortalidade

O modelo demográfico homónimo de Lee Carter desenvolvido em 1992, para projetar as taxas de mortalidade dos Estados Unidos da América para 2065, tem sido até aos dias de hoje um método largamente utilizado por demógrafos e estatísticos na elaboração de projeções estocásticas a médio e longo prazo de taxas de mortalidade e consequentemente da esperança média de vida. Desde o seu desenvolvimento que têm sido inúmeras as extensões e as modificações ao método assim como a aplicação a diversos países para diferentes períodos de tempo: ao Japão por Wilmoth (1996), aos países que constituem o G7 por Tuljapurkar *et al.* (2000), à Suécia por Wang (2007), à Bélgica por Brouhns *et al.* (2002) e também a Portugal por Coelho (2001), entre outros.

4.1.1 Modelo de Lee Carter

O modelo de Lee Carter assenta na combinação de um modelo demográfico com um modelo de séries temporais. Seja m_{at} a matriz das taxas de mortalidade específicas à idade a ($a = 1, \dots, A$), no ano t ($t = 1, \dots, T$) de Portugal. O primeiro passo no método de Lee Carter consiste em separar a componente dependente do tempo (modelo de séries temporais) da componente específica das idades (modelo demográfico), modelando os dados como (Lee and Carter, 1992):

$$\ln(m_{at}) = \alpha_a + \beta_a \gamma_t + \epsilon_{at} \quad (4.1)$$

onde α_a e β_a são parâmetros específicos da idade, γ_t é o parâmetro que varia ao longo do tempo e ϵ_{at} representa os erros aleatórios associados à série histórica, com valor esperado 0 e variância σ_ϵ^2 . Os erros refletem a existência de uma variação irregular entre as idades e o tempo, que deriva principalmente de circunstâncias históricas e que não é capturada pelo modelo (Härdle and Mysickova, 2009). O modelo é sobreparametrizado, visto ser invariante em relação às seguintes transformações (Giroi and King, 2007):

$$\begin{aligned} \beta_a &\rightsquigarrow c\beta_a & \gamma_t &\rightsquigarrow \frac{1}{c}\gamma_t & \forall c \in \mathbb{R}, c \neq 0 \\ \alpha_a &\rightsquigarrow \alpha_a - \beta_a c & \gamma_t &\rightsquigarrow \gamma_t + c & \forall c \in \mathbb{R} \end{aligned}$$

Desta forma torna-se necessário definir restrições aos parâmetros de forma a encontrar uma solução única. As restrições impostas passam por assumir que a soma dos coeficientes β_a somam um ao longo das idades ($\sum_a \beta_a = 1$) e que a soma dos coeficientes γ_t somam zero ao longo do tempo ($\sum_t \gamma_t = 0$). O último pressuposto implica que o parâmetro α_a seja simplesmente a média empírica de $\ln(m_{at})$ ao longo do tempo para cada idade a . Atente-se à seguinte demonstração:

$$\begin{aligned} \sum_t \alpha_a &= \sum_t \ln(m_{at}) - \sum_t \beta_a \gamma_t & (4.2) \\ &= \sum_t \ln(m_{at}) - \beta_a \cancel{\sum_t \gamma_t}^0 \end{aligned}$$

O que leva a que α_a seja dado por:

$$T\alpha_a = \sum_t \ln(m_{at}) \Rightarrow \alpha_a = \frac{1}{T} \sum_t \ln(m_{at}) \quad (4.3)$$

Assim, o parâmetro α_a descreve a forma geral das taxas de mortalidade para cada idade, o parâmetro γ_t representa o índice de nível geral da mortalidade, ou seja, capta as principais tendências temporais da mortalidade para cada ano t , e por fim, o parâmetro β_a descreve as mudanças nas taxas de mortalidade à idade a , em resposta às mudanças no índice geral da mortalidade γ_t ($\frac{\partial \ln(m_{at})}{\partial t} = \beta_a \frac{\partial \gamma_t}{\partial t}$).

Quando o parâmetro β_a é elevado para determinada idade a , então a taxa de mortalidade nessa idade varia consideravelmente quando o índice geral de mortalidade, γ_t , sofre alterações, por outro lado, quando o parâmetro β_a assume um valor baixo para certa idade, as taxas de mortalidade nessa idade sofrem pequenas oscilações em resposta a mudanças de γ_t (Russolillo, 2005).

Um pressuposto deste modelo é o de as taxas de mortalidade específicas por idade aumentarem ou diminuírem todas juntas, visto todas dependerem do índice geral de mortalidade γ_t . Para se garantir que este pressuposto é cumprido, o parâmetro β_a não pode assumir valores negativos em nenhuma das idades, pois se tal acontecer, obtem-se uma dualidade da mortalidade. Na prática isto parece não acontecer, como é afirmado pelos próprios Lee and Carter (1992) e também por Girosi and King (2007). Por outro lado uma das grandes vantagens surge na construção do índice geral de mortalidade γ_t e na posterior projeção deste, visto que, se γ_t tende para infinito negativo, então cada taxa específica por grupo etário tende para zero, impossibilitando a ocorrência de taxas de mortalidade negativas (conforme a equação 4.1).

4.1.1.1 Estimação

Como já foi supramencionado o parâmetro α_a é estimado como a média empírica ao longo do tempo para cada grupo etário. Por outro lado os parâmetros β_a e γ_t podem ser estimados através de várias abordagens, como é o caso da Decomposição de Valores Singulares (SVD)¹, da estimação de máxima versomilhança ou do

¹Acrônimo derivado do inglês: Singular Value Decomposition.

método dos mínimos quadrados. Optou-se por empregar a SVD para estimar β_a e γ_t , por um lado por ter sido a abordagem utilizada no artigo original de Lee e Carter, e por outro, porque segundo Koissi and Högnäs (2005), que realizaram uma comparação das estimativas dos parâmetros do modelo, obtido com estas três formas de estimação, a SVD é a melhor alternativa para a estimação do índice de nível geral de mortalidade γ_t . Koissi and Högnäs (2005) construíram matrizes de erros aleatórios utilizando a técnica de bootstrap para cada um dos métodos e compararam-nos posteriormente através do erro quadrático médio.

Considere-se a matriz M que é definida em termos dos logaritmos das taxas de mortalidade específicas por grupos etários centradas em torno da média

$$M = \ln m_{at} - \alpha_a = \beta_a \gamma_t \quad (4.4)$$

De forma a estimarem-se os parâmetros β_a e γ_t emprega-se a SVD à matriz M . A SVD é um método de factorização que pode ser aplicado a qualquer matriz (real ou complexa) e que tem como principal objetivo a redução da dimensão dos dados assim como a retenção da maioria da variabilidade. Desta forma, obtêm-se a seguinte equação:

$$M = \Gamma \Lambda \Delta^T \simeq \Gamma_{A1} (\Lambda_{11} (\Delta^T)_{1T}) \underbrace{\sum_a \beta_a}_{\hat{\beta}_a} \underbrace{\sum_a \gamma_a}_{\hat{\gamma}_t} \quad (4.5)$$

onde Γ é uma matriz de dimensão $(A \times A)$, Λ é uma matriz diagonal de dimensão $(A \times T)$ e Δ é uma matriz de dimensão $(T \times T)$. O parâmetro β_a e o índice γ_t obtêm-se retendo apenas o primeiro valor próprio, partindo do pressuposto que este contém a maioria da variabilidade da matriz M . Assim, a estimativa de β_a define-se como a primeira coluna normalizada da matriz Γ , e a estimativa do índice γ_t define-se pela multiplicação do primeiro valor próprio com a primeira coluna da matriz Δ , também normalizada. Note-se que as restrições feitas inicialmente aos dois parâmetros, β_a e γ_t , são tidas em conta na estimação.

4.1.1.2 Reestimação

Sendo a primeira estimação do índice de nível geral da mortalidade baseada nos logaritmos das taxas de mortalidade ao invés das próprias taxas de mortalidade, podem ocorrer discrepâncias consideráveis entre o número real e o número previsto de mortes. De forma a garantir que as taxas de mortalidade ajustadas conduzem ao verdadeiro número de mortes, existe a necessidade de reestimar γ_t , tomando os parâmetros α_a e β_a estimados inicialmente. Esta reestimação não é sempre utilizada, sendo que se pode passar diretamente para a modelação do índice γ_t como uma série temporal (secção 4.4). Já Wilmoth (1993) propôs uma técnica baseada na decomposição ponderada de valores singulares que elimina a necessidade da reestimação de γ_t .

A reestimação prende-se com a procura iterativa de γ_t para cada ano t , tal que, o número de mortes observado seja igual ao número de mortes esperado, ou seja:

$$\sum_a d_t = \sum_a p_{at} e^{\alpha_a + \beta_a \gamma_t} \quad (4.6)$$

onde d_t diz respeito ao número total de óbitos observados no ano t e p_{at} corresponde à população observada no ano t com idade a .

De seguida apresentam-se os passos a seguir para reestimar o γ_t , onde γ_t^* representa a sua reestimativa e d representa um valor definido à partida:

1. Comparação do número total de óbitos observados ($\sum_a d_{at}$) com o número total de óbitos preditos pelo modelo. ($\sum_a p_{at}e^{\alpha_a+\beta_a\gamma_t}$), para cada ano t .
2. Da comparação podem obter-se três resultados:
 - (a) Se $\sum_a d_{at} < \sum_a p_{at}e^{\alpha_a+\beta_a\gamma_t}$ é necessário diminuir o número de óbitos esperado, ajustando a nova estimativa para ser $\gamma_t^* = \gamma_t(1 - d)$ se $\gamma_t > 0$ ou $\gamma_t^* = \gamma_t(1 + d)$ se $\gamma_t < 0$;
 - (b) Se $\sum_a d_{at} > \sum_a p_{at}e^{\alpha_a+\beta_a\gamma_t}$ é necessário aumentar o número de óbitos esperado, ajustando a nova estimativa para ser $\gamma_t^* = \gamma_t(1 + d)$ se $\gamma_t > 0$ ou $\gamma_t^* = \gamma_t(1 - d)$ se $\gamma_t < 0$;
 - (c) Se $\sum_a d_{at} = \sum_a p_{at}e^{\alpha_a+\beta_a\gamma_t}$, encontrámos o γ_t^* e podemos parar a iteração.
3. Voltar ao passo inicial.

Uma desvantagem da reestimação deve-se à equação 4.6 poder ter zero, uma ou duas soluções para γ_t , dependendo da estimativa de β_a . Um número de soluções diferente de um, deriva do facto dos valores estimados de β_a não terem todos o mesmo sinal, o que, como já se referiu, acontece raramente pois indica que a mortalidade está a aumentar em alguns grupos etários e a diminuir em outros (Giroso and King, 2007).

4.1.2 Projeção e construção da tábua de mortalidade

Estimados os parâmetros relativos ao modelo de Lee Carter o próximo passo é a projeção de γ_t de acordo com a metodologia de Box-Jenkins que é apresentada mais adiante. Nesta secção, relativa à projeção e construção da tábua de mortalidade, assume-se que já se passou pela fase da projeção de γ_t .

Tendo projetado γ_t para determinado horizonte, é possível agora gerar tabelas de mortalidade para certo ano e calcular quaisquer funções que dela derivem, tal como a probabilidade de morte, a probabilidade de sobrevivência e a esperança média de vida. A tabela de mortalidade é um dos métodos mais antigos de representação de dados relativos à sobrevivência de um grupo de indivíduos e segundo Nazareth (2009) para se projetar a mortalidade no futuro é necessário estimar o efeito da mortalidade em todas as gerações que compõem a população inicial. Seja $\hat{\gamma}_{T+l}$ a projeção de γ_t para o ano l . Inicialmente pretende-se projetar as taxas de mortalidade específicas por idades, atente-se à seguinte equação:

$$\dot{m}_{a,T+l} = m_{a,T}e^{\beta_a(\hat{\gamma}_{T+l}-\gamma_T)} \quad (4.7)$$

que permite fazer projeções das taxas de mortalidade tendo em conta as últimas taxas de mortalidade disponíveis $m_{a,T}$. Note-se que:

$$\begin{aligned} \dot{m}_{a,T+l} &= m_{a,T}e^{\beta_a(\hat{\gamma}_{T+l}-\gamma_T)} \\ &= e^{\alpha_a+\beta_a\gamma_T}e^{\beta_a(\hat{\gamma}_{T+l}-\gamma_T)} \\ &= e^{\alpha_a+\beta_a\gamma_T+\beta_a\hat{\gamma}_{T+l}-\beta_a\gamma_T} \\ &= e^{\alpha_a+\beta_a\hat{\gamma}_{T+l}} \end{aligned} \quad (4.8)$$

O que se está a fazer é a substituição simples de t por $T + l$ na equação 4.1, ou seja:

$$\ln(\dot{m}_{a,T+l}) = \alpha_a + \beta_a \dot{\gamma}_{T+l} \quad (4.9)$$

Apuradas as taxas de mortalidade por idade para cada um dos anos projetados, segue-se a construção da tábua de mortalidade. Fixo um ano t , o próximo passo é converter a taxa de mortalidade m_a em probabilidades de morte, q_a . Segundo Renshaw and Haberman (2003) seja f_a o número médio de anos vividos, dentro do intervalo $[a, a + 1[$, para pessoas que morrem nessa idade. Assume-se aqui que $f_a = \frac{1}{2}$ para todas as idades exceto para os indivíduos com menos de um ano de idade (mortalidade infantil), para estes casos utiliza-se 0,15 para o sexo masculino e 0,16 para o sexo feminino. A probabilidade de morte para uma determinada idade define-se por:

$$q_a \cong \frac{m_a}{1 + f'_a m_a}, \quad a = 1, \dots, A \quad (4.10)$$

onde $f'_a = 1 - f_a$. De forma a completar o cálculo das tabelas de mortalidade, seja p_a a probabilidade de sobrevivência desde a idade a à idade $a + 1$. A probabilidade de sobrevivência é dada por:

$$p_a = 1 - q_a \quad (4.11)$$

A partir da probabilidade de morte q_a e de uma constante arbitrária l_0 (neste caso considera-se 100000) a tabela de mortalidade é construída por aplicação recorrente das seguintes equações:

$$\begin{aligned} l_{a+1} &= l_a(1 - q_a), & a = 1, \dots, A \\ d_a &= l_a - l_{a+1} = l_a q_a, & a = 1, \dots, A \end{aligned} \quad (4.12)$$

onde l_a indica o número de sobreviventes com idade a e d_a corresponde ao número de óbitos com idade a . O tempo total vivido pelos elementos da coorte dos sobreviventes entre as idades exatas $[a, a + 1[$ na tábua da mortalidade, são dados por:

$$L_a = (l_a - f'_a d_a), \quad a = 1, \dots, A \quad (4.13)$$

Sendo os anos completos após a idade a , tal que:

$$T_a = \sum_{a=a}^{A-1} L_a \quad (4.14)$$

o que implica que a esperança média de vida à idade a seja dada por:

$$\dot{e}_a = \frac{T_a}{l_a} \quad (4.15)$$

4.2 Fertilidade

4.2.1 Modelo de Lee Carter modificado

Para tratar a componente da fertilidade vai-se empregar o método de Lee Carter com uma ligeira modificação. Seja f_{at} a matriz das taxas de fecundidade específicas por idade fértil da mulher² a ($a = 1, \dots, A$), no ano t ($t = 1, \dots, T$) de Portugal. O modelo de Lee Carter para a fertilidade é dado por:

$$f_{at} = \theta_a + \delta_a \kappa_t + v_{at} \quad (4.16)$$

onde, de forma similar à componente da mortalidade, θ_a e δ_a são parâmetros específicos da idade e independentes do tempo, κ_t é o parâmetro que varia ao longo do tempo e v_{at} diz respeito a erros aleatórios homocedásticos associados à construção do modelo. Pressupõe-se também que $\sum_a \delta_a = 1$ e que $\sum_t \kappa_t = 0$, o que leva a que θ_a seja a média simples de f_{at} para cada idade a ao longo do tempo ($\theta_a = \frac{1}{T} \sum_t f_{at}$).

Deste modo, θ_a descreve a forma geral das taxas de fecundidade para cada uma das idades, κ_t reflete o nível geral da fertilidade no tempo e δ_a representa o quão rápido ou lentamente a fertilidade varia em resposta a mudanças do nível geral da fertilidade.

4.2.1.1 Estimação

Analogamente à estimação dos parâmetros relativos à mortalidade, aplica-se a decomposição de valores singulares à matriz N que se define à custa da taxa de fecundidade específica por idade (f_{at}) em torno do seu valor médio (θ_a), ou seja:

$$N = f_{at} - \theta_a = \Gamma \Lambda \Delta^T \simeq \hat{\delta}_a \hat{\kappa}_t \quad (4.17)$$

Desta forma a estimativa do parâmetro δ_a define-se como a primeira coluna da matriz Γ normalizada, e a estimativa do índice κ_t , também ele normalizado, define-se pela multiplicação do primeiro valor próprio com a primeira coluna da matriz Δ . Note-se que as restrições feitas inicialmente aos parâmetros, δ_a e κ_t , são tidas em conta na sua estimação.

Posteriormente à estimação de δ_a e κ_t , considere-se a equação 4.16 e some-se, para ambos os lados da equação, os parâmetros ao longo das suas idades, ou seja:

$$\begin{aligned} \sum_a f_{at} &= \sum_a \theta_a + \sum_a \delta_a \kappa_t + \sum_a v_{at} \\ F_t &= \Theta + \kappa_t \sum_a \delta_a + \Upsilon \\ F_t &= \Theta + \kappa_t + \Upsilon \end{aligned} \quad (4.18)$$

²A taxa de fecundidade específica por idade define-se como o número de nados-vivos observados durante um determinado período de tempo, normalmente um ano civil, referido ao efetivo médio de mulheres em idade fértil, no idade considerado, desse período. Considera-se que a idade fértil da mulher varia entre os 15 e os 49 anos de idade.

onde $F_t = \sum_a f_{at}$, $\Theta = \sum_a \theta_a$ e $\Upsilon = \sum_a v_{at}$. O parâmetro obtido F_t , corresponde ao ISF ³, ou seja, ao número médio de crianças por mulher. Assim o índice κ_t pode ser interpretado como um desvio da ISF em relação ao termo médio Θ .

Para construir o índice que posteriormente irá ser modelado como uma série temporal, trabalha-se com os valores ajustados do ISF, isto é, $F_t = \Theta + \kappa_t$, onde κ_t oscila em torno de zero (Härdle and Mysickova, 2009). De forma a construir uma projeção da fertilidade demograficamente plausível incorporam-se limites pré-especificados diretamente na modelação do ISF (Lee, 1993). Seja I e S o limite inferior e superior respetivamente, e defina-se o índice de transformação da fertilidade, g_t como:

$$g_t = \ln \frac{F_t - I}{S - F_t} \quad (4.19)$$

Desta forma, a metodologia de Box-Jenkins aplica-se sobre o índice transformado da fertilidade, g_t , de forma a construir-se uma projeção a longo prazo com os respetivos intervalos de confiança. Só posteriormente é que se apura o ISF assim como as taxas de fecundidade específicas por idade, como se vai ver na secção seguinte.

4.2.2 Projeção da fertilidade

Selecionado o processo de séries temporais que melhor se ajusta ao índice g_t e realizada a sua projeção, interessa agora calcular o ISF, assim como o índice de nível geral da fertilidade κ_t e as taxas de fecundidade específicas por idade, para cada ano do horizonte temporal da projeção. Inicialmente é possível obter o ISF através da aplicação da transformação inversa à função que originou g_t , 5.5.

$$F_t = \frac{S \cdot \exp g_t + I}{1 + \exp g_t} \quad (4.20)$$

Note-se que à medida que g_t vai para infinito, F_t converge para o limite superior S e à medida que g_t vai para infinito negativo, F_t converge para o limite inferior I . Devido a esta característica da transformação logística a projeção do ISF, F_t , e os respetivos intervalos de confiança associados cairão dentro destes limites (Härdle and Mysickova, 2009).

Posteriormente ao apuramento de F_t , é possível obter κ_t partindo da equação $F_t = \Theta + \kappa_t$, visto Θ ser um parâmetro constante invariante no tempo. Desta forma a averiguação das taxas de fecundidade específicas por idade para o ano $T + l$ são dadas pela aplicação da seguinte equação:

$$f_{a,T+l} = \theta_a + \delta_a \kappa_{T+l} \quad (4.21)$$

4.3 População

Após a estimação de modelos para as componentes demográficas: mortalidade, fertilidade e migração é possível construir uma projeção estocástica da população. Aplica-se metodologia do Coorte-Componente que parte de uma população de partida desagregada por sexo e idades e calcula a população futura de cada geração,

³O ISF define-se como o número médio de crianças vivas nascidas por mulher em idade fértil (dos 15 aos 49 anos de idade), admitindo que as mulheres estão submetidas às taxas de fecundidade observadas no momento. Resulta da soma das taxas de fecundidade específicas por grupos etários

aplicando as projeções anteriormente realizadas de cada um dos processos populacionais. Realiza-se o cálculo da população passo-a-passo, sendo um método recursivo o ano $t + 1$ depende invariavelmente do ano t .

Para o efeito cria-se uma matriz de probabilidades de sobrevivência e taxas de fertilidade (denominada matriz de Leslie) à qual se adiciona um vetor respeitante ao saldo migratório. Desta forma dada a população para um ano t , a população feminina para o ano $t + 1$ é dada por:

$$\begin{pmatrix} P_{1,t+1}^F \\ P_{2,t+1}^F \\ \vdots \\ \vdots \\ P_{a,t+1}^F \\ \vdots \\ \vdots \\ P_{A,t+1}^F \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & r.F_{15,t} & \dots & r.F_{49,t} & 0 & \dots & 0 \\ M_{1,t}^F & 0 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & M_{2,t}^F & \ddots & \ddots & & & & & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & & & & \vdots \\ \vdots & \vdots & \ddots & M_{a,t}^F & \ddots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & & \ddots & 0 & 0 & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & M_{A-1,t}^F & 0 \end{pmatrix} \begin{pmatrix} P_{1,t}^F \\ P_{2,t}^F \\ \vdots \\ \vdots \\ P_{a,t}^F \\ \vdots \\ \vdots \\ P_{A,t}^F \end{pmatrix} + \begin{pmatrix} SM_{1,t}^F \\ SM_{2,t}^F \\ \vdots \\ \vdots \\ SM_{a,t}^F \\ \vdots \\ \vdots \\ SM_{A,t}^F \end{pmatrix} \quad (4.22)$$

Onde $P_{a,t}^F$ corresponde aos efetivos populacionais femininos com idade a no ano t , $M_{a,t}^F$ corresponde à probabilidade de sobrevivência de uma mulher com idade a atingir o ano $t + 1$, $F_{a,t}$ é a taxa de fecundidade específica de uma mulher com idade a , no ano t , $SM_{a,t}^F$ diz respeito ao saldo migratório feminino com idade a no ano t e r corresponde à proporção de nados-vivos do sexo feminino.

Para a população masculina, o cálculo da população para o horizonte temporal da projeção é análogo, com a exceção de que não existem taxas de fecundidade na matriz de Leslie, sendo os nados-vivos masculinos calculados da seguinte forma:

$$(1 - r) \sum_{a=15}^{49} F_{a,t} P_{a,t}^F \quad (4.23)$$

4.4 Metodologia de Box-Jenkins

Posteriormente à estimação do índice do nível geral de mortalidade (γ_t) e do índice transformado da fertilidade (g_t), o próximo passo é modelar cada um deles como um processo estocástico de séries temporais. Segundo Wei (1990) uma série temporal é uma sequência ordenada de observações, esta ordenação ocorre usualmente no tempo, em particular com intervalos igualmente espaçados. Nesta dissertação iremos trabalhar apenas com séries temporais univariadas discretas, em intervalos de tempo igualmente espaçados, que salvo referência contrária será o ano.

Analogamente a Lee and Carter (1992) pretende-se identificar o modelo correto dentro da classe geral dos Modelo Misto Autoregressivo e Médias Móveis (ARMA) ou Modelo Misto Integrado Autoregressivo e Médias Móveis (ARIMA). Definindo rigorosamente o primeiro destes processos, diz-se que, a série temporal $\{X_t, t \in \mathbb{Z}\}$ é uma série mista autoregressiva de médias móveis de ordens p e q ($p, q \in \mathbb{N}$) e escreve-se $\{X_t, t \in \mathbb{Z}\} \sim \text{ARMA}(p, q)$, se satisfaz a seguinte equação às diferenças estocástica:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (4.24)$$

que é equivalente a

$$\phi_p(B)X_t = \theta_q(B)\epsilon_t \quad (4.25)$$

onde ϵ_t tem valor esperado 0 e variância σ_ϵ^2 e

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad \text{e} \quad \theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (4.26)$$

não têm raízes comuns. O operador B designa-se por *operador backshift* e define-se por $BX_t = X_{t-1}$. No processo considerado o parâmetro p diz respeito ao número de parâmetros autoregressivos do modelo e q corresponde ao número de parâmetros de médias móveis. Para o processo ser estacionário as raízes de $\phi_p(B) = 0$ devem estar fora do círculo unitário e para o processo ser invertível as raízes de $\theta_q(B) = 0$ devem estar fora do círculo unitário. Já os modelos ARIMA são não estacionários e compreendem três parâmetros, representando-se por $ARIMA(p,d,q)$, onde p e q têm o mesmo significado para os modelos previamente mencionados e d corresponde ao número de vezes que a série precisa de ser diferenciada de forma a resultar num modelo estacionário, ou seja, num modelo ARMA.

Pretende-se então encontrar um modelo que siga os princípios da modelação estatística (Turkman, 2010), isto é, com elevada bondade de ajustamento (o modelo deve-se adaptar aos dados), parcimonioso (de todos os modelos que se adequam aos dados, deve-se utilizar o mais simples) e que tenha interpretação (o modelo deve ser útil e de fácil interpretação face ao que se pretende estudar).

Para o efeito iremos aplicar a metodologia *forecast and control*⁴ (também denominada por metodologia ARIMA) desenvolvida por Box and Jenkins (1976). Os modelos pressupõem estacionaridade fraca⁵, observações em tempos igualmente espaçados e pelo menos 30 a 50 observações.

A metodologia de Box e Jenkins assenta em três grandes etapas:

identificação \implies estimação \implies avaliação do diagnóstico

A primeira etapa prende-se com a análise geral das tendências e dos padrões da série temporal. Assim, pretende-se nesta fase inicial, estacionarizar a série temporal através de um conjunto de transformações de modo que a sucessão resultante possa ser descrita por um processo ARMA. Posteriormente seleciona-se um modelo ARMA. O passo seguinte passa pela estimação dos parâmetros do modelo identificado. Esta fase caracteriza-se pela pouca intervenção do analista, visto as estimativas serem obtidas através de *softwares* estatísticos. Por fim, com a avaliação do diagnóstico pretende-se verificar se o modelo identificado inicialmente e se a estimação dos parâmetros são adequados. Esta fase compreende então duas grandes subetapas: analisar a qualidade estatística do modelo e a qualidade do ajustamento. Caso o modelo não satisfaça os critérios destas duas subetapas deve ser rejeitado (Murteira, B., Müller, D. and Turkman, K., 1993) e deve repetir-se a metodologia de Box-Jenkins, não descurando os conhecimentos previamente adquiridos sobre a série temporal. Os três próximos subcapítulos destinam-se à descrição estatística das etapas mencionadas.

⁴Nota de tradução: previsão e controlo.

⁵Uma série temporal $\{X_t, t \in \mathbb{Z}\}$ diz-se estacionária ou fracamente estacionária se i) $E[X_t^2] < \infty, \forall t \in \mathbb{Z}$, ii) $E[X_t] = \mu, \forall t \in \mathbb{Z}$ e iii) $Cov(X_s, X_t) = Cov(X_{s+h}, X_{t+h}), \forall s, t, h \in \mathbb{Z}$.

4.4.1 Identificação

Nesta etapa pretende-se ter uma ideia dos parâmetros p , q e d dos modelos ARIMA, sendo que o principal objetivo passa por reduzir o número de modelos adequados à série temporal. Num primeiro passo, deve-se representar graficamente a série temporal $\{\gamma_t, t = 1, \dots, T\}$ e escolher as transformações necessárias para a transformar numa série estacionária. Posteriormente devem-se avaliar a Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial (FACP), notadas respetivamente por ρ_k e ϕ_{kk} para um *lag* de amplitude k com $k = 1, \dots, K$, de modo a identificar os parâmetros p e q . É de salientar que quando se trabalha com dados demográficos raramente os parâmetros p e q são superiores a dois (Russolillo, 2005). Considerem-se então os seguintes estágios:

1. Estacionarização da série temporal

- (a) A transformação de Box-Cox ⁶ para a estabilização da variância da série temporal.
- (b) Após estabilização da variância da série temporal, o próximo passo é a estabilização da média, para tal devem-se aplicar as seguintes transformações:
 - i. Operador diferenciação simples de ordem d

$$\nabla^d = (1 - B)^d \quad \text{com} \quad BX_t = X_{t-1} \quad (4.27)$$

Note-se que o número de vezes que a série temporal precisa de ser diferenciada para alcançar a estacionaridade reflete-se no parâmetro d do modelo.

- ii. Operador diferenciação sazonal de ordem D

$$\nabla_s^D = (1 - B^s)^D \quad (4.28)$$

Apesar de referido, não iremos empregar o operador de diferenciação sazonal pois as séries temporais com que vamos trabalhar não o necessitam.

- (c) Aplicação do teste Aumentado de Dickey-Fuller (ADF)
Este teste permite verificar se uma série (ou as suas diferenças) é ou não estacionária sendo a hipótese nula a de não estacionaridade.

2. Análise da FAC e da FACP empíricas da série temporal

A representação gráfica da FAC e da FACP serve de apoio à identificação do número de parâmetros p e q do modelo, para tal, deve-se ter em conta as particularidades descritas na tabela 4.1 dos principais processos estacionários.

⁶Esta transformação foi desenvolvida pelos estatísticos George E. P. Box e David Cox em 1964, num artigo designado por *An Analysis of Transformations*.

Tabela 4.1: Principais propriedades da FAC e da FACP empíricas das séries temporais estacionárias

| | FAC ρ_k | FACP ϕ_{kk} |
|-------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| AR(p) | Decaimento exponencial ou sinusoidal amortecido, para zero | Decaimento brusco para zero a partir de um certo <i>lag</i> k , $k > p$ |
| MA(q) | Decaimento brusco para zero a partir de um certo <i>lag</i> k , $k > q$ | Decaimento exponencial ou sinusoidal amortecido, para zero |
| ARMA(p , q) | Decaimento exponencial ou sinusoidal amortecido para zero a partir da ordem $q + 1$ | Decaimento exponencial ou sinusoidal amortecido para zero a partir da ordem $p + 1$ |

3. Segundo Pacheco (2001) informação adicional que pode ser usada na identificação das ordens p e q de uma série temporal ARMA incluem o uso da função de autocorrelação inversa, a função de autocorrelação estendida e diversos critérios de ajustamento.

4.4.2 Estimação

Como foi mencionado anteriormente, esta etapa da metodologia de Box-Jenkins tem uma participação pouco ativa por parte do analista, visto os parâmetros serem estimados através de *software* estatístico. Desta forma não julgamos necessária uma revisão teórica dos métodos e iremos apenas enunciar os mais relevantes:

- Método dos momentos;
- Método da máxima verosimilhança;
- Método da máxima verosimilhança (ou mínimos quadrados) condicional.

4.4.3 Avaliação do diagnóstico

Posteriormente à identificação de um modelo ARMA e à estimação dos seus parâmetros, a avaliação do diagnóstico é aplicada ao modelo ajustado. O objetivo passa por, avaliar a qualidade estatística e a qualidade do ajustamento do modelo à série, de modo a que haja um compromisso entre os três princípios: adaptação, parcimónia e interpretação. Registe-se que este ponto teve como bibliografia principal Murteira, B., Müller, D. and Turkman, K. (1993) e Pacheco (2001).

4.4.3.1 Avaliação da qualidade estatística do modelo

1. Significância estatística do modelo

Tendo em vista um modelo parcimonioso, devem-se analisar as estimativas dos parâmetros do modelo com o objetivo de eliminar aqueles que são desnecessários, ou seja, aqueles que sejam estatisticamente diferentes de zero. Seja β_i um parâmetro estimado do modelo, pretende-se então realizar o seguinte teste de hipóteses:

$$H_0 : \beta_i = 0 \quad \text{versus} \quad H_1 : \beta_i \neq 0$$

com a respetiva estatística de teste:

$$T_i = \sqrt{n} \frac{\hat{\beta}_i}{\sqrt{v_i(\hat{\beta})}} \stackrel{H_0}{\sim} t_{n-p-q} \quad (4.29)$$

onde $\sqrt{v_i(\hat{\beta})}$ é a variância do parâmetro estimado $\hat{\beta}_i$, n representa o número de observações, p representa o número de parâmetros autoregressivos e q diz respeito ao número de parâmetros de médias móveis. A hipótese nula deve ser rejeitada para valores absolutos da estatística de teste elevados, sendo que a não rejeição da hipótese nula deve consuzir à eliminação do parâmetro β_i do modelo.

2. Estacionaridade e invertibilidade do modelo

O modelo deve ser estacionário caso tenha parâmetros autoregressivos e deve ser invertível caso tenha parâmetros de médias móveis.

3. Estabilidade do modelo estimado

Para a análise da estabilidade do modelo estimado deve-se avaliar a estrutura de correlações entre as estimativas dos parâmetros. Uma correlação elevada indica, em geral, uma má qualidade dos parâmetros estimados e também uma sobreparametrização do modelo. Na prática, consideram-se estimativas altamente correlacionadas quando o correspondente coeficiente de correlação for, em módulo, superior a 0,7. Caso esta situação ocorra devem-se procurar modelos alternativos.

4. Redundância dos parâmetros estimados

Considere-se um modelo misto ARMA. Se multiplicarmos, tanto o operador autoregressivo como o operador de médias móveis desse modelo, pelo o mesmo fator, obtemos um segundo modelo misto, diga-se ARMA* que é idêntico ao inicial pois possuem variância, FAC e FACP muito semelhantes. Nestas circunstâncias diz-se que existe redundância dos parâmetros estimados. Citando Murteira, B., Müller, D. and Turkman, K. (1993) "*modelos que contenham fatores redundantes ou quase redundantes devem ser evitados, não só por respeito ao princípio da parcimónia, como pela má qualidade das estimativas obtidas*".

4.4.3.2 Avaliação da qualidade do ajustamento do modelo

A qualidade do ajustamento de um modelo pode avaliar-se através da análise dos correspondentes resíduos. Suponha-se que o seguinte modelo ARMA(p, q)

$$\phi_p(B)X_t = \theta_q(B)\epsilon_t \quad (4.30)$$

foi identificado e que $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)$ e $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_q)$ representam os estimadores dos p e q parâmetros respetivamente, então as estimativas dos resíduos ϵ_t definem-se por

$$\hat{\epsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)X_t, \quad t = 1, \dots, n \quad (4.31)$$

com $\hat{\phi}_p(B) = 1 - \hat{\phi}_1(B) - \dots - \hat{\phi}_p B^p$ e $\hat{\theta}_q(B) = 1 - \hat{\theta}_1(B) - \dots - \hat{\theta}_q B^q$.

De forma a avaliar a qualidade do ajustamento do modelo, interessa verificar os seguintes pontos

1. As variáveis aleatórias $\hat{\epsilon}_t$ possuem distribuição normal

Para o efeito deve analisar-se a série das estimativas dos resíduos padronizados $\{\hat{\epsilon}_t/S_{\hat{\epsilon}}^2\}$ através do seu cronograma, histograma e do papel de probabilidades normal (ou do Q-Q plot).

2. Variância dos resíduos $\hat{\epsilon}_t$ é constante

Para a verificação deste ponto deve analisar-se o cronograma da série das estimativas dos resíduos (ou dos resíduos padronizados) e analisar o efeito da aplicação de transformações estabilizadoras de variância à mesma. Se se chegar à conclusão que a variância da série dos resíduos não é constante devem procurar-se modelos alternativos.

3. A série temporal $\{\hat{\epsilon}_t, t \in \mathbb{Z}\}$ é uma série de ruído branco

Um processo $X = \{X_t, t \in \mathbb{Z}\}$ diz-se que constitui uma série de ruído branco quando é formado por uma sucessão de variáveis aleatórias com a mesma distribuição, valor esperado constante μ (usualmente $\mu = 0$) e variância σ^2 e escreve-se $X \sim RB(\mu, \sigma^2)$, se

$$Cov(X_t, X_{t+k}) = \begin{cases} \sigma^2 & s = 0 \\ 0 & s \neq 0 \end{cases} \quad \forall t \in \mathbb{Z} \quad (4.32)$$

Para confirmação deste último pressuposto deve-se por um lado, analisar a FAC e a FACP estimadas a partir dos resíduos de modo a confirmar se as correspondentes funções teóricas são nulas para todas as ordens positivas, visto as características que identificam a série dos resíduos como uma série de ruído branco serem:

$$\rho_k = 0 \quad \text{e} \quad \phi_{kk} = 0, \quad k \neq 0 \quad (4.33)$$

Definam-se então $\{\hat{\rho}_k(\hat{\epsilon}), k = 1, \dots, K\}$ e $\{\hat{\phi}_{kk}(\hat{\epsilon}), k = 1, \dots, K\}$ como a FAC e a FACP respetivamente, obtidas a partir da série das estimativas dos resíduos do modelo. De seguida especificam-se os testes a utilizar:

• Teste de Bartlett (1946)

Para cada k ($k = 1, \dots, K$) testa a nulidade da FAC dos resíduos, sendo que a rejeição da hipótese nula leva a concluir pelo mau ajustamento do modelo estimado. Contudo deve ter-se em conta que uma vez que se efetuam sucessivos testes de hipóteses, é de esperar que, mesmo que a série dos resíduos seja uma série de ruído branco, haja rejeição da hipótese nula para algumas ordens, especialmente as de ordem mais baixa.

• Teste de Box-Pierce (1970)

Ao contrário do teste de Bartlett, este teste diz respeito à globalidade dos *lags* da FAC dos resíduos. Se o *lag* máximo for K então a hipótese nula é dada por $H_0 : \rho_1(\hat{\epsilon}) = \rho_2(\hat{\epsilon}) = \dots = \rho_K(\hat{\epsilon}) = 0$.

• Teste de Ljung-Box (1978)

Este teste é análogo ao teste supramencionado, com a melhoria de que a estatística de teste utilizada converge mais rapidamente, sendo por isso preferível quando a série a ser estudada não é muito grande.

• Teste de Jenkins e Daniels (1956)

Análogo ao teste de Bartlett mas com a substituição da FAC pela FACP das estimativas dos resíduos, o presente teste dirige-se então à significância estatística da FACP tendo como hipótese nula $H_0 : \phi_{kk}(\hat{\epsilon}) = 0$

4.4.3.3 Critérios de seleção de modelos

Já se pretendermos selecionar um único modelo entre vários, cujo *feedback* da avaliação do diagnóstico é positivo, existem vários critérios baseados nas estimativas dos resíduos que possibilitam a escolha do modelo que

melhor se adequa à série temporal, destaca-se o Critério de Informação Akaike (AIC) e o Critério Bayesiano de Informação Akaike (BIC). O AIC define-se como

$$AIC = -2 \ln(L) + 2m \quad (4.34)$$

onde se define L como a máxima verosimilhança do modelo ajustado e por m o número de parâmetros ajustados. Como é sabido um modelo é considerado melhor se tem uma verosimilhança elevada, contudo um modelo com menos parâmetros apresenta geralmente verosimilhanças maiores, uma das vantagens do AIC é que este retira o efeito teoricamente positivo dos modelos com menos parâmetros através da integração do termo $2m$. Por outro lado, uma das desvantagens deste critério consiste no facto de não integrar o número de observações (n) da série a que é ajustado o modelo, o que, segundo Bozdogan (1987) conduz a uma falta de consistência. O BIC surge como uma extensão Bayesiana do critério anterior. Defina-se S_X^2 como a variância amostral da série em estudo e S_ε^2 como a variância amostral da série das estimativas dos resíduos, o critério é dado por

$$BIC = n \ln(S_\varepsilon^2) + m \ln(n) - (n - m) \ln\left(1 - \frac{m}{n}\right) + m \ln\left(\frac{1}{m} \left[\frac{S_X^2}{S_\varepsilon^2} - 1\right]\right) \quad (4.35)$$

Registe-se que apenas se podem comparar modelos através destas grandezas se estes estiverem associados a mecanismos análogos de estacionarização da série original, caso contrário não são comparáveis. Procura-se sempre o modelo com menor AIC e BIC.

Aplicação do modelo desenvolvido de projeções estocásticas à realidade portuguesa

5.1 Modelação da componente da mortalidade

Para a análise da componente da mortalidade apuraram-se as taxas de mortalidade centradas desagregadas por sexo e idade de 1970 a 2009, definidas à custa do número de óbitos e da população exposta ao risco:

$$\text{Taxa de mortalidade centrada à idade } a = \frac{\text{Número de óbitos com idade } a}{\text{População exposta ao risco com idade } a} \quad (5.1)$$

Para o número de óbitos empregaram-se, de 1970 a 1995, dados extraídos da *Human Mortality Database* (HMD) de 1996 a 2009 dados extraídos do portal do INE, ambos desagregados por sexo e idade. Para a população exposta ao risco utilizou-se, para todo o período analisado, informação da HMD. Para cada ano analisado, os óbitos de idades desconhecidas foram alocados à idade com maior número de óbitos nesse ano.

Considerou-se a desagregação etária dos 0 aos 100 e mais anos de idade. Inicialmente a opção foi considerar apenas até aos 85 ou mais anos, todavia, uma vez que o número de indivíduos pertencente a este grupo etário está a aumentar cada vez mais, devido a melhorias expressivas na esperança de vida, optou-se por uma desagregação mais fina.

Como se pode observar, pela figura 5.1, a percentagem de população com mais de 85 anos de idade tem vindo a aumentar gradualmente, sendo que no sexo feminino esta proporção subiu 1,7 pontos percentuais entre 1970 para 2009.

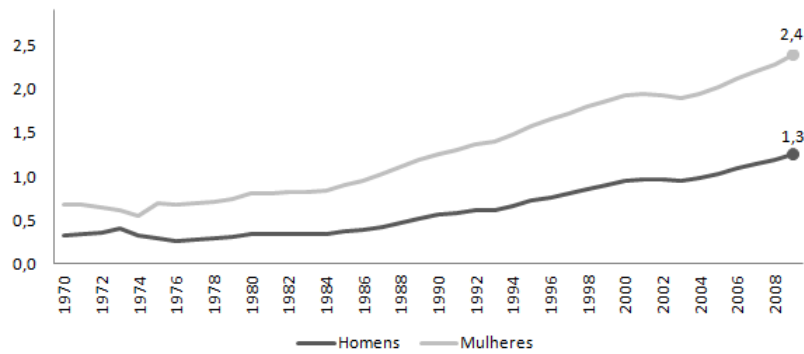


Figura 5.1: Proporção de população com mais de 85 anos de idade na população total por sexo, 1970-2009

5.1.1 O modelo de Lee Carter

O modelo demográfico de Lee Carter decompõe as taxas de mortalidade em três parâmetros, dois respeitantes às idades e fixos ao longo do tempo, α_a e β_a , e um índice representativo do nível geral da mortalidade, γ_t que varia ao longo do tempo.

O parâmetro α_a descreve o perfil geral das taxas de mortalidade específicas por idade. Com exceção das idades mais avançadas, a probabilidade de morte mais elevada encontra-se no grupo etário mais jovem (menos de um ano de idade) dando relevância à mortalidade infantil no período analisado, para ambos os sexos (figura 5.2). O sexo masculino apresenta, em média, uma mortalidade mais elevada que o sexo feminino, observando-se a maior discrepância nas idades compreendidas entre os 17 e os 28 anos de idade, provavelmente efeito das mortes provocadas por acidentes de viação que mais afetam os homens (Torres, 2009). A partir dos 28 esta discrepância tende a diminuir, tornando-se praticamente nula nas idades mais velhas.

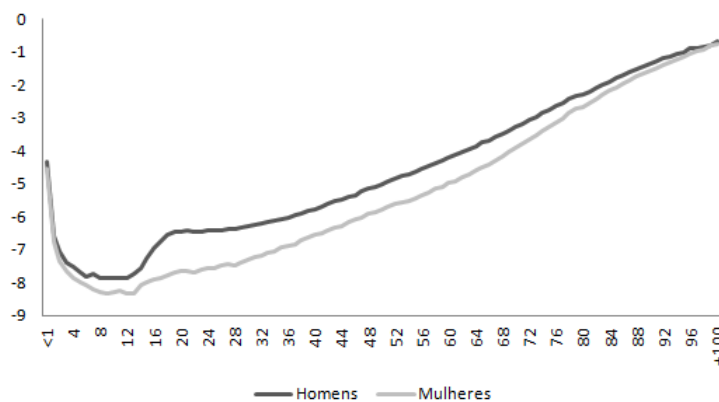


Figura 5.2: Estimativas do parâmetro α_a por sexo e idade, 1970-2009

O parâmetro β_a descreve a tendência da taxa de mortalidade na idade a sofrer maiores oscilações em resposta a oscilações do índice de nível geral γ_t . Pela análise da figura 5.3, observa-se que, para ambos os sexos, β_a tem uma tendência geralmente decrescente assumindo valores mais elevados nas idades mais jovens, o que permite concluir que, quanto mais jovem o grupo etário, maior a sua sensibilidade relativamente a variações de γ_t .

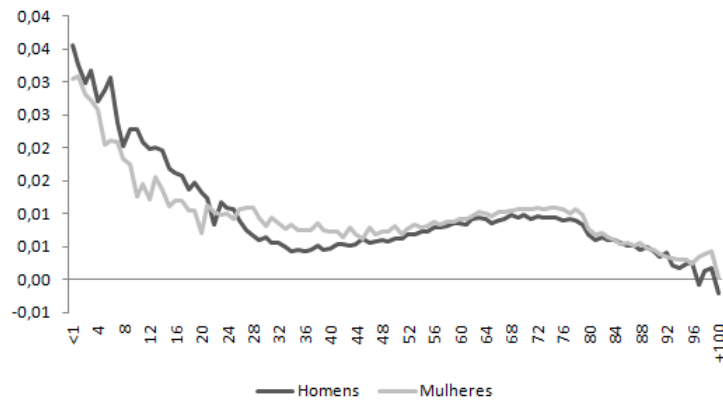


Figura 5.3: Estimativas do parâmetro β_a por sexo e idade, 1970-2009

Uma vez que, a estimação do índice de nível geral da mortalidade γ_t se baseia nos logaritmos das taxas de mortalidade ao invés de se basear nas próprias taxas de mortalidade, recorre-se a uma reestimação deste índice de forma a garantir que o número de óbitos observados anualmente iguale o número de óbitos preditos com o modelo. Na reestimação emprega-se o α_a e o β_a inicialmente estimados.

Em termos práticos permitiu-se que a diferença entre o número de óbitos observados e preditos fosse de um (em valor absoluto), de forma a minimizar o tempo de processamento da função construída para o efeito.

As figuras 5.4 e 5.5 expõem o índice γ_t antes e posteriormente à reestimação, para os homens e as mulheres, respetivamente. Em ambos os sexos, γ_t apresenta um comportamento decrescente com pequenas oscilações, indicativo de que a mortalidade tem vindo a diminuir em Portugal, para o período de 1970 a 2009.

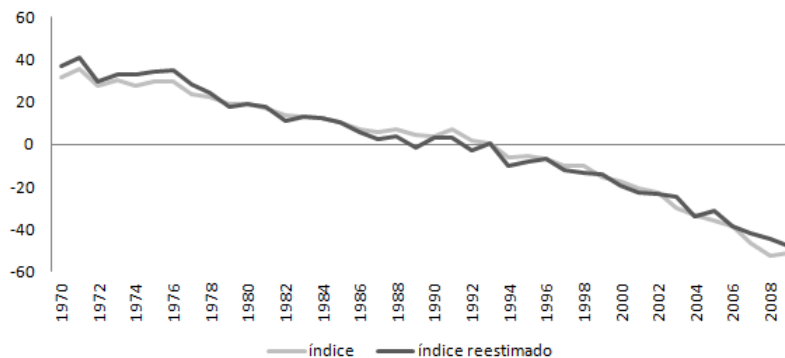


Figura 5.4: Estimativas de γ_t e de γ_t reestimado, Homens, 1970-2009

O coeficiente de determinação (R^2), calculado posteriormente à reestimação de γ_t , indica que o modelo se ajusta bem aos dados, sendo de 99,5% nos homens e de 99,6% nas mulheres.

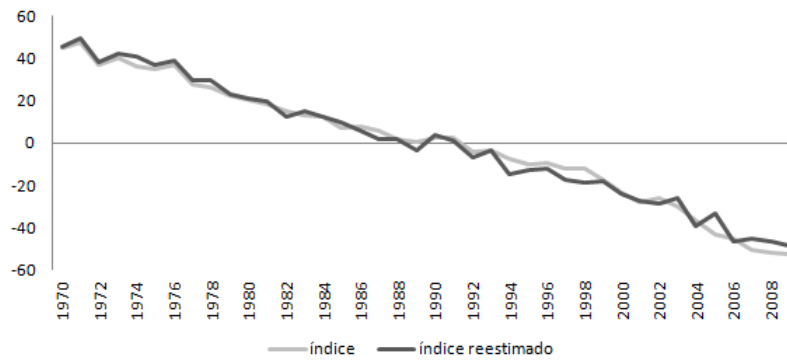


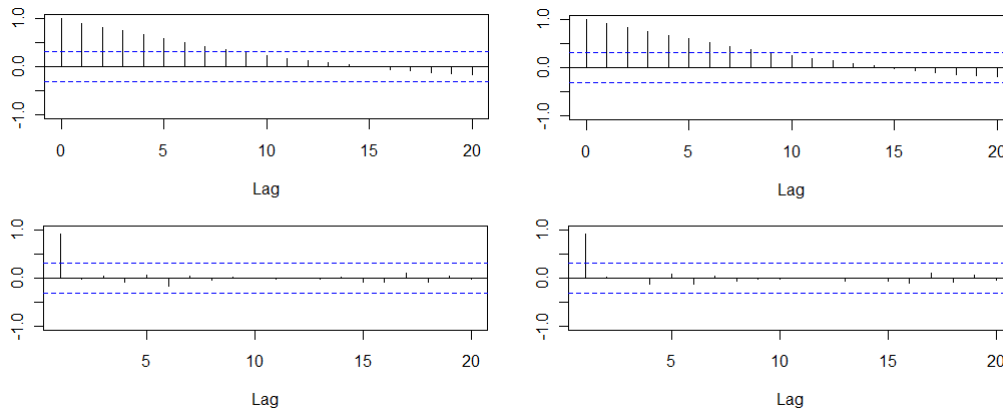
Figura 5.5: Estimativas de γ_t e de γ_t reestimado, Mulheres, 1970-2009

5.1.2 Modelação do índice geral da mortalidade como uma série temporal

5.1.2.1 Identificação e estimação

De forma a identificar o modelo ARIMA, o primeiro passo é verificar se a série temporal γ_t é estacionária, isto é, se tem média, variância e autocorrelação constante ao longo do tempo. O número de vezes que a série é diferenciada reflete-se no parâmetro d do modelo.

A série temporal γ_t tem um comportamento claramente decrescente indicando não estacionaridade em média (conforme figura 5.4 e 5.5). A visualização da FAC e da FACP corroboram este facto, pois a FAC aproxima-se gradualmente de zero e a FACP sofre um corte abrupto no primeiro *lag*, sinal de que se está na presença de uma série não estacionária.



5.6.1: Homens

5.6.2: Mulheres

Figura 5.6: FAC e FACP do índice γ_t

Sendo o índice γ_t não estacionário, construíram-se as suas primeiras diferenças, ilustradas na figura 5.7, para ambos os sexos. As primeiras diferenças da série parecem ter eliminado a não estacionaridade em média, fixando-se esta em torno de -2 .

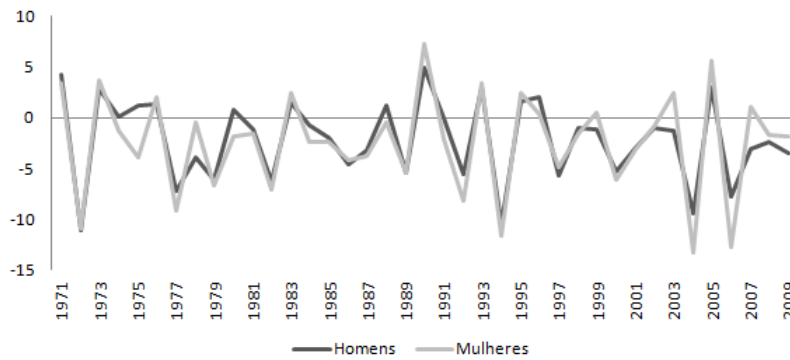
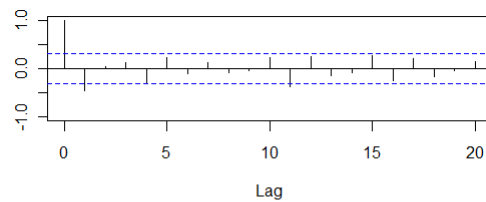
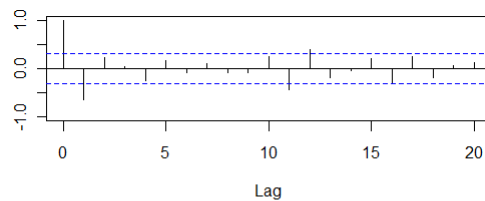


Figura 5.7: Primeiras diferenças de γ_t , 1971-2009

Também a FAC e a FACP das primeiras diferenças, para cada um dos sexos, parecem consistentes com a hipótese de uma série estacionária (figuras 5.8.1 e 5.8.2) pois, ao contrário da FAC FACP do índice geral de mortalidade γ_t sem aplicação das diferenças, não parecem evidenciar um padrão.



5.8.1: Homens



5.8.2: Mulheres

Figura 5.8: FAC e FACP das primeiras diferenças do índice γ_t

A tabela 5.1 apresenta os resultados da aplicação do teste estatístico ADF, em termos de *valor p*. Este teste, corrobora as conclusões a que se chegou através da observação dos gráficos das diferenças do índice geral da mortalidade, assim como através da representação gráfica da FAC e da FACP. Por um lado, não rejeita a hipótese nula de não estacionariedade do índice γ_t , em ambos os sexos e, por outro, rejeita a hipótese de não estacionariedade das primeiras diferenças de γ_t (*valor p* de 0,01 para os homens e de 0,02 para as mulheres).

Tabela 5.1: Resultados (valor p) do teste estatístico ADF

| | Homens | Mulheres |
|------------------------------------|--------|----------|
| γ_t | 0,68 | 0,28 |
| Primeiras diferenças de γ_t | 0,01 | 0,02 |

Findo o processo de estacionarização das séries, conclui-se que ambas as série devem ser diferenciadas uma vez, sendo o parâmetro d do modelo ARIMA igual a 1. Por conseguinte, interessa averiguar os parâmetros p e q do modelo a utilizar tanto para os homens como para as mulheres. Na prática, no que toca a séries temporais demográficas raramente o número destes parâmetros é superior a dois (Russolillo, 2005).

Assim, estimaram-se 9 modelos para cada um dos índices, com p e q a variar entre 0 a 2 e o parâmetro d igual a 1. Incluiu-se também um parâmetro de (*drift*) devido à tendência decrescente dos índices. Utilizou-se o método dos mínimos quadrados para encontrar os valores iniciais, seguido do método de máxima verosimilhança.

5.1.2.2 Avaliação de diagnóstico

Após a estimação dos 18 modelos, designadamente 9 para o sexo masculino e 9 para o sexo feminino, procura-se seleccionar aquele que melhor se ajusta aos dados. Para a seleção do melhor modelo deve-se avaliar, por uma lado, a sua qualidade estatística e, por outro, a sua qualidade do ajustamento. Assim e de acordo com a secção 4.4.3 definiu-se a metodologia a seguir:

1. Excluir os modelos com parâmetros, cujas correlações entre si, são superiores a 0,7;
2. Excluir os modelos com pelo menos um parâmetro estatisticamente não significativo;
3. Excluir os modelos em que se rejeita a hipótese nula do teste de Ljung-Box;
4. Seleccionar, dos modelos restantes, o que tiver menor AIC e/ou BIC;
5. Avaliar a estacionaridade do modelo assim como os seus resíduos e caso não se reúnam as condições necessárias voltar ao ponto anterior.

A tabela 5.2 e a tabela 5.3 apresentam a designação dos modelos construídos, assim como, resultados que permitem avaliar sobre a qualidade estatística e a qualidade de ajustamento de cada um dos modelos à série, para o sexo masculino e feminino, respetivamente ¹.

Os resultados para os homens permitem a exclusão imediata dos modelos m_H^3 , m_H^5 , m_H^6 , m_H^7 , m_H^8 e m_H^9 por terem parâmetros com correlações entre si superiores a 0,7, indicativo de uma má qualidade da estimativa dos parâmetros e de uma sobreparametrização do modelo, ou por terem parâmetros estatisticamente não significativos, ou seja, parâmetros cuja estatística de teste não rejeita a hipótese de serem nulos.

¹Utiliza-se a sigla NA para representar resultados não aplicáveis.

Posto isto, permanecem os modelos m_H^1 , m_H^2 e m_H^4 , dos quais se seleciona o modelo m_H^2 por ser o que tem menor AIC. Desta forma o índice geral de mortalidade masculino apenas tem parâmetros de médias móveis sendo modelado por um ARIMA(0,1,1) com a inclusão de *drift*.

Tabela 5.2: Designação dos modelos e respetivos resultados, Homens

| Designação do modelo | p (AR) | q (MA) | AIC | BIC | Correlações entre parâmetros > 0,7 | Parâmetros não significativos |
|----------------------|--------|--------|--------|--------|------------------------------------|-------------------------------|
| m_H^1 | 0 | 0 | 222,59 | 225,97 | NA | NA |
| m_H^2 | 0 | 1 | 214,26 | 219,33 | não | não |
| m_H^3 | 0 | 2 | 215,96 | 222,72 | não | sim |
| m_H^4 | 1 | 0 | 214,83 | 219,90 | não | não |
| m_H^5 | 1 | 1 | 215,87 | 222,62 | sim | sim |
| m_H^6 | 1 | 2 | 217,91 | 226,36 | sim | não |
| m_H^7 | 2 | 0 | 215,64 | 222,40 | não | sim |
| m_H^8 | 2 | 1 | 215,24 | 223,69 | sim | não |
| m_H^9 | 2 | 2 | 216,83 | 226,96 | sim | sim |

Analogamente para o caso feminino, são rejeitados os modelos m_M^3 , m_M^5 , m_M^6 , m_M^8 e m_M^9 pelos motivos acima enunciados. Dos restantes modelos seleciona-se o m_M^2 por ser o que tem menor AIC. Contudo, numa análise prévia, averigua-se que este modelo não cumpre as condições de invertabilidade, pois a estimativa do seu parâmetro é exatamente -1 . Excluído o modelo m_M^2 , seleciona-se o modelo m_M^7 por ser o próximo com menor AIC. Este modelo é constituído por dois parâmetros autoregressivos, sendo definido por um ARIMA(2,1,0) com *drift*.

Tabela 5.3: Designação dos modelos e respetivos resultados, Mulheres

| Designação do modelo | p (AR) | q (MA) | AIC | BIC | Correlações entre parâmetros > 0,7 | Parâmetros não significativos |
|----------------------|--------|--------|--------|--------|------------------------------------|-------------------------------|
| m_M^1 | 0 | 0 | 238,66 | 242,04 | NA | NA |
| m_M^2 | 0 | 1 | 217,13 | 222,19 | não | não |
| m_M^3 | 0 | 2 | 218,29 | 225,04 | não | sim |
| m_M^4 | 1 | 0 | 219,10 | 224,16 | não | não |
| m_M^5 | 1 | 1 | 217,36 | 224,12 | sim | não |
| m_M^6 | 1 | 2 | 219,23 | 227,68 | sim | sim |
| m_M^7 | 2 | 0 | 217,43 | 224,18 | não | não |
| m_M^8 | 2 | 1 | 218,80 | 227,24 | sim | sim |
| m_M^9 | 2 | 2 | 220,93 | 231,06 | sim | sim |

Selecionados os modelos para cada uma dos sexos, interessa agora avaliar os seus resíduos, estes devem seguir o padrão de uma série de ruído branco. Desta forma, aplicou-se o teste estatístico de Ljung-Box aos resíduos dos modelos selecionado (5.4) para aferir sobre a adequabilidade dos mesmos. Observando os seus resultados conclui-se que o modelo é adequado à série temporal analisada, visto para nenhum dos *lags* analisados se rejeitar a hipótese nula de adequabilidade.

Tabela 5.4: Resultados da aplicação do teste de Ljung-Box (*valor p*)

| | Ljung-Box | | | |
|---------|---------------|---------------|---------------|----------------|
| | <i>lag</i> =4 | <i>lag</i> =6 | <i>lag</i> =8 | <i>lag</i> =10 |
| m_H^2 | 0,42 | 0,56 | 0,69 | 0,81 |
| m_M^7 | 0,07 | 0,15 | 0,24 | 0,40 |

Interessa também analisar a série dos resíduos estandardizados, $\{\hat{\epsilon}_t/S_\epsilon^2\}$, cuja distribuição deve ser normal. Para ambos os modelos, a análise dos Q-Q plot dos resíduos indicia que estes têm uma distribuição normal visto toda a série se distribuir dentro dos intervalos de confiança (figuras 5.9.1 e 5.9.2), resultado que foi confirmado pelo teste estatístico para a normalidade Shapiro-Wilks.

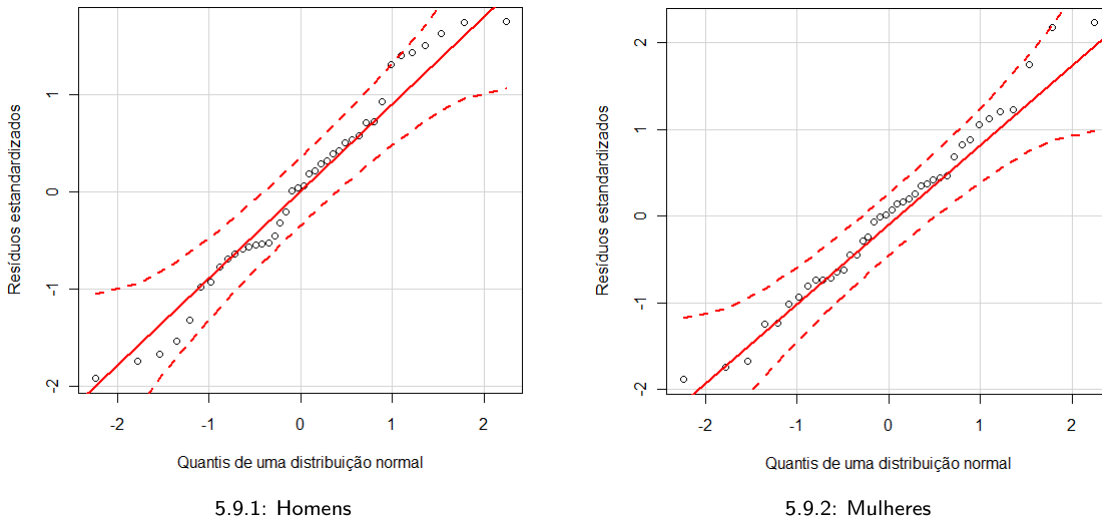


Figura 5.9: Q-Q plot dos resíduos estandardizados

Os resultados anteriores e a observação da série dos resíduos estandardizados (figura 5.10), que tem o perfil de uma série de ruído branco, para cada um dos modelos permite concluir que os modelos seleccionados se ajustam adequadamente as séries históricas analisadas, designadamente o modelo ARIMA(0,1,1) para os homens e um modelo ARIMA(2,1,0) para as mulheres, ambos com a inclusão de uma constante (*drift*).

Desta forma o índice de nível geral de mortalidade, para o sexo masculino, é modelado através de um ARIMA(0,1,1) com a inclusão de *drift*, ou seja:

$$\gamma_t = \lambda + \gamma_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1} \tag{5.2}$$

onde o termo constante λ representa a variação média anual de γ_t , θ_1 representa o parâmetro de médias móveis e ϵ a série dos resíduos do modelo.

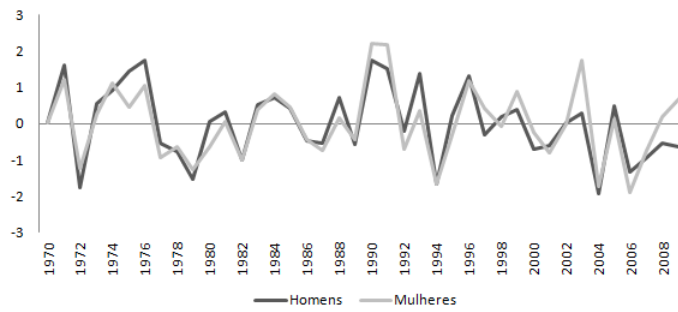


Figura 5.10: Resíduos dos modelos

Por sua vez, o índice de nível geral de mortalidade do sexo feminino é modelado através de um ARIMA(2,1,0) com a inclusão de uma constante (*drift*), isto é:

$$\gamma_t = \lambda + \phi_1\gamma_{t-1} + \phi_2\gamma_{t-2} + \gamma_{t-1} - \phi_1\gamma_{t-2} - \phi_2\gamma_{t-3} + \epsilon_t \quad (5.3)$$

onde ϕ_1 , ϕ_2 correspondem aos parâmetros autoregressivos, λ representa a variação média anual de γ_t e ϵ a série dos resíduos. Os parâmetros estimados e os respetivos erros padrões apresentam-se nas tabelas 5.5 e 5.6 para o sexo masculino e feminino, respetivamente .

Tabela 5.5: Estimativa dos parâmetros para o modelo ARIMA(0,1,1) com *drift*, Homens

| Parâmetro | Coefficiente | Erro padrão | Valor-p |
|------------|--------------|-------------|---------|
| θ_1 | -0,52 | 0,14 | 0 |
| λ | -2,19 | 0,28 | 0 |

Uma vez que o modelo ARIMA(0,1,1) é apenas constituído por parâmetros de médias móveis, tem-se somente de verificar as condições de invertabilidade, de facto $|\theta_1| < 1$. Por outro lado o modelo selecionado para o índice de mortalidade feminino, ARIMA(2,1,0), apenas tem parâmetros autoregressivos e deve verificar unicamente as condições de estacionaridade, ou seja, que $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ e $\phi_2 < 1$.

Tabela 5.6: Estimativa dos parâmetros para o modelo ARIMA(2,1,0) com *drift*, Mulheres

| Parâmetro | Coefficiente | Erro padrão | Valor-p |
|-----------|--------------|-------------|---------|
| ϕ_1 | -0,85 | 0,15 | 0 |
| ϕ_2 | -0,30 | 0,15 | 0,01 |
| λ | -2,49 | 0,27 | 0 |

5.1.3 Projeção do índice γ_t

Posteriormente à estimação dos modelos ARIMA para os índices da mortalidade de cada um dos sexos, segue-se agora a sua projeção. Utilizam-se os modelos ARIMA(0,1,1) e ARIMA(2,1,0) para gerar projeções para os próximos 41 anos (2010 a 2050), baseados nas observações do período de 1970 a 2009. As figuras 5.11 e 5.12 exibem as observações do índice de mortalidade juntamente com as suas projeções e intervalos de confiança a 80% e a 95%, para o sexo masculino e feminino, respetivamente. Para ambos os sexos, as projeções de γ_t seguem a tendência decrescente da série histórica e representativa de uma, cada vez mais, baixa mortalidade. Repare-se igualmente que os intervalos de confiança associados têm pequenas amplitudes, representando uma boa estimação da projeção.

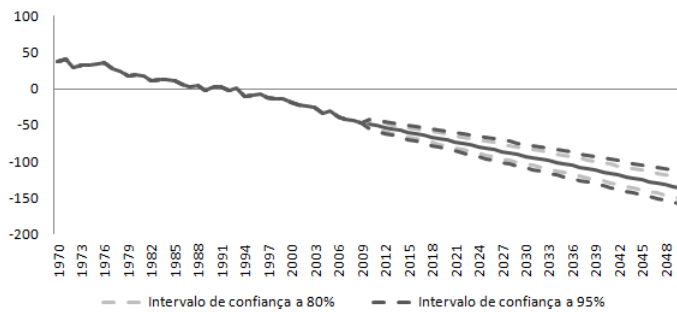


Figura 5.11: Projeção do índice de mortalidade γ_t , Homens, 2010-2050

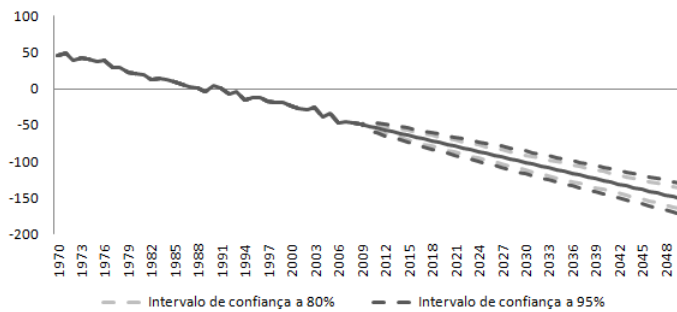


Figura 5.12: Projeção do índice de mortalidade γ_t , Mulheres, 2010-2050

Note-se que a aplicação do método de Lee Carter com vista a obter a projeção das taxas de mortalidade por idades, para o período de 2010 a 2050, fixa os parâmetros α_a e β_a , variando unicamente o índice γ_t . Para a generalidade das idades as taxas de mortalidade projetadas são bastantes semelhantes entre os dois sexos, com exceção das idades mais avançadas onde a taxa de mortalidade dos homens assume valores superiores (figuras 5.13.1 e 5.13.2), reflexo da diferença de esperança média de vida entre os dois sexos.

Posteriormente ao cálculo das taxas de mortalidade prepararam-se tábuas de mortalidade de acordo com a metodologia apresentada na subsecção 4.1.2. Apurou-se, para cada ano e para cada uma das idades consideradas, a probabilidade de sobrevivência (número de sobreviventes), a probabilidade de morte (número de óbitos), os anos completos após cada uma das idades e, por fim, a esperança média de vida à nascença assim como a esperança média de vida aos 65 anos de idade.

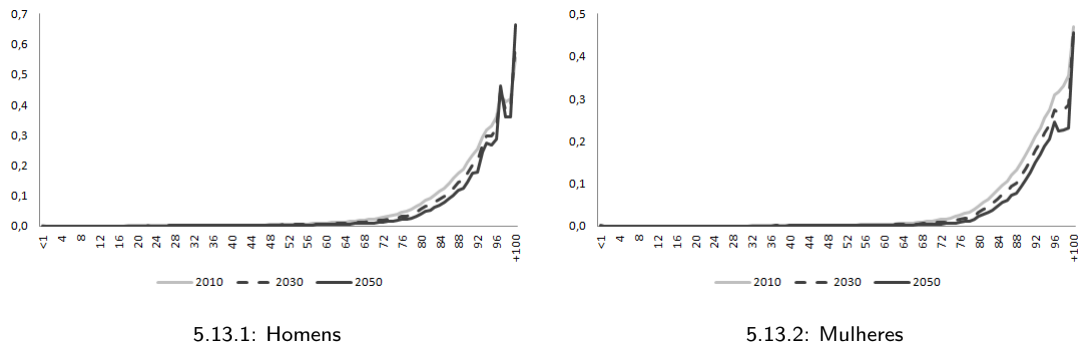


Figura 5.13: Projeção das taxas de mortalidade específicas por idade para 2010, 2030 e 2050

A esperança média de vida à nascença que em 2009 era de 76,1 anos para os homens e de 82,1 anos para as mulheres aumentou em 2050 para 83 anos nos homens e 88,5 anos para as mulheres, representando ganhos de 6,9 anos para os homens e de 6,4 anos para as mulheres. Nos homens, este ganho foi reflexo de um crescimento constante ao longo de toda a projeção, tendo uma média de ganhos de cerca de 0,2 anos por ano até 2041 e de 0,1 nos restantes anos até 2050. Para as mulheres, os ganhos na esperança média de vida não se mostraram tão otimistas, sendo de 0,2 nos primeiros 19 anos da projeção (até 2028) e de 0,1 nos anos remanescentes. A superioridade dos homens nos ganhos da esperança de vida à nascença indicia uma futura aproximação entre os dois sexos. As figuras 5.14.1 e 5.14.2 apresentam a projeção da esperança média de vida para o horizonte temporal de 2050, assim como os respetivos intervalos de confiança.

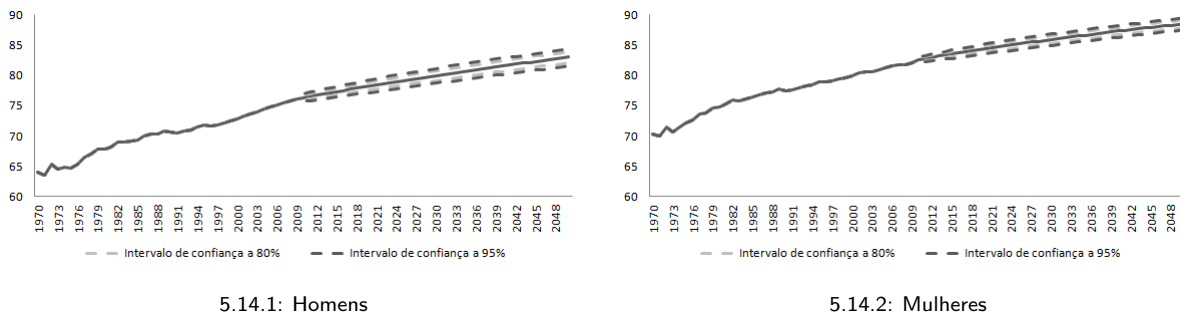


Figura 5.14: Projeção da esperança média de vida à nascença, 2010-2050

Os resultados do INE para a projeção da esperança média de vida à nascença são de 79 anos para os homens e de 84,7 anos para as mulheres, já no caso da projeção realizada pelo Eurostat são de 80,4 anos para os homens e de 86,6 para as mulheres. Ambos os resultados, apesar de próximos, são ligeiramente inferiores aos resultados que se obteram com o método de Lee Carter o que pode estar diretamente relacionado com a série histórica utilizada assim como com o momento de partida. Nesta dissertação projetaram-se as taxas a partir de 2010, por outro lado, as projeções oficiais têm como população de base o ano de 2000 e o exercício do Eurostat parte de 2004.

5.2 Modelação da componente da fertilidade

Emprega-se o modelo demográfico, como foi descrito na secção 4.2.1, às taxas de fecundidade específicas por idade fértil da mulher para o período de 1980 a 2009. Para o efeito, utilizou-se informação oficial fornecida pelo INE, o indicador nados-vivos por sexo e idade da mãe e o indicador população média anual residente por sexo e idade. Os nados-vivos de mães com idade ignorada foram alocados, em cada ano, à idade da mãe com mais nados-vivos. Esta imputação é na generalidade insignificante, sendo exceção o ano de 2008 com 940 nados-vivos de mães com idades ignoradas.

A taxa de fecundidade específica por idade da mãe define-se pelo quociente entre o número de nados-vivos de mães com idade a pelo efetivo médio de mulheres com essa idade:

$$\text{Taxa de fecundidade específica à idade } a = \frac{\text{Número de nados-vivos de mães com idade } a}{\text{Número médio de mulheres à idade } a} \quad (5.4)$$

As idades consideradas foram desde as mães com menos ou 14 anos de idade até às mães com 50 ou mais anos de idade. Note-se que oficialmente a idade fértil da mulher é dos 15 aos 49 anos de idade, contudo uma vez que existem nados-vivos de mães com idades fora deste intervalo, decidiu-se contemplar mais estas duas classes etárias.

5.2.1 O modelo de Lee Carter modificado

O modelo de Lee Carter modificado aplicou-se à matriz de taxas de fecundidade específicas por ano e idade da mãe, para o período de 1980 a 2009. O parâmetro θ_a traduz a forma geral do perfil da fertilidade por idade da mãe para os anos analisados (figura 5.15).

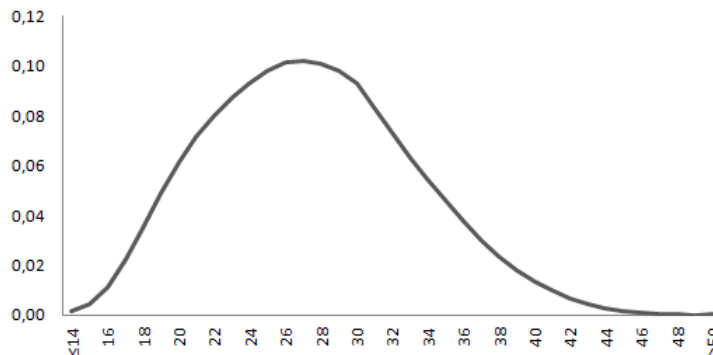


Figura 5.15: Estimativas do parâmetro θ_a por idade fértil da mulher, 1980-2009

Como se pode observar as estimativas de θ_a parecem seguir um comportamento normal com uma cauda ligeiramente mais alongada à direita, correspondendo às idades mais avançadas. Para o período analisado a fertilidade média é mais elevada nas idades compreendidas entre os 22 e os 31 anos de idade, tornando-se praticamente nula a partir dos 45 anos.

Já o coeficiente δ_a apresenta um comportamento bastante oscilatório (figura 5.16), com o seu máximo a ser atingido aos 23 anos de idade e o seu mínimo aos 33 anos de idade. Note-se que este parâmetro, apesar de não ter significado específico descreve a sensibilidade das taxas de fertilidade relativamente a variações no parâmetro κ_t .

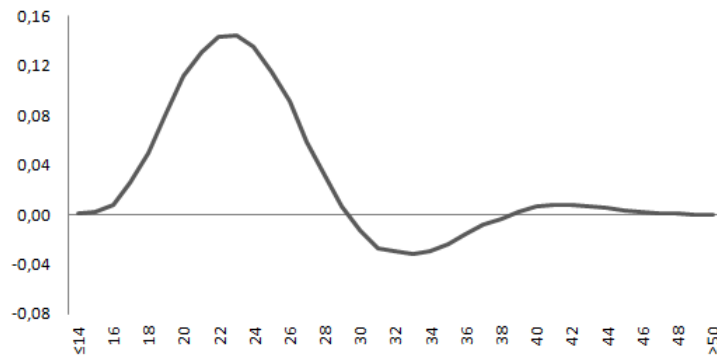


Figura 5.16: Estimativas do parâmetro δ_a por idade fértil da mulher, 1980-2009

É possível concluir que, em termos gerais, as idades compreendidas entre os 18 e os 27 anos de idade tem uma maior sensibilidade relativamente a variações de κ_t e que por outro lado, a sensibilidade das taxas de fertilidade a partir dos 40 anos relativamente a κ_t tende a aproximar-se de zero.

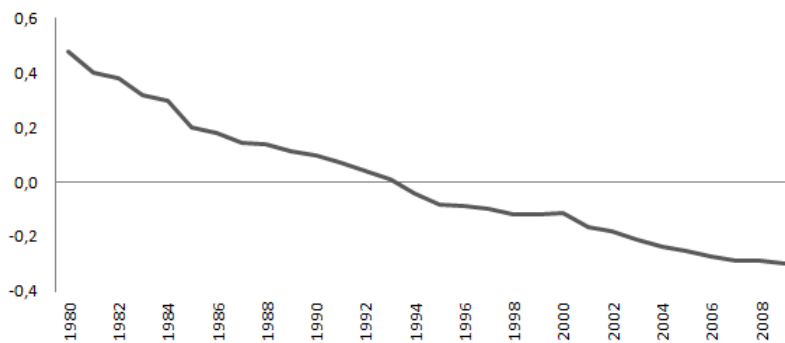


Figura 5.17: Estimativas do parâmetro κ_t , 1980-2009

As estimativas do índice geral de fertilidade, κ_t (figura 5.17) apresentam uma tendência claramente decrescente com pequenas oscilações, traduzindo uma diminuição gradual das taxas de fertilidade ao longo dos últimos anos.

Posteriormente à estimação dos parâmetros θ_a , δ_a e κ_t , segue-se o apuramento da estimativa do ISF de forma a construir-se o índice de fertilidade transformado g_t :

$$g_t = \ln \frac{F_t - I}{S - F_t} \quad (5.5)$$

Tanto o limite superior (S) como o limite inferior (I) tem de ser pré-especificados. Sendo o ISF máximo para o período analisado de 2,25 crianças por mulher em 1980 e tendo este uma tendência geralmente decrescente, é de esperar, que mesmo existindo uma recuperação do ISF, este não volte a atingir novamente aquele máximo. Assim estabeleceu-se o limite superior em 2,3 crianças por mulher. Visto a transformação g utilizar a função logaritmo, o limite superior tem de ser sempre maior que o máximo do ISF observado, caso assim não fosse, obteriam-se valores indeterminados.

Por outro lado, a determinação do limite inferior revestiu-se de alguma dificuldade. Num primeiro ensaio definiu-se o limite inferior como 1, ou seja, estabeleceu-se que, para a projeção de 2009 a 2050 o ISF nunca poderia ser menor que uma criança por mulher. Esta escolha baseou-se na opinião generalizada de diversos especialistas, de que o ISF poderá experienciar uma ligeira recuperação devido ao adiamento da natalidade para idades mais avançadas da mulher, não se esperando que atinja valores inferiores a 1. Outro fator que contribuiu para esta decisão foram os resultados das projeções oficiais para 2050, que relativamente à componente da fertilidade assumem, para todos os ensaios, que o ISF nunca é inferior a um, aliás para o cenário inferior este é de 1,26.

Contudo, a utilização deste limite inferior, produziu taxas de fecundidade específicas negativas para os intervalos de confiança inferiores a 80% e a 95%, nas idades compreendidas entre os 20 e os 23 anos de idade a partir de 2021 e também nas idades superiores e iguais a 45 anos a partir de 2012. Para o primeiro grupo etário este facto deve-se ao rápido decréscimo das taxas de fertilidade nestas idades no período de base da projeção, como é possível constatar na figura 5.18. Já nas idades mais elevadas deve-se às taxas de fertilidade terem também um comportamento decrescente, assim como, compreenderem constantemente valores extremamente baixos (entre 0 e 0,0015 a partir de 1990).

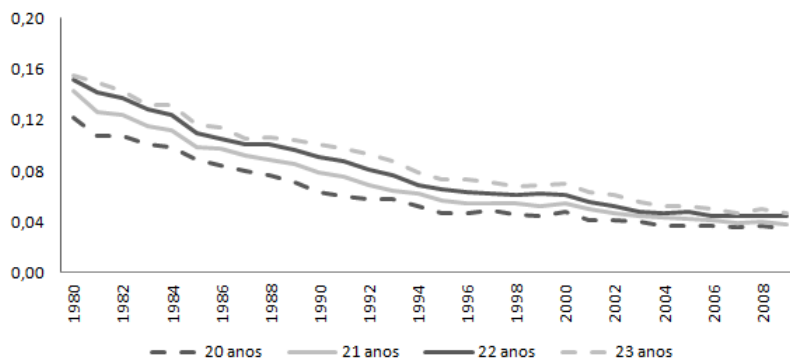


Figura 5.18: Taxas de fertilidade específicas dos 20 aos 23 anos, 1980-2009

Pelas razões enunciadas elevou-se o limite inferior do ISF para 1.1 crianças por mulher, o que resolveu o problema das taxas de fecundidade negativas nas idades entre os 20 e os 23 anos, contudo nas idades mais avançadas estas manteram-se negativas. Assim, por cada ano da projeção, decidiu-se apurar a média das taxas de fecundidade específicas do grupo etário constituído pelas idades entre os 40 e os 50 anos, e aplicar essa média em todo o grupo etário considerado.

Apesar de se reconhecer que não é o método mais viável, apenas se está a evitar que existam nascimentos negativos, pois em termos totais a diferença do número de nados-vivos não se revela significativa. A título de exemplo, aplicaram-se as taxas de fecundidade específicas de 2014 originalmente projetadas à população feminina em idade fecunda de 2013 e também as taxas de fecundidade específicas com a aplicação da média à mesma população. A diferença foi de menos 47 nados-vivos. Todas estas questões enunciam um problema de ajustamento e adequação do método ao caso português que irá ser explorado mais tarde.

Estabelecidos os limites inferior e superior é possível averiguar o índice transformado da fertilidade g_t (figura 5.19), como se pode observar, tal como o parâmetro κ_t e o ISF, este tem uma tendência decrescente. Note-se que, a utilização das estimativas do ISF ao invés do próprio ISF, para a estimação do índice transformado de fertilidade g_t , tem como objetivo minorar variações irregulares provenientes de circunstâncias históricas, assim como, o alcance de uma função mais linear, que seja facilmente modelada e projetada através da metodologia de Box-Jenkins.

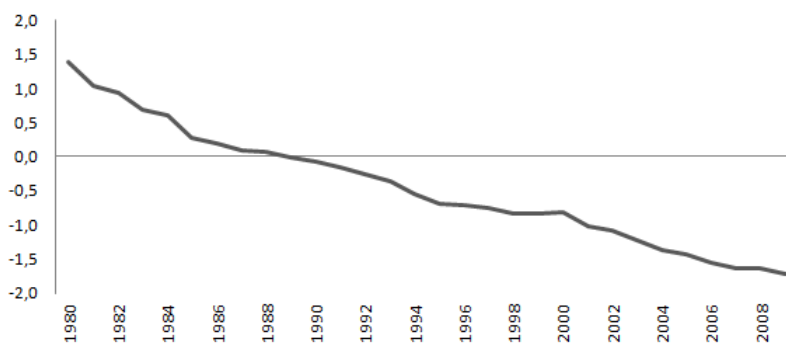


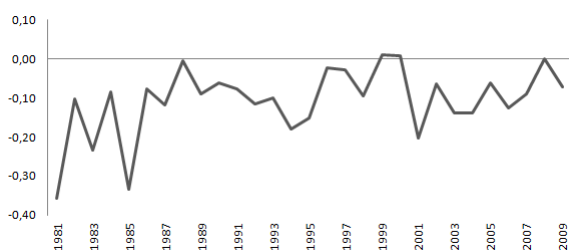
Figura 5.19: Estimativas do índice transformado da fertilidade g_t , 1980-2009

5.2.2 Modelação do índice transformado da fertilidade como uma série temporal

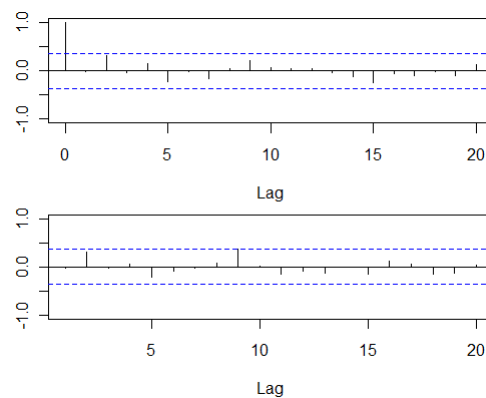
Analogamente à componente da mortalidade, utilizou-se a metodologia de Box-Jenkins para proceder à modelação e projeção do índice transformado de fertilidade g_t como uma série temporal. Posteriormente, partiu-se da sua projeção para estimar o ISF, assim como as taxas de fecundidade específicas por idade fértil da mulher.

5.2.2.1 Identificação e estimação

Pretende-se nesta fase identificar o número de vezes que a série g_t deve ser diferenciada, ou seja, pretende-se definir o parâmetro d do modelo ARIMA a utilizar. Conforme a figura 5.19, observa-se uma tendência claramente decrescente com pequenas oscilações, indicando que a série temporal g_t não é estacionária em torno da média, por essa razão apuraram-se as primeiras diferenças de g_t e respetiva FAC e FACP.



5.20.1: Série temporal



5.20.2: FAC e FACP

Figura 5.20: Primeiras diferenças de g_t , 2010-2050

Na análise das primeiras diferenças de g_t ainda não é visível a estacionaridade em média (figura 5.20.1). Congruentemente a FAC e a FACP parecem assumir um padrão (figura 5.20.2), indiciando também a não estacionaridade da série. Assim, diferenciou-se a série uma segunda vez com o objetivo de a tornar estacionária.

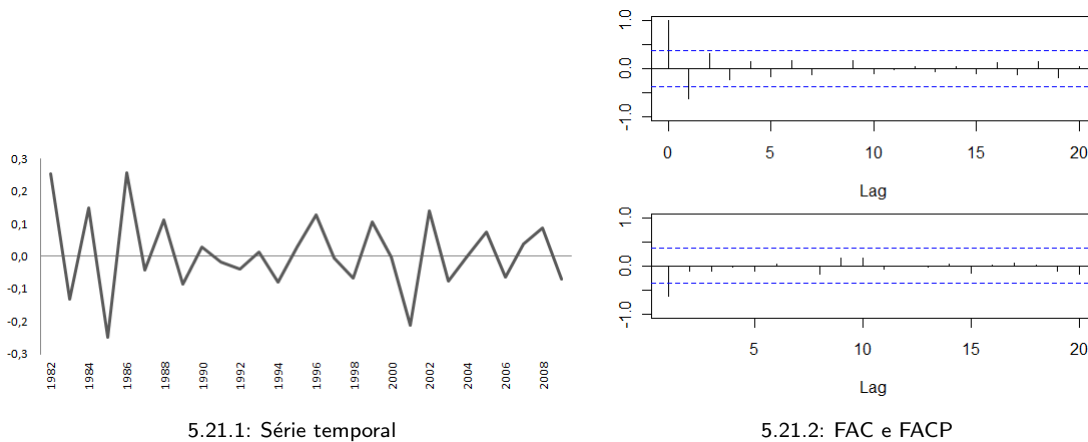


Figura 5.21: Segundas diferenças de g_t , 2010-2050

Analisando os gráficos relativos à série das segundas diferenças de g_t , observa-se que a série já parece ser estacionária com média zero (figura 5.21.1) o que é corroborado pela representação gráfica da sua FAC e FACP (figura 5.21.2).

É então plausível concluir, que o parâmetro d do modelo ARIMA a aplicar na modelação do índice transformado da fertilidade é de 2. Posto isto resolveu-se estimar 9 modelos com combinações de p e q entre 0 e 2 e o parâmetro d igual a 2. Não se inclui nenhum parâmetro *drift* (como se incluiu nos modelos relativos à mortalidade) por, apesar de g_t ser decrescente, não ter um declive que se considere expressivo. Utilizou-se o método dos mínimos quadrados para encontrar os valores iniciais, seguido do método da máxima verosimilhança.

5.2.2.2 Avaliação do diagnóstico

Posteriormente à estimação dos 9 modelos interessa agora identificar aquele que melhor se ajusta à série g_t . Seguiram-se os mesmos passos relatados para a componente da mortalidade. Na tabela 5.7 encontra-se a designação de cada um dos modelos estimados, o número de parâmetros correspondentes e também, diversas medidas que permitem concluir sobre a qualidade e ajustamento do modelo à série histórica de 1980 a 2009.

Imediatamente se excluem os modelos m_3 , m_5 , m_6 , m_7 , m_8 e m_9 por terem ou parâmetros com correlações entre si superiores a 0,7 ou por terem pelo menos um parâmetro não significativo. Desta forma, a escolha recai sobre os modelos m_1 , m_2 e m_4 , dos quais se seleciona o modelo m_4 por ser o que tem menor AIC assim como menor BIC. A série g_t é portanto duas vezes diferenciada sendo estimada por um modelo ARIMA com um parâmetros autoregressivos, ARIMA(1,2,0).

Selecionado o modelo, segue-se a análise dos seus resíduos de forma a aferir sobre a sua adequabilidade à série histórica. A tabela 5.8 expõe os resultados da aplicação do teste de Ljung-Box. Para todos os *lags* analisados não se rejeita a hipótese de adequabilidade do modelo à série (todos os *valores p* são superiores a 0,05).

Examinando a série dos resíduos estandardizados, $\{\hat{\epsilon}_t/S_{\hat{\epsilon}}^2\}$, (figura 5.22.1) e respetivo Q-Q plot (figura 5.22.2), constata-se que estes seguem o padrão de uma série de ruído branco e que se distribuem normalmente. O último

Tabela 5.7: Designação dos modelos e respetivos resultados

| Designação do modelo | p (AR) | q (MA) | AIC | BIC | Correlações entre parâmetros > 0,7 | Parâmetros não significativos |
|----------------------|--------|--------|--------|--------|------------------------------------|-------------------------------|
| m_1 | 0 | 0 | -38,16 | -36,76 | NA | NA |
| m_2 | 0 | 1 | -51,08 | -48,27 | NA | não |
| m_3 | 0 | 2 | -50,17 | -45,97 | não | sim |
| m_4 | 1 | 0 | -52,57 | -49,77 | não | não |
| m_5 | 1 | 1 | -51,55 | -47,35 | sim | sim |
| m_6 | 1 | 2 | -51,05 | -45,44 | não | sim |
| m_7 | 2 | 0 | -51,12 | -46,92 | não | sim |
| m_8 | 2 | 1 | -50,42 | -44,82 | sim | sim |
| m_9 | 2 | 3 | -49,10 | -42,09 | sim | sim |

Tabela 5.8: Resultados da aplicação do teste de Ljung-Box (*valor p*) aos resíduos dos modelos seleccionados

| | Ljung-Box | | | |
|-------|-----------|-------|-------|--------|
| | lag=4 | lag=6 | lag=8 | lag=10 |
| m_4 | 0,47 | 0,76 | 0,79 | 0,59 |

facto foi corroborado pelo resultado do teste estatístico de Shapiro-Wilks, que é o de não rejeitar a hipótese nula de normalidade (*valor p* de 0,42).

Por conseguinte, conclui-se que o modelo ARIMA(1,2,0) se ajusta ao índice transformado da fertilidade g_t . A forma geral deste modelo é:

$$g_t = \phi_1 g_{t-1} + 2g_{t-1} - 2\phi_1 g_{t-2} - g_{t-2} + \phi_1 g_{t-3} + \epsilon_t \quad (5.6)$$

onde ϕ_1 corresponde ao parâmetro autoregressivo do modelo e ϵ à série dos resíduos do modelo. A estimativa do parâmetro do modelo e respetivo erro padrão encontra-se na tabela 5.9. O modelo cumpre as condições de estacionaridade, visto $|\phi_1| < 1$.

Tabela 5.9: Estimativa dos parâmetros para o modelo ARIMA(1,2,0)

| Parâmetro | Coefficiente | Erro padrão | Valor-p |
|-----------|--------------|-------------|---------|
| ϕ_1 | -0,71 | 0,14 | 0 |

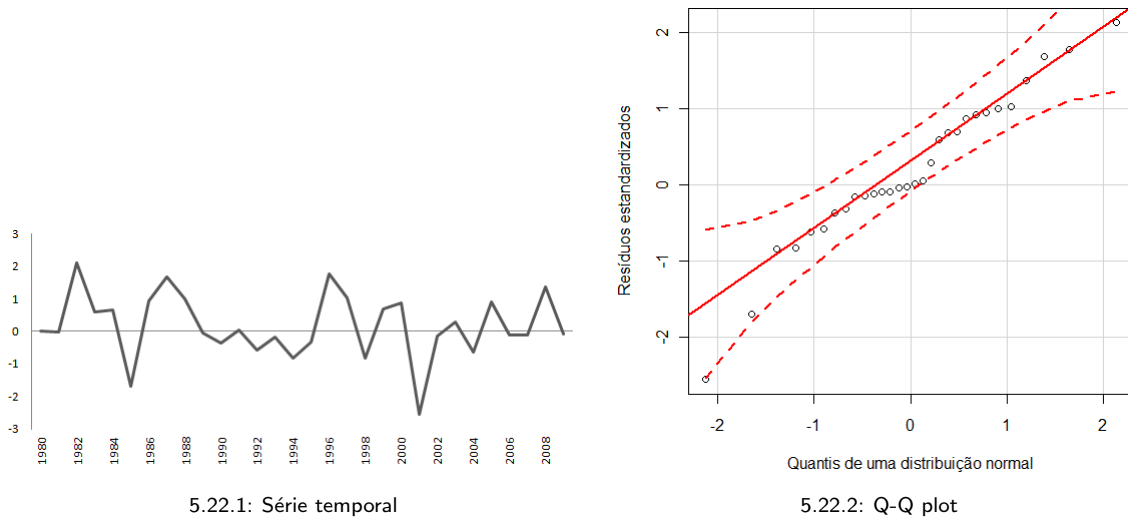


Figura 5.22: Resíduos do modelo

5.2.3 Projeção da série g_t

Após a estimação do modelo ARIMA que melhor se ajusta ao índice transformado da fertilidade, construíram-se projeções para os próximos 41 anos, desde 2010 a 2050.

As figuras 5.23.1 e 5.23.2 expõem a projeção do índice transformado da fertilidade g_t e do índice geral da fertilidade κ_t , respetivamente. Como se pode observar ambos apresentam um comportamento decrescente, sendo limitados aos limites inferior e superior pré-definidos. Analisando atentamente a projeção de g_t , esse facto não se parece evidenciar, contudo analisando κ_t o limite inferior parece reputar influência na sua projeção, nomeadamente nos anos mais avançados.

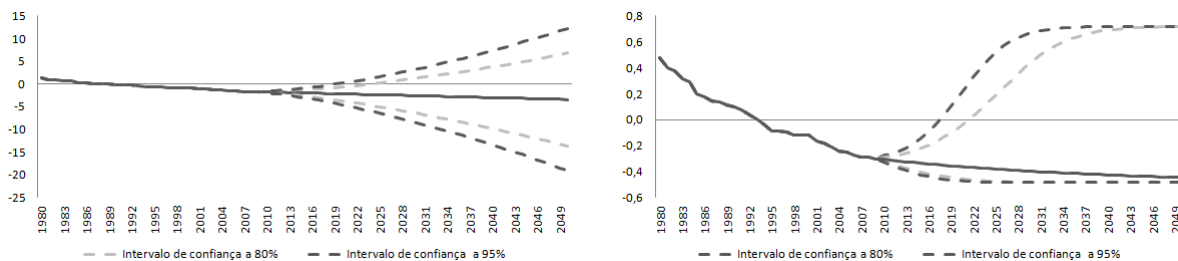


Figura 5.23: Projeções de 2010 a 2050

A projeção do ISF ajustado obtém-se a partir da projeção de g_t , aplicando a transformação inversa à função 5.5, ou seja:

$$F_t = \frac{S \cdot \exp g_t + I}{1 + \exp g_t} \quad (5.7)$$

A série histórica do ISF ajustado e a sua projeção apresenta-se no gráfico 5.24. Como se pode verificar a projeção de F_t segue a tendência decrescente dos dados de base aproximando-se gradualmente dos intervalos de confiança inferiores e consequentemente do limite inferior pré-definido, atingindo em 2050 o valor de 1,14 crianças por mulher.

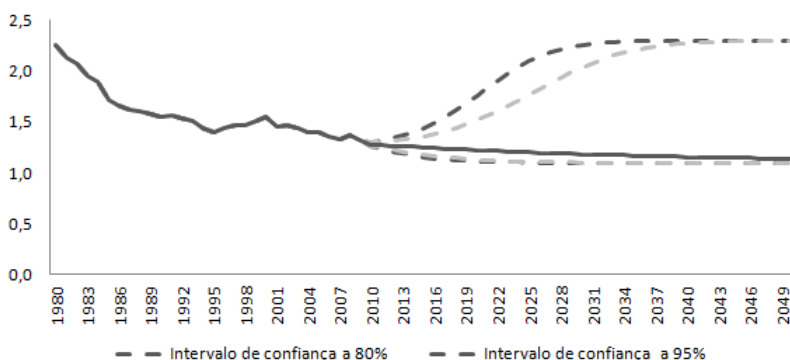


Figura 5.24: Projeção do ISF, 1980-2009

É clara a influência do limite inferior na sua projeção, caso não se estabelecesse este limite o ISF seria com forte probabilidade inferior a 1,1, já o limite superior estabelecido não parece ter qualquer importância na sua projeção.

Os resultados das projeções para o ISF do INE e do Eurostat são superiores aos obtidos neste exercício, sendo de 1,7 e de 1,6 crianças por mulher, respetivamente. Tendo ambos os exercícios uma origem determinística e existindo na altura e ainda hoje uma opinião generalizada de que, com o adiamento da maternidade para idades da mulher mais avançadas o ISF irá sofrer uma recuperação, é natural que as projeções reflitam esse facto. Para além disso, tanto em 2000 como em 2004, os momentos de partida, o ISF era de 1,6 e de 1,4 respetivamente, sendo no momento de partida desta projeção de 1,3 (2009). Contudo, apesar de muitos demógrafos argumentarem que o declínio da fecundidade é uma consequência do adiamento da maternidade, o nível do ISF mantém-se constantemente baixo há já alguns anos, e atualmente não existem razões para esperar uma recuperação considerável (Tuljapurkar, 2006).

5.2.4 Adequação do modelo ao caso português

Como se referiu anteriormente, a projeção das taxas de fecundidade específicas por idade produz valores negativos o que forçou a alteração do limite inferior do ISF de 1 para 1,1. O método de Lee Carter, baseia-se no pressuposto de que os padrões ocorridos ao longo da série histórica, por idade, se repercutem de forma semelhante no futuro (através dos parâmetros θ_a e δ_a). Na mortalidade este pressuposto produz resultados plausíveis, visto a tendência ser sempre decrescente, seja qual for a idade, e não é de todo irreal assumir que a *distância* da mortalidade entre idades é sempre constante. Contudo a fertilidade não é uma componente etariamente estável,

se em 1980 a idade média da mulher ao nascimento de um filho era de 27,2 anos, em 2009 era já de 29,5 anos e portanto torna-se inadequado, utilizar na projeção, parâmetros etários baseados em séries históricas.

A análise do cruzamento das taxas de fecundidade específicas entre o ano de partida (2009) e alguns anos da projeção, revela uma descida acentuadíssima destas nas idades compreendidas entre os 18 e os 24 anos. Apesar de se esperar um decréscimo da fertilidade nestas idades, este não deveria ser tão profundo. Por outro lado, era de esperar uma maior subida nas idades entre os 28 e os 32 anos (figura 5.25).

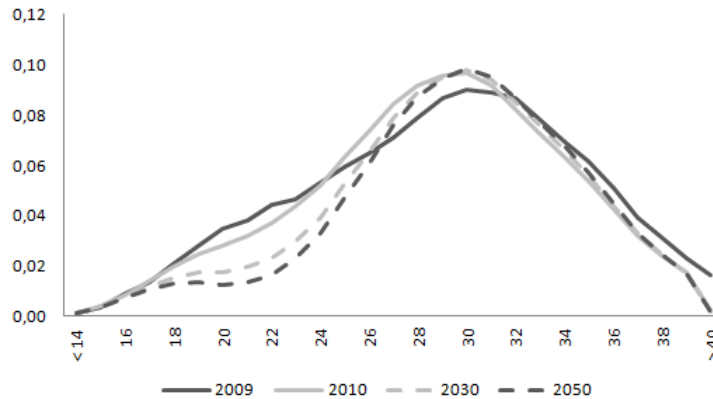


Figura 5.25: Taxas de fertilidade específicas por mulher, 2009, 2010, 2030 e 2050

Em consequência do comportamento das taxas de fecundidade específicas por idade, a idade média da mulher ao nascimento de um filho sofre um ligeiro aumento, fixando-se em 30,4 anos em 2050. Apesar deste indicador, como seria de esperar, aumentar, a sua trajetória não é de todo linear, passando de 29,7 em 2009 para 29,5 em 2010, revelando novamente uma clara inadequação do método utilizado à realidade portuguesa. Ainda para este indicador, os limites impostos ao ISF parecem travar o seu desenvolvimento, encontrando-se a projeção bastante próxima dos intervalos de confiança superiores (figura 5.26).

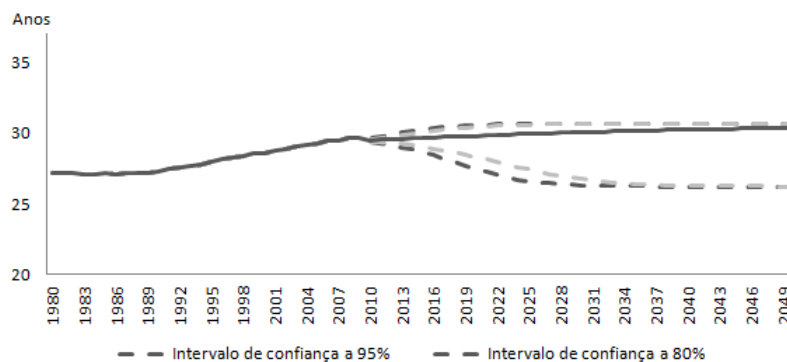


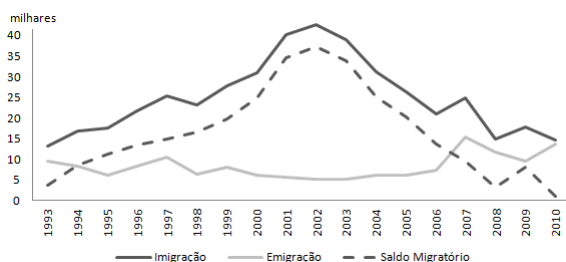
Figura 5.26: Projeção da idade média da mulher ao nascimento do primeiro filho, 2010-2050

5.3 Modelação da componente migratória

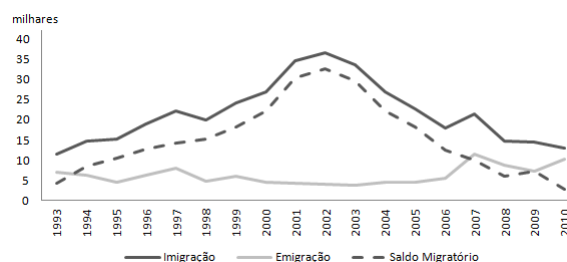
A componente migratória é a componente mais frágil do modelo de projeções da população, por um lado pela ausência de dados com qualidade e que reflitam o verdadeiro estado da migração portuguesa. Por outro, por ser tão diretamente e rapidamente influenciada por decisões de cariz político, económico e demográfico, nomeadamente políticas da imigração e atratividade do mercado de trabalho. Por estas razões a sua projeção tanto a médio como a longo prazo revela-se de extrema dificuldade.

Utilizaram-se estimativas da emigração e imigração desagregados por sexo, desde 1993 a 2010, disponibilizados pelo INE.

O número total de imigrantes, emigrantes e respetivo saldo migratório, representam-se na figura 5.27.1 e 5.27.2 para os homens e mulheres, respetivamente. O padrão dos três processos é semelhante entre sexos, com valores mais elevados no caso do sexo masculino. Em ambos os sexos, observam-se grandes flutuações no que toca à imigração, atingindo o seu valor máximo em 2002 com 42687 imigrantes do sexo masculino e 36613 do sexo feminino. Já a emigração apresenta um comportamento bastante linear relativamente à imigração, variando os seus valores entre cerca de 5000 e 15000 emigrantes para os homens e 4000 e 11000 emigrantes para as mulheres, durante o período analisado.



5.27.1: Homens



5.27.2: Mulheres

Figura 5.27: Processos migratórios por sexo, 1993-2009

Examinando as séries mais atentamente destaca-se o ano de 2001, onde a imigração sofreu um rápido crescimento para ambos os sexos (de 2000 para 2001 aumentou 9425 e 7715, nos homens e mulheres, respetivamente), assim como o ano de 2007 onde se vivenciou um aumento repentino tanto da imigração como da emigração.

Ambos os picos são consequência da tomada de medidas em relação à migração em Portugal. Em 2001 foi criado um novo mecanismo de legalização, as autorizações de residência (Costa, 2004). Este mecanismo tornou visível a crescente emergência do fluxo migratório do Leste da Europa, bem como o crescimento tendencial da população brasileira. Em 2007 foi criada uma nova lei da imigração², que prevê que os imigrantes não abrangidos por processos de regularização extraordinários tenham a oportunidade de obter uma autorização de residência mediante a apresentação de um contrato de trabalho e inscrição na segurança social. Esta lei visa também o alargamento do reagrupamento familiar (Nunes, Nunes).

As leis referidas, nomeadamente a lei de 2007, permitiu que muitos imigrantes que viviam ilegalmente em Portugal se legalisassem, entende-se que é por essa razão que existe um aumento tão significativo nesse ano. Por outro lado, deduz-se que a subida repentina na emigração foi também devido a esta lei, por nem todos os

²Lei n.º 37/2007, de 4 de julho .

imigrantes poderem ficar, muitos foram obrigados a sair. Consequentemente, entende-se que esta subida repentina, aconteceu gradualmente ao longo dos anos e não subitamente, apesar de tal não se reconhecer através da análise da informação.

Tendo em conta este facto, considera-se razoável aplicar a técnica do alisamento³ sobre os três processos migratórios, de forma, a obterem-se séries com padrões mais lineares e também que transpareçam o melhor possível a realidade portuguesa. Para o efeito utilizaram-se splines cúbicos.

A ideia dos splines cúbicos é construir polinómios de terceiro grau por subintervalos, de tal forma, que quaisquer dois deles, definidos em subintervalos contíguos $[x_k, x_{k+1}]$ e $[x_{k+1}, x_{k+2}]$ coincidam em x_{k+1} . Esta coincidência ocorre, não só na função, mas também na primeira e segunda derivada, com o objetivo de alcançar suavidade na própria função, minimizando a sua curvatura.

Devido às dificuldades de modelação da migração, arquitetaram-se 4 ensaios de projeção, pretendendo-se selecionar aquele que, com base em conhecimento empírico, melhor se adequa ao futuro. Para todos os ensaios utiliza-se a informação desagregada por sexos e para a sua projeção a metodologia de Box-Jenkins. Os ensaios considerados são os seguintes:

- **Ensaio 1:** projeção da imigração e emigração sem aplicação de *smoothing*, por sexo;
- **Ensaio 2:** projeção da imigração e emigração com aplicação de *smoothing*, por sexo;
- **Ensaio 3:** projeção do saldo migratório sem aplicação de *smoothing*, por sexo;
- **Ensaio 4:** projeção do saldo migratório com aplicação de *smoothing*, por sexo.

Apesar de, para os dois primeiros ensaios se projetar a imigração e emigração separadamente, em termos de resultados apresenta-se sempre a projeção do saldo migratório por sexo.

Em todos os ensaios construídos a projeção está associada a largos intervalos de confiança, o que remete para a curta série histórica que serve de *input*. De facto, a série considerada tem apenas 18 observações anuais, o que é um número bastante reduzido tendo em conta o número de anos que se quer projetar. Desta forma é de explicitar que apesar de se selecionar o *melhor* ensaio, todos eles são inadequados.

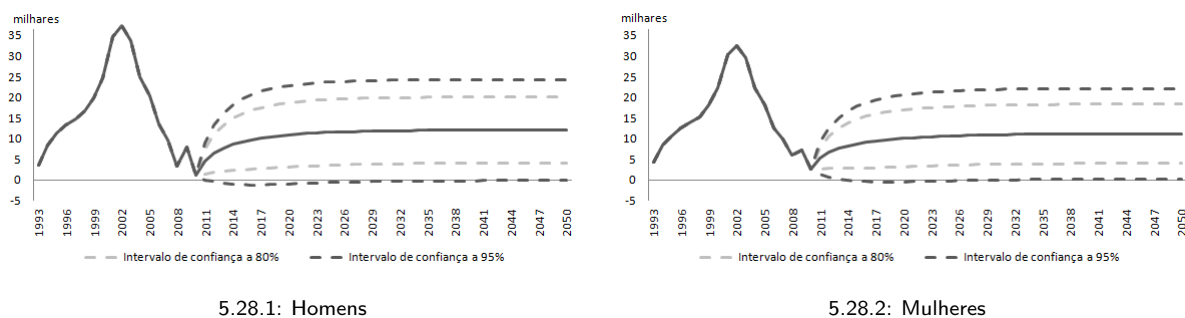
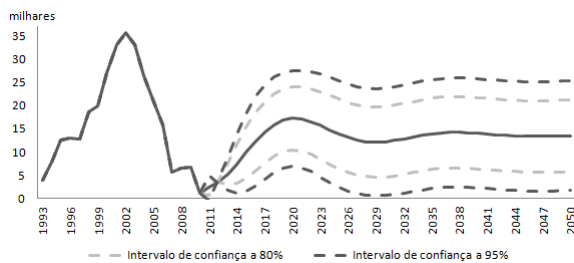
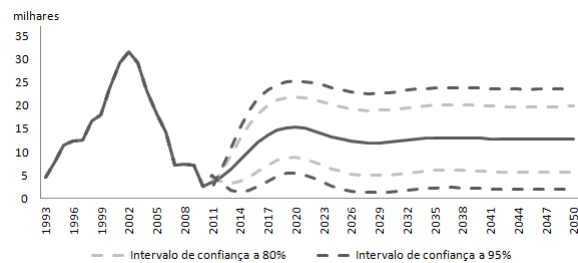


Figura 5.28: Projeção do saldo migratório, ensaio 1, 2010-2050

³Traduzido do inglês *smoothing*.



5.29.1: Homens

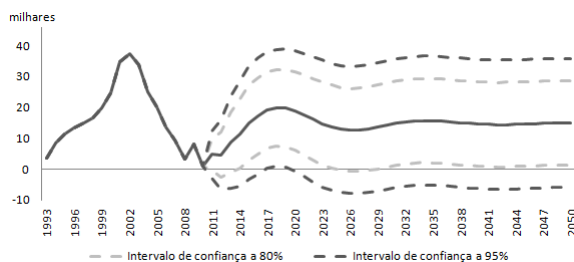


5.29.2: Mulheres

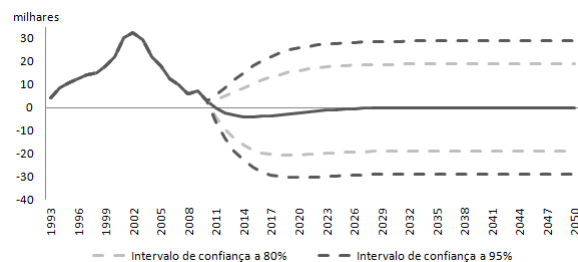
Figura 5.29: Projeção do saldo migratório, ensaio 2, 2010-2050

A projeção do saldo migratório dos ensaios 1 (figura 5.28) e 2 (figura 5.29) é sempre positiva, traduzindo uma imigração bastante superior à emigração, com valores superiores no sexo masculino. Apesar de se aplicar o *smoothing* no ensaio 2, este não surte qualquer efeito a longo prazo. Também nestes ensaios é visível o não seguimento da tendência migratória que, pelas observações mais recentes, deveria decrescer. Pelas razões enunciadas considera-se que estas projeções não são o mais adequadas para a componente migratória.

Já o ensaio 3 (figura 5.30), produz projeções positivas para os homens e projeções inicialmente negativas para as mulheres, que tendem a aproximarem-se de zero à medida que se avança no tempo.



5.30.1: Homens



5.30.2: Mulheres

Figura 5.30: Projeção do saldo migratório, ensaio 3, 2010-2050

Por fim, o ensaio 4 (figura 5.31) produz projeções do saldo migratório bastante semelhantes para os dois sexos. Inicialmente, o modelo projeta saldos migratórios negativos, indicando um maior número de saídas do que entradas. O valor mais baixo do saldo migratório é atingido em 2015 para os homens (saldo migratório de -10581) e em 2014 para as mulheres (saldo migratório de -11278), a partir destes anos o saldo vai recuperando ligeiramente até se fixar em torno de 0, indicando igualdade entre o número de imigrantes e o número de emigrantes, o que ocorre a partir de 2033 para os homens e de 2039 para as mulheres.

Reconhece-se este ensaio como o mais adequado para a componente migratória, tendo em conta a conjuntura económica atual, é razoável admitir que no futuro próximo, Portugal passará por uma fase em que sairão mais pessoas do que as que entram.

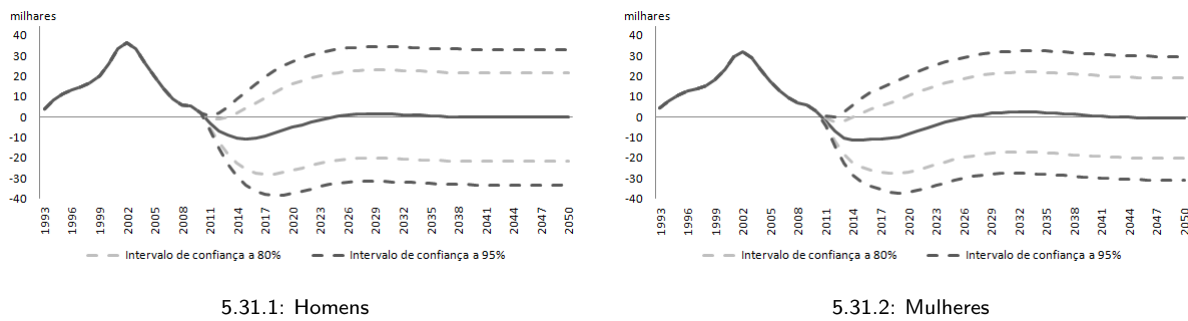


Figura 5.31: Projeção do saldo migratório, ensaio 4, 2010-2050

Ambas as séries representativas do ensaio 4 são estacionárias, não precisando por isso de serem diferenciadas. Assim trabalha-se com modelos ARMA ao invés de modelos ARIMA. Designe-se por X_t a série do saldo migratório masculino e por Y_t a série do saldo migratório feminino. O saldo migratório masculino é estimado através de um modelo com dois parâmetros autoregressivos e um parâmetro de médias móveis, ARMA(2,1), ou seja:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1} \quad (5.8)$$

Onde ϕ_1 e ϕ_2 representam os parâmetros autoregressivos, θ_1 o parâmetro de médias móveis e ϵ_t a série dos resíduos associados a X_t . Por sua vez, o saldo migratório feminino é estimado através de um modelo autoregressivo, com quatro parâmetros, AR(4), isto é:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \epsilon_t \quad (5.9)$$

Onde ϕ_1, ϕ_2, ϕ_3 e ϕ_4 representam os parâmetros autoregressivos do modelo.

As estimativas dos parâmetros e respetivos erros-padrões estão apresentados na tabela 5.10 e na tabela 5.11, para o modelo do saldo migratório masculino e feminino, respetivamente .

Tabela 5.10: Estimativa dos parâmetros do modelo ARMA(2,1), Homens

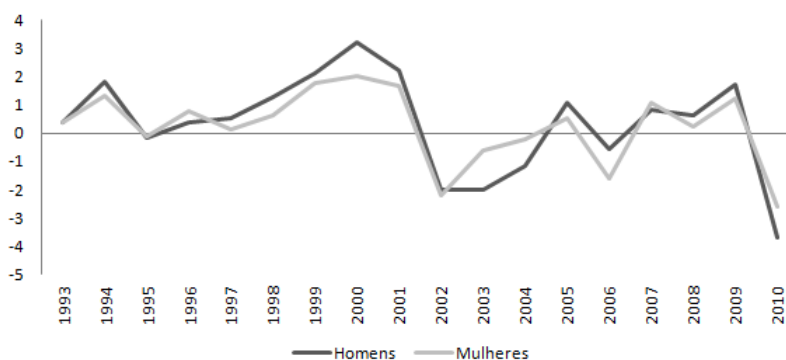
| Parâmetro | Coefficiente | Erro padrão | Valor-p |
|------------|--------------|-------------|---------|
| ϕ_1 | 1,70 | 0,14 | 0 |
| ϕ_2 | -0,76 | 0,14 | 0 |
| θ_1 | 0,68 | 0,18 | 0 |

O modelo estimado para o saldo migratório masculino cumpre as condições de estacionaridade através dos seus parâmetros autoregressivos, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$ e $|\phi_2| < 1$ e também as condições de invertabilidade visto $|\theta_1| < 1$. Também o modelo estimado para o saldo migratório feminino cumpre as condições de estacionaridade requeridas.

Tabela 5.11: Estimativa dos parâmetros do modelo ARMA(4), Mulheres

| Parâmetro | Coefficiente | Erro padrão | Valor-p |
|-----------|--------------|-------------|---------|
| ϕ_1 | 2,63 | 0,22 | 0 |
| ϕ_2 | -2,93 | 0,55 | 0 |
| ϕ_3 | 1,76 | 0,54 | 0 |
| ϕ_4 | -0,49 | 0,21 | 0,01 |

Analisando os resíduos produzidos pelos dois modelos, conforme figura 5.32 realiza-se que ambos se comportam como uma série de ruído branco. Aplicou-se também à séries dos resíduos estandardizados, o teste de normalidade de Shapiro-Wilks, no qual, para ambas as séries, não se rejeitou a hipótese de normalidade dos dados (*valor p* de 0,58 para os homens e de 0,17 para as mulheres).

**Figura 5.32:** Resíduos dos modelos respeitantes ao saldo migratório, ensaio 4

Selecionados e projetados os modelos para o saldo migratório masculino e feminino, pretende-se criar um processo de forma a desagregar o seu total anual por idades. Para o efeito recorreu-se às estimativas de entradas e saídas por idades, dos 0 até aos 85 ou mais anos, por sexo, disponibilizadas pelo INE, para os anos de 2008 e 2009.

Apurou-se, separadamente para as entradas e para as saídas, a média dos dois anos por sexo e idade, obtendo-se assim quatro indicadores: entradas e saídas para o sexo masculino e entradas e saídas para o sexo feminino. Posteriormente, construíram-se proporções para cada uma das idades em relação ao seu total, para cada um dos indicadores. A título de exemplo, a proporção de entradas de população do sexo feminino com 15 anos de idade, é igual, ao quociente entre o número de entradas de mulheres com 15 e o total de entradas do sexo feminino.

Posto isto, decidiu-se que, caso a projeção do saldo migratório seja positiva, isto é caso o número de entradas seja superior ao número de saídas, aplica-se a estrutura etária das entradas e, caso a projeção do saldo migratório seja negativa, isto é, o número de saídas superior ao número de entradas, aplica-se a estrutura etária das saídas (conforme esquema da figura 5.33).

É de notar que com este método, caso a projeção do saldo migratório seja positiva, então esta vai ser positiva em toda a estrutura etária e analogamente para o caso em que é negativa. Reconhecemos que este não é, com toda a certeza, o melhor método, e por isso é importante voltar a referir que a componente migratória sofre de um grande *lack* de informação, o que torna a sua modelação bastante complicada.

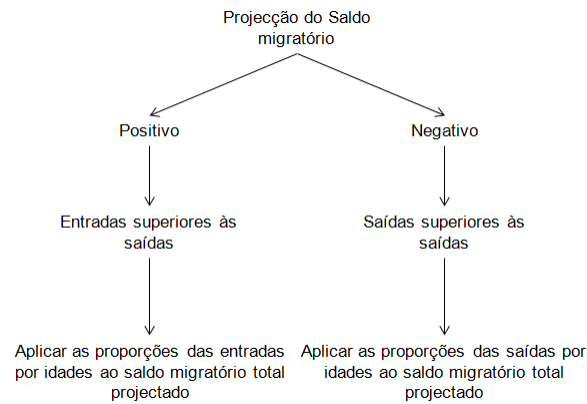
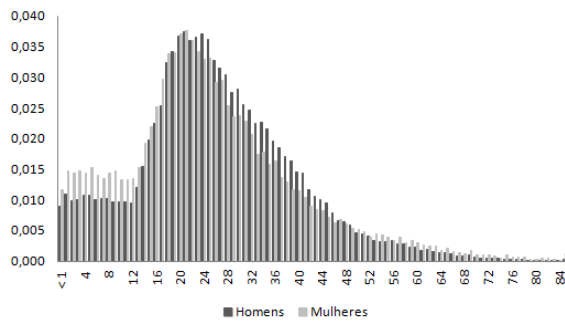
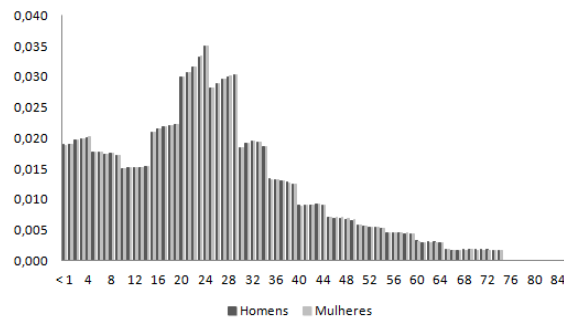


Figura 5.33: Esquema do método utilizado para desagregação da projeção do saldo migratório por idades

Os gráficos 5.34.1 e 5.34.2 mostram a proporção das entradas e das saídas por idades respetivamente, para ambos os sexos. Tanto o padrão das entradas assim como o das saídas é bastante semelhante entre os dois sexos.



5.34.1: Entradas



5.34.2: Saídas

Figura 5.34: Distribuição da proporção de entradas e saídas por sexo e idade, 2008/2009

A maior concentração de entradas encontra-se, nas idades compreendidas entre os 16 e os 32 anos de idade, decrescendo gradualmente a partir destas. Nas idades mais jovens, o sexo feminino, destaca-se com proporções de entradas ligeiramente superiores ao sexo masculino. As saídas de Portugal apresentam um comportamento menos variável ao longo da estrutura etária relativamente às entradas, e também bastante semelhante entre sexos. Saliente-se o facto de não existirem saídas a partir dos 75 anos de idade e o maior número de saídas estar entre os 20 e os 30 anos, seguindo-se as idades adjacentes.

5.4 Projeção da população

Realizada a modelação e a projeção das três componentes populacionais: a mortalidade, a fertilidade e o saldo migratório segue-se o apuramento da projeção da população. Executou-se a projeção da população passo-a-passo de acordo com a matriz 4.22. Tratou-se inicialmente a população feminina, calculando os nascimentos de ambos os sexos para todo o horizonte da projeção, e só posteriormente a população masculina.

No modelo aplicado assume-se que todos os nascimentos, óbitos e migrações acontecem no último dia do ano. Sendo uma projeção a longo prazo este pressuposto provoca apenas alterações insignificantes. Além disso o nosso modelo assume que as populações imigrantes adotam diretamente as condições de fertilidade e mortalidade portuguesas.

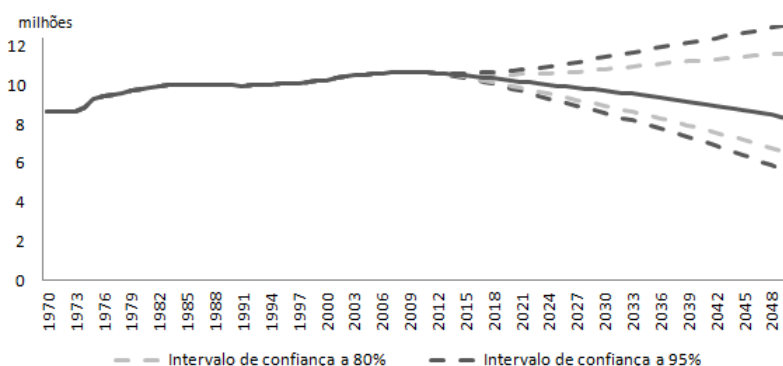


Figura 5.35: Projeção da população total portuguesa, 2010-2050

Parte-se da população residente em Portugal para 2009 por sexo e idade, disponível no portal do INE. Estando esta população desagregada até aos 85 ou mais anos de idade surge a necessidade de uma maior desagregação, nomeadamente até aos 100 ou mais anos de idade, de forma a corresponder às idades para as quais foram projetadas as probabilidades de sobrevivência. Para o efeito, utilizou-se informação da HMD garantindo que o número de indivíduos com 85 anos ou mais é exatamente o mesmo que foi extraído do portal do INE.

Segundo a projeção para 2050 (figura 5.35) a população decai substancialmente em 40 anos, passando de cerca de 10 milhões de habitantes para 8 milhões. Como é possível observar os intervalos de confiança superiores da projeção da população estão a uma distancia superior da média projetada relativamente aos intervalos de confiança inferiores, o que reflete as decisões tomadas para a componente da fertilidade conjugadas com uma menor mortalidade e um saldo migratório inicialmente negativo. É de reparar também que a projeção da população para 2050 tem sensivelmente o mesmo número de habitantes que existiam em 1970 em Portugal, podendo ser indicativo de um ciclo populacional com o seu máximo a ser atingido em 2010.

Os resultados da projeção da população para 2050 estão abaixo dos resultados obtidos pelo INE e pelo Eurostat. Para o primeiro, o valor da projeção ronda os 9 milhões e 300 mil habitantes, e para o segundo, 9 milhões e 900 mil habitantes. Esta diferença, é consequência das projeções individuais das componentes populacionais. Contudo, é de notar que, os valores de ambas as projeções oficiais estão contidas no intervalo de confiança da projeção da população que é realizada nesta dissertação.

A taxa de crescimento efetivo populacional (figura 5.36) repercute o comportamento populacional ao longo do horizonte temporal da projeção. Assume sempre valores negativos e o seu comportamento é geralmente decrescente, sendo exceção os anos de 2020 a 2024. De 2009 para 2050 decresce cerca de 1,2 pontos percentuais.

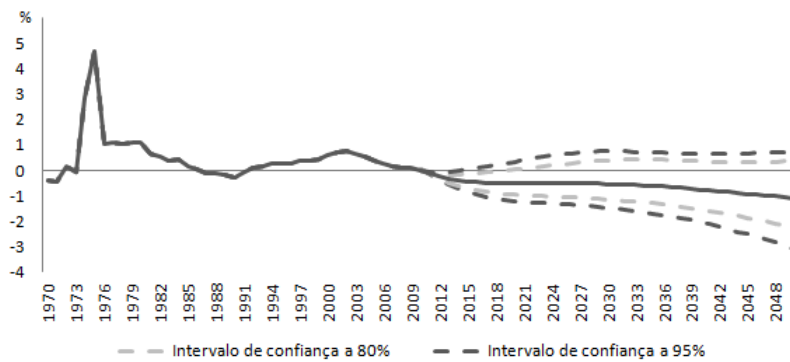


Figura 5.36: Taxa de crescimento efetiva populacional, 2010-2050

A figura 5.37 expõe a série histórica do saldo total, natural e migratório de 1970 a 2009 e a sua projeção a partir de 2010 (por uma questão de facilidade de leitura não se representam os intervalos de confiança). Como se pode observar, desde 1970, que o saldo natural segue uma trajetória geralmente decrescente, marcando o ano de 2007 o primeiro momento que experiencia valores negativos, isto é, a primeira vez que o número de óbitos ultrapassa o número de nados-vivos. Por outro lado, o saldo migratório apresenta um comportamento com grandes variações e é, ao longo da série histórica, a componente que claramente determina o saldo total, todavia, a partir do momento da projeção tal deixa de acontecer, sendo o saldo total induzido pelo saldo natural.

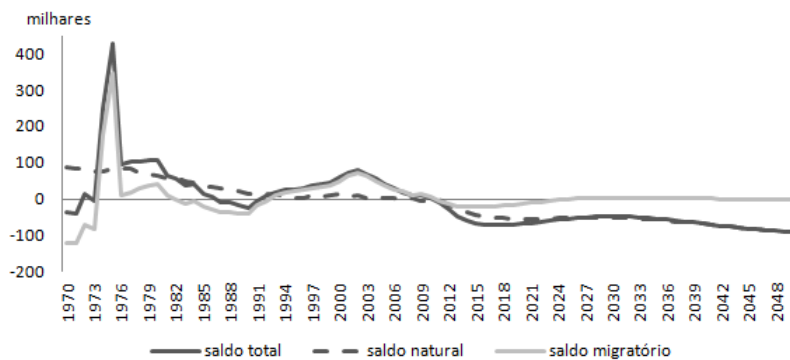


Figura 5.37: Saldo total, natural e migratório, 2010-2050

Apesar do decréscimo da população ser bastante equilibrado entre sexos, não o é na estrutura etária esperando-se uma acentuada diminuição da população jovem e um aumento gradual da proporção de população idosa.

De facto, observando o cruzamento da pirâmide etária de 2009 com a pirâmide etária de 2050 (figura 5.38), observa-se um grande aumento da proporção da população nas idades a partir dos 60 anos. Por outro lado existe uma diminuição substancial na proporção de nados-vivos, jovens e indivíduos em idade ativa. Este é um fenómeno relativamente recente que surgiu nas sociedades desenvolvidas - o duplo envelhecimento demográfico - temos por um lado o envelhecimento na base da pirâmide demográfica, caracterizado por uma diminuição significativa da percentagem de jovens, e por outro, o envelhecimento de topo, efeito de um aumento da percentagem de idosos, fazendo assim com que a topo da pirâmide comece a alargar (Nazareth, 2009). Este envelhecimento está claramente evidenciado na alteração do perfil das pirâmides demográficas entre 2009 e 2050.

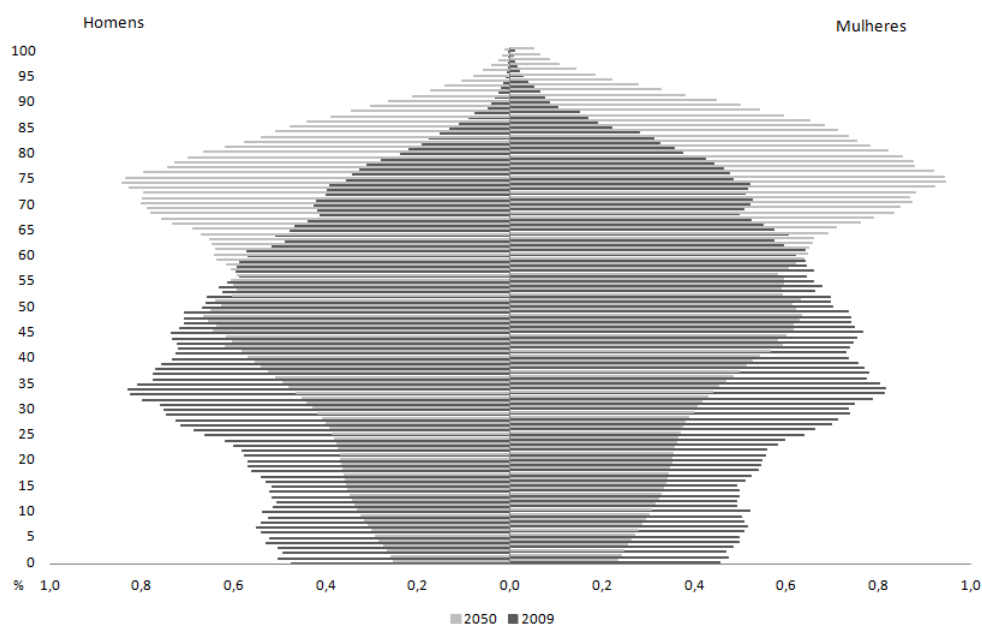
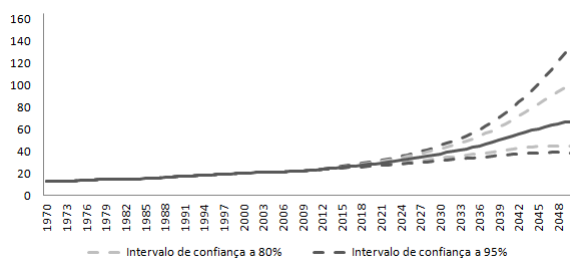
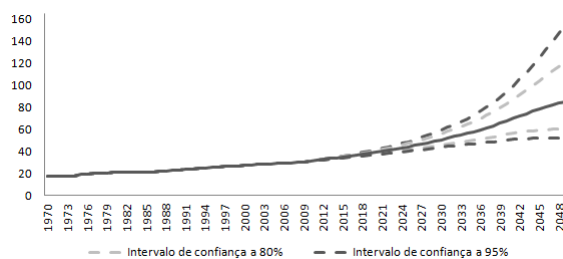


Figura 5.38: Pirâmide etária da população portuguesa, 2009 e 2050

O índice de dependência de idosos⁴ (figura 5.39) é ligeiramente superior nas mulheres relativamente aos homens, tanto na série histórica como na sua projeção, contudo em termos de evolução esta é bastante semelhante, aumentado cerca de 54 idosos por 100 indivíduos em idade ativa de 2009 para 2050 em ambos os sexos. A superioridade do sexo feminino, isto é, um maior número de mulheres com mais de 65 anos, é consequência de uma maior esperança de vida.



5.39.1: Homens



5.39.2: Mulheres

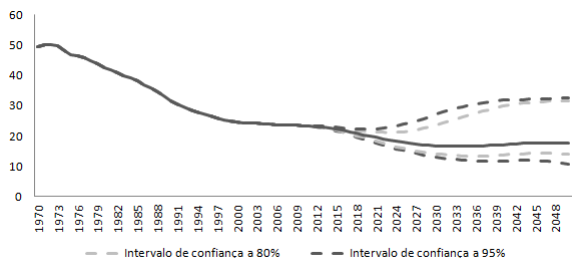
Figura 5.39: Projeção do índice de dependência de idosos por sexo, 2010-2050

Assimetricamente encontra-se o índice de dependência dos jovens⁵ que sofre inicialmente um decréscimo bastante acentuado para se tornar praticamente invariável a partir de 2030, fixando-se por volta dos 17 jovens por 100 indivíduos em idade ativa (figura 5.40) tanto no sexo masculino como no sexo feminino. Analogamente

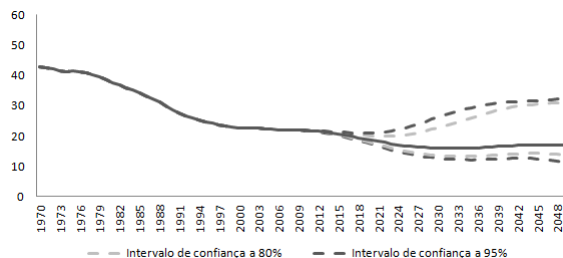
⁴O índice de dependência de idosos representa o número de idosos (indivíduos com 65 ou mais anos de idade) por cada 100 indivíduos na idade ativa (a idade ativa compreende-se entre os 15 e os 64 anos).

⁵O índice de dependência dos jovens representa o número de jovens (indivíduos com 14 ou menos anos de idade) por cada 100 indivíduos na idade ativa

à projeção da população, a distância entre a projeção média e os intervalos de confiança superiores do índice de dependência dos jovens é bastante mais elevada relativamente aos intervalos de de confiança inferiores, o que é uma consequência direta dos limites impostos à projeção do ISF. A assimetria existente entre o índice de dependência de idosos e o índice de dependência dos jovens é um sinal claro do envelhecimento populacional.



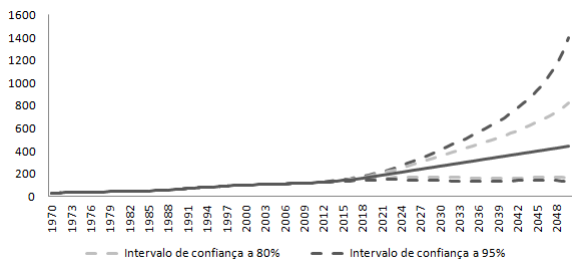
5.40.1: Homens



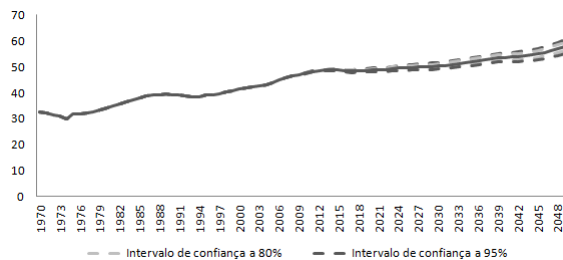
5.40.2: Mulheres

Figura 5.40: Projeção do índice de dependência dos jovens por sexo, 2010-2050

As tendências observadas no índice de dependência dos idosos assim como no dos jovens reflete-se num aumento considerável do índice de envelhecimento. De facto, este índice mostra um aumento brutal do número de idosos por 100 jovens, sendo em 2009 de 117, é projetado para 2050 com o valor extraordinário de 445 idosos por 100 jovens (figura 5.41.1).



5.41.1: Índice de envelhecimento



5.41.2: Índice de longevidade

Figura 5.41: Projeção do índice de envelhecimento e longevidade, 2010-2050

Também o índice de longevidade sofre uma subida considerável (figura 5.41.2), reflexo de uma maior esperança de vida, sendo que o número de pessoas com mais de 75 anos aumenta mais do dobro, correspondendo a uma evolução populacional de 117%.

Conclusão

O principal objetivo desta tese - realização de uma projeção estocástica para a população portuguesa - foi alcançado, revelando o potencial dos métodos probabilísticos na construção de projeções.

Privilegiou-se a utilização do método de Lee Carter que combina um modelo demográfico com um modelo de séries temporais. Esta metodologia mostrou-se bastante satisfatória com a mortalidade, fornecendo resultados plausíveis e coerentes com as estatísticas oficiais, produzindo intervalos de confiança de pequena amplitude, reflexo de uma boa estimação e ajustamento do modelo.

Em contraste com a mortalidade, o método de Lee Carter aplicado à fertilidade revelou algumas limitações e mostrou-se ineficaz na projeção das taxas de fecundidade específicas por idade. Tal aconteceu devido às diferentes tendências da fertilidade por idade, salientando-se a subida da maternidade tardia e, assimetricamente, a descida da maternidade precoce, nos últimos anos que se imputou no modelo. Conclui-se que o método não se adequa à realidade portuguesa no período temporal em análise. Como trabalho futuro, sugere-se a aplicação de métodos alternativos ao desenvolvido nesta dissertação, como o desenvolvido por Schmertmann (2003) ou a adaptação de distribuições gama ajustadas às taxas de fertilidade.

Na última componente populacional, a migração, o constrangimento da falta de informação obrigou ao desenvolvimento de vários ensaios para a projeção do saldo migratório. Estes ensaios diferenciam-se entre si pelo facto de se modelar o saldo migratório directamente ou em alternativa a emigração e imigração separadamente e também pela aplicação (ou não) da técnica de *smoothing*. Neste contexto foi, com base em conhecimento empírico, que se seleccionou aquele que se considerou adequar melhor ao caso português.

Segundo a projeção desenvolvida para 2050, a população decaí substancialmente em 40 anos chegando a cerca de 8 milhões de habitantes revelando um Portugal cada vez mais envelhecido, com a população idosa (mais de 65 anos) a atingir cerca de 40% do total da população. Os resultados obtidos mostram-se inferiores aos produzidos tanto pelo Eurostat como pelo INE, todavia ambos estão contidos no intervalo de confiança a 80% para a população, revelando uma das vantagens óbvias das projecções estocásticas, a associação de probabilidades à projeção da população. A divergência de resultados é efeito da utilização de diferentes séries históricas, do

momento de partida da projeção e igualmente das metodologias aplicadas.

O envelhecimento demográfico em Portugal, analogamente aos seus congéneres europeus, é efeito de três grandes tendências (Comissão das Comunidades Europeias, 2005):

- Aumento contínuo da longevidade, resultado de progressos consideráveis nos cuidados de saúde e na qualidade de vida;
- Ininterrupto crescimento no número de trabalhadores com mais de 60 anos;
- Taxas de natalidade continuamente baixas, devido a vários fatores, nomeadamente, dificuldades em encontrar emprego, o custo da habitação, a idade dos pais ao nascimento do primeiro filho, a falta de incentivos à natalidade assim como as escolhas na vida profissional e familiar.

As políticas públicas tomadas devem levar estas mudanças demográficas em séria consideração, em todas as áreas de intervenção. As mudanças demográficas estão a criar uma nova sociedade cada vez com menos jovens e jovens adultos, trabalhadores cada vez mais velhos, pensionistas e pessoas muito idosas (mais de 75 anos de idade). Novas formas de solidariedade entre gerações devem ser desenvolvidas com base em apoio mútuo e transferência de competências, conhecimentos e experiência (Comissão das Comunidades Europeias, 2005).

A alternativa dos governos deve ser a de resolver o problema quando este se revelar ou tentar antecipar os resultados negativos e preveni-los. O envelhecimento demográfico tem várias implicações nos sistemas fiscais. O impacto mais óbvio é na despesa pública com as reformas altamente correlacionadas com a dimensão e percentagem da população acima dos 65 anos de idade (idade atual de passagem à reforma) (Muenz, 2005). Cuidados de saúde e cuidados de longa duração são dois outros tipos de despesas do sistema operacional que são influenciados pela idade média crescente da população (Bogaert, 2007).

No início do século XXI, Portugal gastava cerca de 9,8 por cento do Produto Interno Bruto (PIB) em pensões da segurança social. De acordo com as projeções do Comité de Política Económica (EPC)¹ este valor aumentará para 13,2 por cento do PIB em 2050, aumentando cerca de 35% em 50 anos.

Vários fatores podem contribuir para o aumento (ou eventualmente decréscimo) da despesa pública em pensões, todavia a evolução demográfica é provavelmente o fator com mais impacto. Ainda segundo a análise do EPC, caso apenas se considere uma mudança de estrutura demográfica, com todos os restantes parâmetros considerados constantes, o aumento passa de 3,4 para 6,7 pontos percentuais (como percentagem do PIB) entre 2000 e 2050.

Estes números mostram o impacto das mudanças demográficas na sustentabilidade das pensões públicas e a crescente importância da disponibilidade de projecções da população, desagregadas por sexo e idade, de forma a facilitar decisões de carácter social, económico e político. Segundo Tuljapurkar (2006) as projecções estocásticas são vantajosas em contextos que envolvam transferências intergeracionais e que portanto exijam horizontes temporais de longo prazo. Naturalmente a incerteza associada à projeção irá crescer ao longo do tempo, e todas as decisões devem ser avaliadas em termos probabilísticos o que é conseguido somente com projecções estocásticas da população.

¹Acrónimo derivado do inglês, Economic Policy Committee.

Referências Bibliográficas

- Alho, J. (1999). On Probabilistic Forecasts of Population and Their Uses. Conference in 52nd ISI Session.
- Alho, J. and Spencer, B. (2005). *Statistical Demography and Forecasting*. New York: Springer Science and Business Media.
- Almeida, A. N., André, I. M. and Lalande, P. (2002). Novos padrões e outros cenários para a fecundidade em Portugal. *Análise Social* (163), 371–409.
- Bandeira, M. L. (1996a). *Demografia e Modernidade. Família e Transição Demográfica em Portugal*. Lisboa: Imprensa Nacional Casa da Moeda.
- Bandeira, M. L. (1996b). Teorias da População e Modernidade: o caso português. *Análise social* **36**(135), 7–43.
- Bertino, S., Sonnino, E. and Lanzieri, G. (2010). Combining Deterministic and Stochastic Population Projections. p. 485–490. Work session on demographic projections: Eurostat.
- Bogaert, H. (2007). Long-Term Population Projections in Europe: how they influence policies and accelerate reforms. Paper prepared for the Joint Eurostat-UNECE Work Session on demographic Projections.
- Box, G. and Jenkins, G. (1976). *Time Series Analysis: forecasting and control* (2 ed.). Holden-Day.
- Bozdogan, H. (1987). Model-selection and Akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika* **52**, 345–370.
- Brouhns, N., Denuit, M. and Vermunt, J. K. (2002). Measuring the longevity risk in mortality projections. In *Bulletin of the Swiss Association of Actuaries*, Volume 2, p. 105–130.
- Carrilho, M. and Lurdes, P. (2002). A situação demográfica recente em Portugal. *Revista de Estudos Demográficos* **32**, 147–175.
- Carrilho, M. J. (2005). Metodologias de cálculo das projecções demográficas: aplicação em Portugal. *Revista de Estudos Demográficos* (37), 5–24.
- Castles, S. and Miller, M. J. (2003). *The Age of Migration: International Population Movements in the Modern World* (3 ed.). Guilford Press.

- Coelho, E. (2001). O método de Lee-Carter para Previsão da Mortalidade. *Revista de Estudos Demográficos* **37**(2), 25–34.
- Coleman, D. (2006). Immigration and ethnic change in low-fertility countries: a third demographic transition. *Population and Development Review* **32**(3), 401–446.
- Comissão das Comunidades Europeias (2005). Green Paper - Confronting demographic change: a new solidarity between the generations. Communication from the comission.
- Conim, C. (1996). Perspectivas de evolução da população - Portugal, 1994-2024. *Prospectiva e Planeamento* **2**, 143–181.
- Costa, P. M. (2004). *Políticas de Imigração e as Novas Dinâmicas da Cidadania em Portugal*. Instituto Piaget.
- Cunha, V. (2007). *O lugar dos filhos: ideais, práticas e significados*. Lisboa: Imprensa de Ciências Sociais.
- Eurostat (2004). Summary note on assumptions and methodology for international migration. Technical report, Luxembourg. Working Paper for the Ageing Working Group of the EPC - ESTAT/F1-/POP/19(2004)/GL.
- Ferreira, V. S. and Nunes, C. (2010). Les marqueurs de passage à l' âge adulte en Europe. *Politiques Sociales et Familiales* (102), 21–38.
- Gans, H. A. (1999). *Population Forecasting 1895-1945 - The Transition to Modernity (European Studies of Population)* (1 ed.). Kluwer Academic Publishers.
- Giroso, F. and King, G. (2007). Understanding the Lee-Carter Mortality Forecasting Method.
- Härdle, W. and Mysickova, A. (2009). Stochastic Population Forecast for Germany and its Consequence for the German Pension System. Sfb 649 discussion papers, Humboldt University, Collaborative Research Center 649.
- Haug, W., Crompton, P. and Courbage, Y. (2002). The demographic characteristics of Immigrant Population.
- INE, I. P. (2010). Retrato Territorial de Portugal 2009.
- INE, I.P. (2003). Projecções da População Residente 2000-2050. Technical report.
- Keifitz, N. (1981). The limits of population forecasting. *Population and Development Review* **7**(4), 579–593.
- Keilman, N. (2008). Using deterministic and Probabilistic Population Forecasts. p. 22–28. Complexity Interdisciplinary Communications: Centre for Advanced Study. Norwegian Academy of Science and Letters.
- Keilman, N., Pham, D. Q. and Hetland, A. (2002). Why Population Forecasts should be Probabilistic - illustrated by the case of Norway. *Demographic Research - Max Plank Institute* **6**(15), 408–454.
- Koissi, M. Shapiro, A. and Högnäs, G. (2005). Fitting and Forecasting Mortality Rates for Nordic Countries Using the Lee-Carter method.
- Lee, R. (1998). Probabilistic Approaches to Population Forecasting. *Population and Development Review* **24**, 156–190. Supplement: Frontiers of Population Forecasting.
- Lee, R. and Carter, L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* **87**(419), 659–671.
- Lee, R. and Tuljapurkar, S. (1994). Population Forecasts for the United States: Beyond High, Medium, and Low. *Journal of the American Statistical Association* **89**(428), 1175–1189.

- Lee, R. and Tuljapurkar, S. (2000). Population Forecasting for Fiscal Planning: Issues and Innovations. *CEDA Papers*.
- Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting* **9**, 187–202.
- Lesthaeghe, R. and Van de Kaa, D. J. (1986). Two demographic transitions? *Population Growth and Decline*, 9–24.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, J. E. (1993). Theories of international migration: A review and appraisal. *Population and Development Review* **19**(3), 431–466.
- McNown, R., Rogers, A. and Knudsen, C. (1993). Projections of Fertility, Mortality and the Population of the United States: 1990-2050. paper presented at the Population Association of American Annual Meetings.
- Muenz, R. (2005). Dimensions and impacts of demographic aging: the case of Europe's public pension system. Paper for the Conference: The Long-term Budget Challenge - Public Finance and Fiscal Sustainability in the G7.
- Murteira, B., Müller, D. and Turkman, K. (1993). *Análise de Sucessões Cronológicas*. McGraw-Hill.
- Nazareth, J. M. (2009). *Demografia - A Ciência da População* (3 ed.). Editorial Presença.
- Nunes, C. D. Percursos Migratórios no Feminino: Mulheres Brasileiras em Portugal, school=Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, address=Lisboa, year=2009. Master Thesis.
- Oliveira, I. T. (2008). Fecundidade das populações e das gerações em Portugal, 1960-2005. *Análise Social* **186**, 29–53.
- Oliveira, I. T. and Mendes, M. F. (2010). A diferença de esperança de vida entre homens e mulheres: Portugal de 1940 a 2007. *Análise Social* **194**, 115–138.
- O'Neill, B. C., Balk, D., Brickman, M. and Ezra, M. (2001). A Guide to Global Projections. *Demographic Research* **4**(8), 203–288.
- Pacheco, A. (2001). Notas de séries temporais, 2000/01. Notas de apoio da disciplina de Séries Temporais da Licenciatura em Matemática Aplicada e Computação.
- Pappámikail, L., Pereira, M. L. and Marques, T. (2010). *Necessidades em Portugal: Tradição e tendências emergentes*, Chapter Portugal: na encruzilhada entre mudança e continuidade, p. 77–137. Lisboa: Tinta da China.
- Renshaw, A. and Haberman, A. (2003). Lee-Carter mortality forecasting: a parallel generalised linear modelling approach for England and Wales mortality projections. *Applied Statistics* **52**(1), 119–137.
- Rosa, M. J. V., Seabra, H. and Santos, T. (2004). Contributos dos imigrantes na demografia portuguesa. Technical report, Lisboa.
- Russolillo, M. (2005). *Lee-Carter Mortality Forecasting: Methodological and Computational Issues*. PhD Thesis, Università degli Studi di Napoli Federico II, Napoli, Italia.
- Scherbov, S., Mamolo, M. and Lutz, W. (2006). Probabilistic Population Projections for the 27 EU member States Bases on Eurostat Assumptions.

- Schmertmann, C. P. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic Research* **9**, 81–110.
- Stoto, M. (1983). The accuracy of population projections. *Journal of the American Statistical Association* **78**(381), 13–20.
- Torres, A. R. (2009). Algumas projecções de mortalidade em Portugal. Master Thesis, Faculdade de Ciências da Universidade de Lisboa, Lisboa.
- Tuljapurkar, S. (2006). Population Forecasts, Fiscal Policy, and Risk. Final paper for the conference "Government Spending on the Elderly" at The Levy Economics Institute of Bard College.
- Tuljapurkar, S., Li, N. and Boe, C. (2000). A Universal Pattern of mortality Decline in the G7 Countries. *Nature* (405), 789–792.
- Turkman, M. A. A. (2010). Modelação estatística e meta-análise. Notas de apoio da disciplina de Complementos de Estatística do Mestrado em Bioestatística.
- Wang, J. Z. (2007). Fitting and Forecasting Mortality for Sweden: Applying the Lee-Carter Model. Technical report, Stockholm University.
- Wei, W. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley.
- Wilmoth, J. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical report, U.C. Berkeley.
- Wilmoth, J. (1996). Mortality projections for Japan: A comparison of four methods. In G. Caselli and A. Lopez (Eds.), *Health and Mortality among Elderly Populations*, p. 266–287. Oxford University Press.
- Wilson, T. and Rees, P. (2005). Recent Developments in Population Projections Methodology: A Review. *Population, Space and Place* **11**, 337–360.

Apêndices

A.1 Mortalidade

1. Aplicação do método de Lee Carter onde mc é a matriz M , m é a matriz das taxas de mortalidade específicas por idade, a é o vector por idades α_a , b é o vector por idades β_a , k é o índice de mortalidade γ_t .

```
a<-0
mc<-matrix(0,nlinhas,ncolunas)
for(i in 1:nlinhas)
{for(j in 1:ncolunas)
{a[i]<-mean(m[i,])
mc[i,j]<-m[i,j]-a[i]
}} svd<-svd(mc)
b<-svd$u[,1]/(sum(svd$u[,1]))
vt<-t(svd$v)
k<-svd$d[1]*vt[1,]*sum(svd$u[,1])
```

2. Função re-estimação do índice de mortalidade γ_t , recebe o número de óbitos observados por ano (obitos0), a população observada por ano e idade (populacao0), a população predita com o modelo por ano (obitosP) e o parâmetro α_a (a), β_a (b) e γ_t (k).

```
reestimacao<-function(obitos0, populacao0, obitosP, a, b, k)
{
#constante
d<-0.1
#numero de grupos etários
n_idades<-length(a)
#Diferença entre o número de óbitos observado e o número de óbitos previsto
```

```

dif<-obitos0-obitosP
n_dif<-length(dif)
cat("Dif", dif, "\n")
cat("n_dif", n_dif, "\n")
#indice k
novok<-k
cat("novok", novok, "\n")
#constante auxiliar quando o k está em torno de zero
m<-0
w<-1
for(w in 1:n_dif)
{
#é melhor não ser exactamente zero mas sim uma aproximação
while(dif[w] < -1 || dif[w] > 1)
{
cat("w=", w, "; dif=", dif[w], "\n")
#Diferença negativa e k positivo
while(dif[w] < -1 && novok[w] > 0)
{
if (novok[w] < 0.5 && m > novok[w]) {novok[w] <- novok[w]-1}
else {m<-novok[w]}
novok[w]<-novok[w]*(1-d)}
cat("w=", w, "; dif=", dif[w], "; novok[w]=",novok[w],"\n")
aux<-0
for(i in 1:n_idades)
{aux[i]<-populacao0[i,w]*exp(a[i]+b[i]*novok[w])}
obitosP[w]<-sum(aux)}
dif[w]<-obitos0[w]-obitosP[w]
}
#Diferença negativa e k negativo
while(dif[w] < -1 && novok[w] < 0)
{
if(novok[w] > -0.5 && m < novok[w]) {novok[w] <- novok[w]+1}
else {m<-novok[w]}
novok[w]<-novok[w]*(1+d)}
cat("w=", w, "; dif=", dif[w], "; novok[w]=",novok[w],"\n")
aux<-0
for(i in 1:n_idades)
{aux[i]<-populacao0[i,w]*exp(a[i]+b[i]*novok[w])}
obitosP[w]<-sum(aux)}
dif[w]<-obitos0[w]-obitosP[w]
}
#Diferença positiva e k positivo
while(dif[w] > 1 && novok[w] > 0)
{
if (novok[w] < 0.5 && m > novok[w]) {novok[w] <- novok[w]-1}

```

```

else {m<-novok[w]
novok[w]<-novok[w]*(1+d)}
cat("w=", w, "; dif=", dif[w], "; novok[w]=",novok[w],"\n")
aux<-0
for(i in 1:n_idades)
{aux[i]<-populacao0[i,w]*exp(a[i]+b[i]*novok[w])
obitosP[w]<-sum(aux)}
dif[w]<-obitos0[w]-obitosP[w]
}
#Diferença positiva e k negativo
while(dif[w] > 1 && novok[w] < 0)
{
if(novok[w] > -0.5 && m < novok[w]) {novok[w] <- novok[w]+1}
else {m<-novok[w]
novok[w]<-novok[w]*(1-d)}
cat("w=", w, "; dif=", dif[w], "; novok[w]=",novok[w],"\n")
aux<-0
for(i in 1:n_idades)
{aux[i]<-populacao0[i,w]*exp(a[i]+b[i]*novok[w])
obitosP[w]<-sum(aux)}
dif[w]<-obitos0[w]-obitosP[w]
}
}}
novok<-novok
dif<-dif
}

```

3. Todos os modelos iniciais foram construídos com a função Arima e posteriormente comparados de forma a seleccionar o que melhor se adequa. A título de exemplo encontra-se um modelo com p parâmetros autoregressivos, q parâmetros de médias móveis que é d vezes diferenciado.

```
modelo<-Arima(k, c(p,d,q))
```

4. Testes de diagnóstico que se realizaram a cada um dos modelos estimados, o teste de Ljung-Box e Box-Pierce (função `Box.test`), o cálculo das correlações entre os parâmetros do modelo (`correlacoes`), a significância de cada um dos parâmetros do modelo (`valores_p`) e o valor da bondade de ajustamento (S) e a soma dos quadrados dos resíduos (SSE).

```

Box.test(m1$residuals, type="Ljung-Box", fitdf=p+q)
Box.test(m1$residuals, type="Box-Pierce", fitdf=p+q)
correlacoes<-round(cov2cor(modelo$var.coef),2)
valores_p<-round((1-pnorm(abs(modelo$coef)/sqrt(diag(modelo$var.coef)))))/2,2)
SSE<-sum(modelo$residuals^2)
S<-sqrt(SSE/(70-p+q))

```

5. Testes que se realizaram aos resíduos do modelo seleccionado. Construção do Q-Q plot, do FAC e do FACP e do teste de normalidade Shapiro-Wilks.

```
qq.plot(modelo$residuals/sd(modelo$residuals), xlab="Quantis de uma distribuição normal", ylab="Resíduos estandardizados")
acf(modelo$residuals, 20, ylim=c(-1,1), main=, ylab=)
pacf(modelo$residuals, 20, ylim=c(-1,1), main=, ylab=)
shapiro.test(modelo$residuals/sd(modelo$residuals))
```

6. Projeção de acordo com o modelo seleccionado onde h representa o horizonte da projecção. Seleção da projecção de $\gamma_t(k)$ e respectivos intervalos de confiança, apenas se exemplifica a construção da projecção das taxas de mortalidade específicas por idade para a projecção média.

```
projeccao<-forecast.Arima(modelo, h=41)
proj_k<-as.vector(projeccao$mean)
proj_k_I_80<-projeccao$lower[,1]
proj_k_I_95<-projeccao$lower[,2]
proj_k_S_80<-projeccao$upper[,1]
proj_k_S_95<-projeccao$upper[,2]
proj_taxas<-matrix(0, nrow=dim(m)[1], ncol=41)
nlinhas<-dim(proj_taxas)[1]
ncolunas<-dim(proj_taxas)[2]
for(i in 1:nlinhas)
{for(j in 1:ncolunas)
{proj_taxas[i,j]<-exp(a[i]+b[i]*proj_k[j])}}
```

A.2 Fertilidade

1. Aplicação do método de Lee Carter onde $fertilidadeC$ corresponde à matriz N , fertilidade é a matriz das taxas de fertilidade específicas por idade, a é o vector por idades θ_a , b é o vector por idades δ_a , k é o índice de fertilidade κ_t .

```
a<-0
for(i in 1:nlinhas)
{a[i]<-mean(fertilidade[i,])}
fertilidadeC<-matrix(0,nrow=nlinhas, ncol=ncolunas)
for(j in 1:ncolunas)
{fertilidadeC[,j]<-fertilidade[,j]-a}
svd<-svd(fertilidadeC)
b<-svd$u[,1]/(sum(svd$u[,1]))
vt<-t(svd$v)
k<-svd$d[1]*vt[1,]*sum(svd$u[,1])
```

2. Construção do índice transformado da fertilidade $g_t(g)$, onde Θ é A , e ISF_ajust corresponde ao ISF ajustado construído a partir da soma de Θ com κ_t .

```
A<-sum(a)
BK<-k
```

```
ISF_ajust<-A+BK
limiteI<-1.1
limiteS<-2.3
g<-log((ISF_ajust-limiteI)/(limiteS-ISF_ajust))
```

3. Segue-se a construção dos modelos, avaliação de diagnósticos e análise dos resíduos do modelo seleccionado, todas estas etapas são semelhantes à componente da mortalidade.
4. Projecção de acordo com o modelo seleccionado onde h representa o horizonte da projecção. Projecção de g e respectivos intervalos de confiança, do ISF e do índice κ_t . Apenas se exemplifica a construção da projecção das taxas de fertilidade específicas por idade para a projecção média.

```
projeccao<-forecast.Arima(modelo, h=41)
proj_g<-as.vector(projeccao$mean)
proj_S_95<-projeccao$upper[,2]
proj_S_80<-projeccao$upper[,1]
proj_I_95<-projeccao$lower[,2]
proj_I_80<-projeccao$lower[,1]
ISF_proj <- (limiteS * exp(proj_g) + limiteI)/(1+exp(proj_g))
k_proj<-ISF_proj-A
fertilidade_proj<-matrix(0, nrow=dim(fertilidade)[1], ncol=41)
nlinhas<-dim(fertilidade_proj)[1]
ncolunas<-dim(fertilidade_proj)[2]
for(i in 1:nlinhas)
{for(j in 1:ncolunas)
{fertilidade_proj[i,j]<-a[i]+b[i]*%*%k_proj[j]}}
```

A.3 Migrações

Neste ponto apresenta-se o código em R para a construção do ensaio 4 - saldo migratório por sexo com *smoothing* - sendo o código semelhante para os outros ensaios.

1. Aplicação de splines cúbicos através da função `smooth.splines` ao saldo migratório masculino (`saldo_H`) e ao saldo migratório feminino (`saldo_M`).

```
sp_saldo_H<-smooth.spline(saldo_H)$y
sp_saldo_M<-smooth.spline(saldo_M)$y
```

2. Para encontrar o modelo de séries temporais que melhor se adequa utilizou-se a função `auto.arima`.

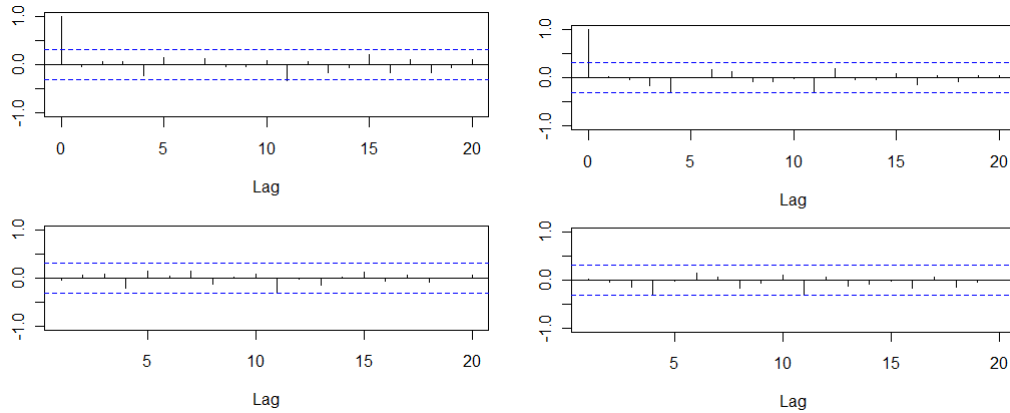
```
modelo_saldo_H_sp<-auto.arima(sp_saldo_H)
modelo_saldo_M_sp<-auto.arima(sp_saldo_M)
```

3. Projecção do saldo migratório por sexo.

```
proj_saldo_H_sp<-forecast.Arima(modelo_saldo_H_sp, h=40)  
proj_saldo_M_sp<-forecast.Arima(modelo_saldo_M_sp, h=40)
```

APÊNDICE B

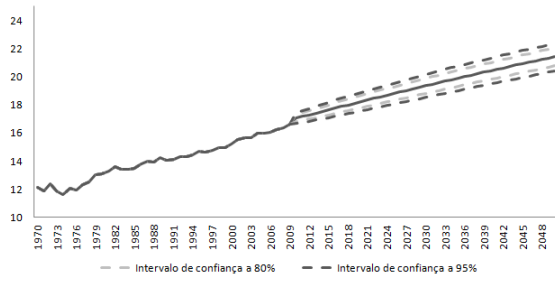
Gráficos e Quadros



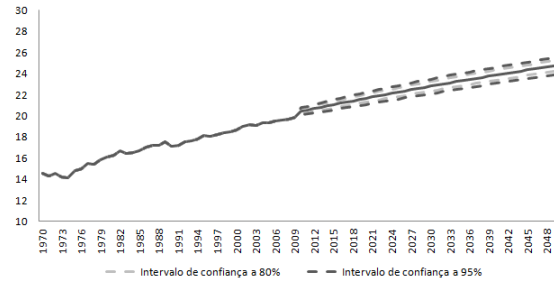
B.1.1: Homens

B.1.2: Mulheres

Figura B.1: FAC e FACP dos resíduos dos modelos ajustados ao índice geral de mortalidade γ_t , por sexo



B.2.1: Homens



B.2.2: Mulheres

Figura B.2: Projeção da esperança média de vida aos 65 anos de idade, 2010-2050

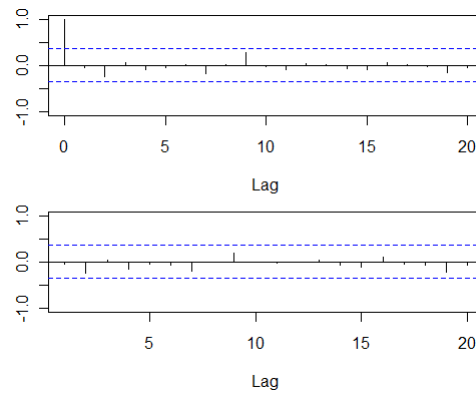
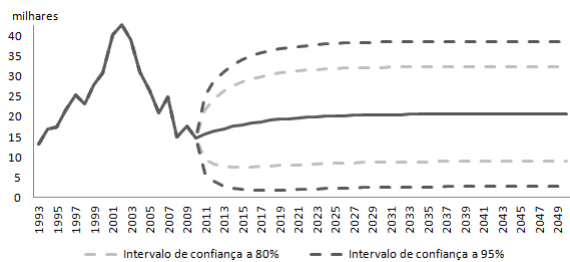
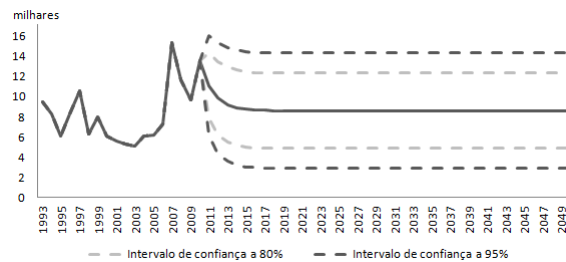


Figura B.3: FAC e FACP dos resíduos dos modelos ajustados ao índice transformado da fertilidade, g_t

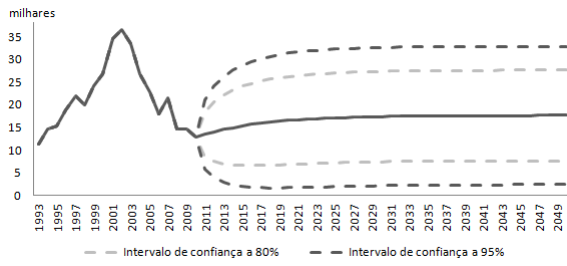


B.4.1: Imigração

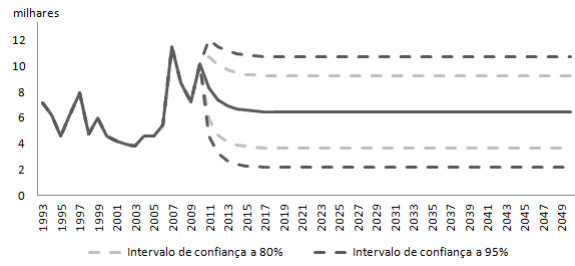


B.4.2: Emigração

Figura B.4: Projeção da imigração e emigração dos homens, ensaio 1, 2010-2050

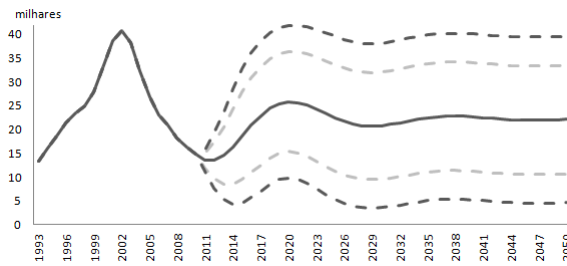


B.5.1: Imigração

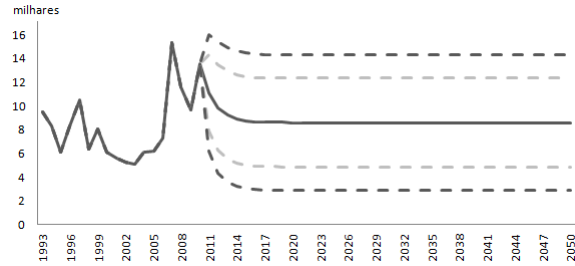


B.5.2: Emigração

Figura B.5: Projecção da imigração e emigração das mulheres, ensaio 1, 2010-2050

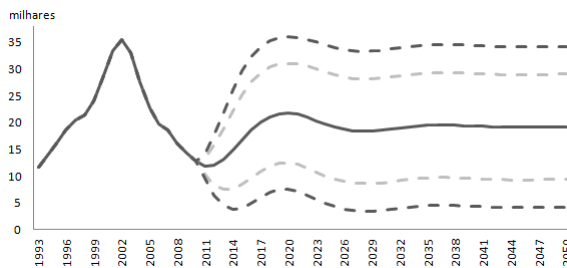


B.6.1: Imigração

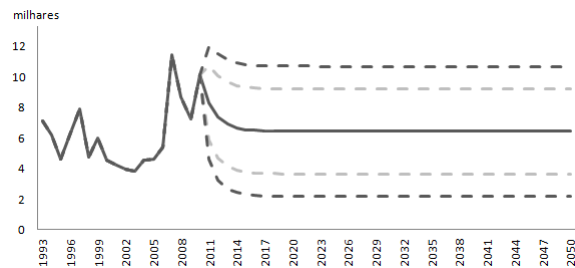


B.6.2: Emigração

Figura B.6: Projecção da imigração e emigração dos homens, ensaio 2, 2010-2050

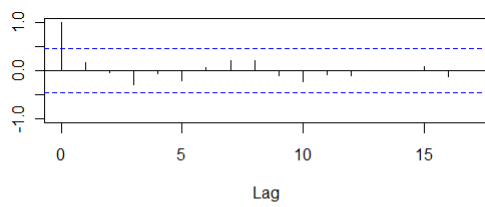


B.7.1: Imigração

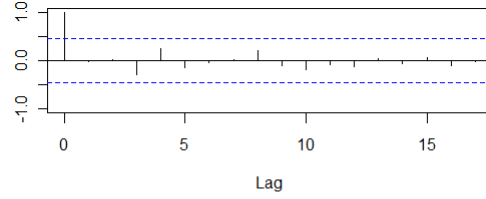


B.7.2: Emigração

Figura B.7: Projecção da imigração e emigração das mulheres, ensaio 2, 2010-2050

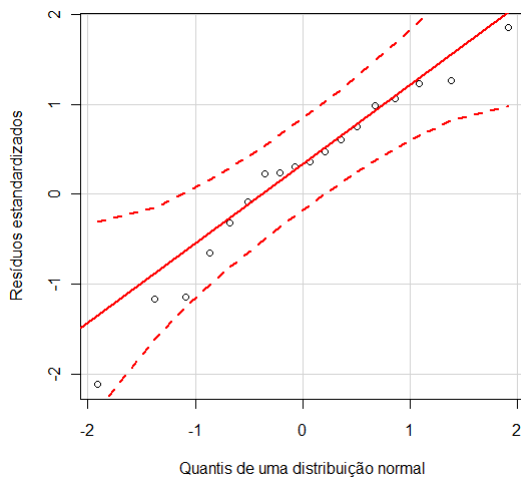


B.8.1: Homens

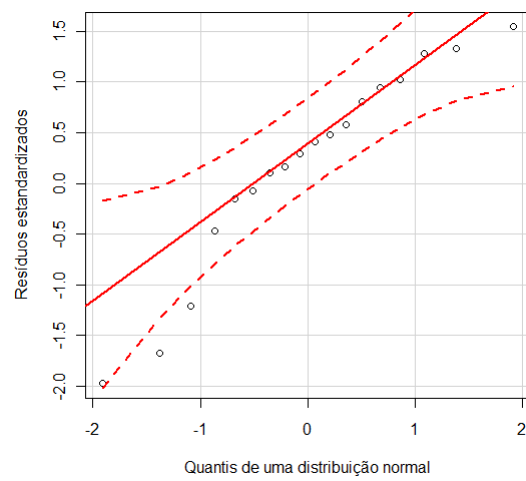


B.8.2: Mulheres

Figura B.8: FAC e FACP dos resíduos dos modelos ajustados ao saldo migratório, por sexo



B.9.1: Homens



B.9.2: Mulheres

Figura B.9: Q-Q plot dos resíduos estandarizados, dos modelos ajustados ao saldo migratório, por sexo

Tabela B.1: Projeção dos principais indicadores demográficos

| Ano Projeção | População | Esperança média de vida - Homens | Esperança média de vida - Mulheres | ISF | Idade média ao nasc.de um filho | Migração Homens | Migração Mulheres |
|-----------------|-----------|-------------------------------------|---------------------------------------|-----|------------------------------------|--------------------|----------------------|
| 2009 | 10637712 | 76,1 | 82,1 | 1,3 | 29,7 | 5295 | 6139 |
| 2010 | 10638036 | 76,4 | 82,6 | 1,3 | 29,5 | 2046 | 3150 |
| 2011 | 10625250 | 76,6 | 82,8 | 1,3 | 29,5 | -3022 | -1914 |
| 2012 | 10600277 | 76,8 | 83 | 1,3 | 29,6 | -6670 | -6947 |
| 2013 | 10565864 | 77,0 | 83,2 | 1,3 | 29,6 | -9025 | -10154 |
| 2014 | 10525207 | 77,2 | 83,4 | 1,3 | 29,6 | -10263 | -11278 |
| 2015 | 10480376 | 77,3 | 83,5 | 1,2 | 29,7 | -10581 | -11203 |
| 2016 | 10432472 | 77,5 | 83,7 | 1,2 | 29,7 | -10186 | -10880 |
| 2017 | 10382082 | 77,7 | 83,9 | 1,2 | 29,7 | -9274 | -10653 |
| 2018 | 10329927 | 77,9 | 84,1 | 1,2 | 29,8 | -8028 | -10321 |
| 2019 | 10276874 | 78,1 | 84,2 | 1,2 | 29,8 | -6603 | -9578 |
| 2020 | 10223703 | 78,3 | 84,4 | 1,2 | 29,8 | -5130 | -8357 |
| 2021 | 10170853 | 78,4 | 84,6 | 1,2 | 29,9 | -3708 | -6846 |
| 2022 | 10118329 | 78,6 | 84,7 | 1,2 | 29,9 | -2411 | -5305 |
| 2023 | 10066345 | 78,8 | 84,9 | 1,2 | 29,9 | -1285 | -3894 |
| 2024 | 10014765 | 79,0 | 85 | 1,2 | 29,9 | -357 | -2638 |
| 2025 | 9963442 | 79,1 | 85,2 | 1,2 | 30 | 365 | -1502 |
| 2026 | 9912218 | 79,3 | 85,3 | 1,2 | 30 | 889 | -463 |
| 2027 | 9861109 | 79,5 | 85,5 | 1,2 | 30 | 1232 | 453 |
| 2028 | 9809999 | 79,7 | 85,6 | 1,2 | 30 | 1417 | 1204 |
| 2029 | 9758493 | 79,8 | 85,8 | 1,2 | 30,1 | 1471 | 1766 |
| 2030 | 9706432 | 80,0 | 85,9 | 1,2 | 30,1 | 1424 | 2141 |
| 2031 | 9653466 | 80,2 | 86,1 | 1,2 | 30,1 | 1303 | 2357 |
| 2032 | 9599495 | 80,3 | 86,2 | 1,2 | 30,1 | 1133 | 2443 |
| 2033 | 9544316 | 80,5 | 86,4 | 1,2 | 30,1 | 936 | 2421 |
| 2034 | 9487757 | 80,6 | 86,5 | 1,2 | 30,2 | 731 | 2306 |
| 2035 | 9429599 | 80,8 | 86,6 | 1,2 | 30,2 | 533 | 2115 |
| 2036 | 9369448 | 81,0 | 86,8 | 1,2 | 30,2 | 350 | 1865 |
| 2037 | 9307347 | 81,1 | 86,9 | 1,2 | 30,2 | 192 | 1579 |
| 2038 | 9243149 | 81,3 | 87,1 | 1,2 | 30,2 | 60 | 1277 |
| 2039 | 9176614 | 81,4 | 87,2 | 1,2 | 30,2 | -43 | 975 |
| 2040 | 9107727 | 81,6 | 87,3 | 1,2 | 30,3 | -118 | 685 |
| 2041 | 9036487 | 81,7 | 87,4 | 1,2 | 30,3 | -168 | 417 |
| 2042 | 8962836 | 81,9 | 87,6 | 1,2 | 30,3 | -195 | 176 |
| 2043 | 8886638 | 82,0 | 87,7 | 1,2 | 30,3 | -204 | -31 |
| 2044 | 8808009 | 82,2 | 87,8 | 1,1 | 30,3 | -199 | -201 |
| 2045 | 8727015 | 82,3 | 87,9 | 1,1 | 30,3 | -183 | -334 |
| 2046 | 8643697 | 82,5 | 88,1 | 1,1 | 30,3 | -160 | -430 |
| 2047 | 8558563 | 82,6 | 88,2 | 1,1 | 30,4 | -133 | -492 |
| 2048 | 8471361 | 82,7 | 88,3 | 1,1 | 30,4 | -104 | -523 |
| 2049 | 8382028 | 82,9 | 88,4 | 1,1 | 30,4 | -76 | -527 |
| 2050 | 8291169 | 83,0 | 88,5 | 1,1 | 30,4 | -51 | -509 |

Tabela B.2: Projeção dos principais indicadores demográficos, intervalo de confiança superior a 95%

| Ano Projeção | População | Esperança média de vida - Homens | Esperança média de vida - Mulheres | ISF | Idade média ao nasc.de um filho | Migração Homens | Migração Mulheres |
|-----------------|-----------|-------------------------------------|---------------------------------------|-----|------------------------------------|--------------------|----------------------|
| 2009 | 10637712 | 76,1 | 82,1 | 1,3 | 29,7 | 5295 | 6139 |
| 2010 | 10634710 | 77 | 83,2 | 1,3 | 29,6 | 2046 | 3150 |
| 2011 | 10625443 | 77,2 | 83,3 | 1,3 | 29,8 | 365 | 639 |
| 2012 | 10615370 | 77,5 | 83,6 | 1,3 | 29,9 | 2049 | 242 |
| 2013 | 10608236 | 77,7 | 83,8 | 1,4 | 30 | 5059 | 2338 |
| 2014 | 10606836 | 78 | 84 | 1,4 | 30,2 | 8692 | 5824 |
| 2015 | 10611660 | 78,2 | 84,2 | 1,4 | 30,3 | 12488 | 9335 |
| 2016 | 10622736 | 78,4 | 84,4 | 1,5 | 30,4 | 16153 | 12125 |
| 2017 | 10639730 | 78,6 | 84,6 | 1,6 | 30,4 | 19511 | 14228 |
| 2018 | 10662926 | 78,9 | 84,8 | 1,6 | 30,5 | 22474 | 16082 |
| 2019 | 10692803 | 79,1 | 85 | 1,7 | 30,5 | 25014 | 18046 |
| 2020 | 10729870 | 79,3 | 85,2 | 1,8 | 30,6 | 27148 | 20158 |
| 2021 | 10774234 | 79,5 | 85,3 | 1,8 | 30,6 | 28913 | 22230 |
| 2022 | 10825628 | 79,7 | 85,5 | 1,9 | 30,6 | 30358 | 24072 |
| 2023 | 10883862 | 79,9 | 85,7 | 2 | 30,6 | 31526 | 25619 |
| 2024 | 10948352 | 80,1 | 85,9 | 2 | 30,7 | 32458 | 26921 |
| 2025 | 11018368 | 80,3 | 86 | 2,1 | 30,7 | 33183 | 28066 |
| 2026 | 11093094 | 80,4 | 86,2 | 2,2 | 30,7 | 33725 | 29104 |
| 2027 | 11171758 | 80,6 | 86,4 | 2,2 | 30,7 | 34102 | 30033 |
| 2028 | 11253512 | 80,8 | 86,5 | 2,2 | 30,7 | 34333 | 30822 |
| 2029 | 11337302 | 81 | 86,7 | 2,2 | 30,7 | 34438 | 31446 |
| 2030 | 11422344 | 81,2 | 86,8 | 2,3 | 30,7 | 34439 | 31902 |
| 2031 | 11507802 | 81,4 | 87 | 2,3 | 30,7 | 34358 | 32207 |
| 2032 | 11593195 | 81,5 | 87,1 | 2,3 | 30,7 | 34218 | 32382 |
| 2033 | 11678093 | 81,7 | 87,3 | 2,3 | 30,7 | 34043 | 32441 |
| 2034 | 11762225 | 81,9 | 87,4 | 2,3 | 30,7 | 33851 | 32397 |
| 2035 | 11845432 | 82 | 87,6 | 2,3 | 30,7 | 33659 | 32261 |
| 2036 | 11927508 | 82,2 | 87,7 | 2,3 | 30,7 | 33480 | 32053 |
| 2037 | 12008717 | 82,4 | 87,8 | 2,3 | 30,7 | 33322 | 31795 |
| 2038 | 12089233 | 82,5 | 88 | 2,3 | 30,7 | 33191 | 31510 |
| 2039 | 12169241 | 82,7 | 88,1 | 2,3 | 30,7 | 33088 | 31217 |
| 2040 | 12249108 | 82,9 | 88,3 | 2,3 | 30,7 | 33013 | 30931 |
| 2041 | 12329268 | 83 | 88,4 | 2,3 | 30,7 | 32964 | 30663 |
| 2042 | 12410097 | 83,2 | 88,5 | 2,3 | 30,7 | 32937 | 30423 |
| 2043 | 12491897 | 83,3 | 88,6 | 2,3 | 30,7 | 32929 | 30216 |
| 2044 | 12575123 | 83,5 | 88,8 | 2,3 | 30,7 | 32936 | 30047 |
| 2045 | 12660139 | 83,6 | 88,9 | 2,3 | 30,7 | 32952 | 29916 |
| 2046 | 12747207 | 83,8 | 89 | 2,3 | 30,7 | 32976 | 29823 |
| 2047 | 12836839 | 83,9 | 89,1 | 2,3 | 30,7 | 33004 | 29765 |
| 2048 | 12928871 | 84,1 | 89,3 | 2,3 | 30,7 | 33032 | 29739 |
| 2049 | 13023187 | 84,2 | 89,4 | 2,3 | 30,7 | 33060 | 29738 |
| 2050 | 13120100 | 84,4 | 89,5 | 2,3 | 30,7 | 33086 | 29760 |

Tabela B.3: Projeção dos principais indicadores demográficos, intervalo de confiança superior a 80%

| Ano Projeção | População | Esperança média de vida - Homens | Esperança média de vida - Mulheres | ISF | Idade média ao nasc.de um filho | Migração Homens | Migração Mulheres |
|-----------------|-----------|-------------------------------------|---------------------------------------|-----|------------------------------------|--------------------|----------------------|
| 2009 | 10637712 | 76,1 | 82,1 | 1,3 | 29,7 | 5295 | 6139 |
| 2010 | 10635871 | 76,8 | 83 | 1,3 | 29,6 | 2046 | 3150 |
| 2011 | 10625362 | 77 | 83,1 | 1,3 | 29,7 | -807 | -245 |
| 2012 | 10610020 | 77,2 | 83,4 | 1,3 | 29,8 | -969 | -2246 |
| 2013 | 10593201 | 77,5 | 83,6 | 1,3 | 29,9 | 184 | -1986 |
| 2014 | 10577746 | 77,7 | 83,8 | 1,3 | 30 | 2131 | -96 |
| 2015 | 10564612 | 77,9 | 84 | 1,4 | 30,1 | 4503 | 2226 |
| 2016 | 10554085 | 78,1 | 84,2 | 1,4 | 30,2 | 7037 | 4162 |
| 2017 | 10546053 | 78,3 | 84,4 | 1,4 | 30,3 | 9548 | 5615 |
| 2018 | 10540864 | 78,5 | 84,5 | 1,4 | 30,4 | 11916 | 6943 |
| 2019 | 10539088 | 78,7 | 84,7 | 1,5 | 30,4 | 14070 | 8485 |
| 2020 | 10541348 | 78,9 | 84,9 | 1,5 | 30,5 | 15975 | 10288 |
| 2021 | 10547981 | 79,1 | 85,1 | 1,6 | 30,5 | 17622 | 12166 |
| 2022 | 10559043 | 79,3 | 85,3 | 1,6 | 30,5 | 19015 | 13904 |
| 2023 | 10574817 | 79,5 | 85,4 | 1,7 | 30,6 | 20169 | 15403 |
| 2024 | 10595292 | 79,7 | 85,6 | 1,7 | 30,6 | 21099 | 16689 |
| 2025 | 10620388 | 79,9 | 85,7 | 1,8 | 30,6 | 21824 | 17831 |
| 2026 | 10649941 | 80,1 | 85,9 | 1,8 | 30,6 | 22359 | 18870 |
| 2027 | 10683791 | 80,2 | 86,1 | 1,9 | 30,6 | 22724 | 19795 |
| 2028 | 10721562 | 80,4 | 86,2 | 1,9 | 30,7 | 22940 | 20571 |
| 2029 | 10762509 | 80,6 | 86,4 | 2 | 30,7 | 23027 | 21173 |
| 2030 | 10806008 | 80,8 | 86,5 | 2 | 30,7 | 23011 | 21601 |
| 2031 | 10851219 | 80,9 | 86,7 | 2,1 | 30,7 | 22916 | 21875 |
| 2032 | 10897553 | 81,1 | 86,8 | 2,1 | 30,7 | 22766 | 22019 |
| 2033 | 10944368 | 81,3 | 87 | 2,2 | 30,7 | 22583 | 22050 |
| 2034 | 10991131 | 81,5 | 87,1 | 2,2 | 30,7 | 22387 | 21981 |
| 2035 | 11037378 | 81,6 | 87,3 | 2,2 | 30,7 | 22193 | 21827 |
| 2036 | 11082577 | 81,8 | 87,4 | 2,2 | 30,7 | 22013 | 21604 |
| 2037 | 11126688 | 81,9 | 87,5 | 2,2 | 30,7 | 21855 | 21337 |
| 2038 | 11169578 | 82,1 | 87,7 | 2,3 | 30,7 | 21723 | 21045 |
| 2039 | 11211126 | 82,3 | 87,8 | 2,3 | 30,7 | 21620 | 20749 |
| 2040 | 11251439 | 82,4 | 87,9 | 2,3 | 30,7 | 21545 | 20462 |
| 2041 | 11290715 | 82,6 | 88,1 | 2,3 | 30,7 | 21496 | 20194 |
| 2042 | 11329138 | 82,7 | 88,2 | 2,3 | 30,7 | 21469 | 19954 |
| 2043 | 11366860 | 82,9 | 88,3 | 2,3 | 30,7 | 21460 | 19747 |
| 2044 | 11404274 | 83 | 88,4 | 2,3 | 30,7 | 21467 | 19577 |
| 2045 | 11441743 | 83,2 | 88,6 | 2,3 | 30,7 | 21483 | 19445 |
| 2046 | 11479605 | 83,3 | 88,7 | 2,3 | 30,7 | 21507 | 19352 |
| 2047 | 11518544 | 83,5 | 88,8 | 2,3 | 30,7 | 21534 | 19292 |
| 2048 | 11558575 | 83,6 | 88,9 | 2,3 | 30,7 | 21563 | 19264 |
| 2049 | 11599816 | 83,8 | 89,1 | 2,3 | 30,7 | 21591 | 19262 |
| 2050 | 11642889 | 83,9 | 89,2 | 2,3 | 30,7 | 21616 | 19283 |

Tabela B.4: Projeção dos principais indicadores demográficos, intervalo de confiança inferior a 95%

| Ano Projeção | População | Esperança média de vida - Homens | Esperança média de vida - Mulheres | ISF | Idade média ao nasc.de um filho | Migração Homens | Migração Mulheres |
|-----------------|-----------|-------------------------------------|---------------------------------------|-----|------------------------------------|--------------------|----------------------|
| 2009 | 10637712 | 76,1 | 82,1 | 1,3 | 29,7 | 5295 | 6139 |
| 2010 | 10641280 | 75,7 | 82,1 | 1,3 | 29,3 | 2046 | 3150 |
| 2011 | 10625174 | 75,9 | 82,3 | 1,2 | 29,3 | -6408 | -4466 |
| 2012 | 10586188 | 76 | 82,4 | 1,2 | 29,1 | -15388 | -14137 |
| 2013 | 10526375 | 76,1 | 82,5 | 1,2 | 29 | -23110 | -22646 |
| 2014 | 10449927 | 76,3 | 82,7 | 1,2 | 28,8 | -29218 | -28381 |
| 2015 | 10360942 | 76,4 | 82,8 | 1,2 | 28,7 | -33651 | -31742 |
| 2016 | 10262263 | 76,6 | 83 | 1,1 | 28,4 | -36525 | -33885 |
| 2017 | 10155938 | 76,7 | 83,1 | 1,1 | 28,2 | -38060 | -35534 |
| 2018 | 10043757 | 76,9 | 83,3 | 1,1 | 28 | -38530 | -36723 |
| 2019 | 9927497 | 77,1 | 83,4 | 1,1 | 27,7 | -38221 | -37202 |
| 2020 | 9808699 | 77,2 | 83,6 | 1,1 | 27,5 | -37408 | -36872 |
| 2021 | 9688438 | 77,4 | 83,7 | 1,1 | 27,3 | -36330 | -35923 |
| 2022 | 9567140 | 77,5 | 83,9 | 1,1 | 27,1 | -35179 | -34683 |
| 2023 | 9445324 | 77,7 | 84 | 1,1 | 26,9 | -34097 | -33406 |
| 2024 | 9322993 | 77,8 | 84,1 | 1,1 | 26,7 | -33173 | -32197 |
| 2025 | 9200068 | 78 | 84,3 | 1,1 | 26,6 | -32453 | -31069 |
| 2026 | 9076407 | 78,1 | 84,4 | 1,1 | 26,5 | -31946 | -30031 |
| 2027 | 8952075 | 78,3 | 84,6 | 1,1 | 26,5 | -31638 | -29128 |
| 2028 | 8826994 | 78,4 | 84,7 | 1,1 | 26,4 | -31500 | -28413 |
| 2029 | 8700780 | 78,6 | 84,8 | 1,1 | 26,3 | -31496 | -27914 |
| 2030 | 8573338 | 78,7 | 85 | 1,1 | 26,3 | -31591 | -27619 |
| 2031 | 8444362 | 78,9 | 85,1 | 1,1 | 26,3 | -31753 | -27492 |
| 2032 | 8313840 | 79 | 85,3 | 1,1 | 26,3 | -31953 | -27495 |
| 2033 | 8181633 | 79,2 | 85,4 | 1,1 | 26,3 | -32171 | -27599 |
| 2034 | 8047633 | 79,3 | 85,5 | 1,1 | 26,3 | -32389 | -27784 |
| 2035 | 7911658 | 79,5 | 85,7 | 1,1 | 26,3 | -32594 | -28032 |
| 2036 | 7773295 | 79,6 | 85,8 | 1,1 | 26,3 | -32779 | -28322 |
| 2037 | 7632633 | 79,7 | 85,9 | 1,1 | 26,2 | -32939 | -28637 |
| 2038 | 7489518 | 79,9 | 86 | 1,1 | 26,2 | -33071 | -28956 |
| 2039 | 7343657 | 80 | 86,2 | 1,1 | 26,2 | -33174 | -29267 |
| 2040 | 7195030 | 80,2 | 86,3 | 1,1 | 26,2 | -33249 | -29560 |
| 2041 | 7043623 | 80,3 | 86,4 | 1,1 | 26,2 | -33300 | -29830 |
| 2042 | 6889347 | 80,5 | 86,5 | 1,1 | 26,2 | -33328 | -30070 |
| 2043 | 6732011 | 80,6 | 86,7 | 1,1 | 26,2 | -33338 | -30277 |
| 2044 | 6571749 | 80,7 | 86,8 | 1,1 | 26,2 | -33334 | -30449 |
| 2045 | 6408628 | 80,9 | 86,9 | 1,1 | 26,2 | -33318 | -30585 |
| 2046 | 6242688 | 81 | 87 | 1,1 | 26,2 | -33296 | -30684 |
| 2047 | 6074571 | 81,2 | 87,2 | 1,1 | 26,2 | -33269 | -30750 |
| 2048 | 5903960 | 81,3 | 87,3 | 1,1 | 26,2 | -33241 | -30785 |
| 2049 | 5730771 | 81,4 | 87,4 | 1,1 | 26,2 | -33213 | -30793 |
| 2050 | 5555770 | 81,6 | 87,5 | 1,1 | 26,2 | -33188 | -30778 |

Tabela B.5: Projeção dos principais indicadores demográficos, intervalo de confiança inferior a 80%

| Ano Projeção | População | Esperança média de vida - Homens | Esperança média de vida - Mulheres | ISF | Idade média ao nasc.de um filho | Migração Homens | Migração Mulheres |
|-----------------|-----------|-------------------------------------|---------------------------------------|-----|------------------------------------|--------------------|----------------------|
| 2009 | 10637712 | 76,1 | 82,1 | 1,3 | 29,7 | 5295 | 6139 |
| 2010 | 10640166 | 76 | 82,3 | 1,3 | 29,4 | 2046 | 3150 |
| 2011 | 10625189 | 76,1 | 82,5 | 1,2 | 29,4 | -5236 | -3582 |
| 2012 | 10590965 | 76,3 | 82,6 | 1,2 | 29,3 | -12370 | -11648 |
| 2013 | 10539769 | 76,4 | 82,8 | 1,2 | 29,2 | -18235 | -18322 |
| 2014 | 10475408 | 76,6 | 82,9 | 1,2 | 29,1 | -22657 | -22461 |
| 2015 | 10401268 | 76,8 | 83,1 | 1,2 | 29 | -25666 | -24633 |
| 2016 | 10319571 | 76,9 | 83,2 | 1,2 | 28,9 | -27408 | -25922 |
| 2017 | 10231863 | 77,1 | 83,4 | 1,2 | 28,8 | -28096 | -26922 |
| 2018 | 10139567 | 77,3 | 83,5 | 1,1 | 28,6 | -27972 | -27584 |
| 2019 | 10044160 | 77,4 | 83,7 | 1,1 | 28,5 | -27277 | -27641 |
| 2020 | 9946933 | 77,6 | 83,9 | 1,1 | 28,3 | -26236 | -27002 |
| 2021 | 9848761 | 77,7 | 84 | 1,1 | 28,2 | -25038 | -25859 |
| 2022 | 9749943 | 77,9 | 84,2 | 1,1 | 28 | -23837 | -24514 |
| 2023 | 9650911 | 78,1 | 84,3 | 1,1 | 27,8 | -22739 | -23191 |
| 2024 | 9551641 | 78,2 | 84,5 | 1,1 | 27,6 | -21814 | -21966 |
| 2025 | 9452050 | 78,4 | 84,6 | 1,1 | 27,5 | -21093 | -20835 |
| 2026 | 9352010 | 78,5 | 84,8 | 1,1 | 27,3 | -20581 | -19797 |
| 2027 | 9251588 | 78,7 | 84,9 | 1,1 | 27,1 | -20261 | -18889 |
| 2028 | 9150708 | 78,9 | 85 | 1,1 | 27 | -20106 | -18162 |
| 2029 | 9048995 | 79 | 85,2 | 1,1 | 26,9 | -20085 | -17641 |
| 2030 | 8946346 | 79,2 | 85,3 | 1,1 | 26,8 | -20163 | -17318 |
| 2031 | 8842445 | 79,3 | 85,5 | 1,1 | 26,7 | -20311 | -17160 |
| 2032 | 8737257 | 79,5 | 85,6 | 1,1 | 26,6 | -20501 | -17132 |
| 2033 | 8630622 | 79,6 | 85,7 | 1,1 | 26,5 | -20711 | -17208 |
| 2034 | 8522411 | 79,8 | 85,9 | 1,1 | 26,5 | -20925 | -17369 |
| 2035 | 8412425 | 79,9 | 86 | 1,1 | 26,4 | -21128 | -17597 |
| 2036 | 8300255 | 80,1 | 86,1 | 1,1 | 26,4 | -21312 | -17873 |
| 2037 | 8185968 | 80,2 | 86,3 | 1,1 | 26,3 | -21471 | -18178 |
| 2038 | 8069404 | 80,4 | 86,4 | 1,1 | 26,3 | -21603 | -18491 |
| 2039 | 7950284 | 80,5 | 86,5 | 1,1 | 26,3 | -21706 | -18799 |
| 2040 | 7828580 | 80,7 | 86,7 | 1,1 | 26,3 | -21781 | -19091 |
| 2041 | 7704272 | 80,8 | 86,8 | 1,1 | 26,3 | -21832 | -19360 |
| 2042 | 7577274 | 81 | 86,9 | 1,1 | 26,3 | -21860 | -19601 |
| 2043 | 7447405 | 81,1 | 87 | 1,1 | 26,3 | -21869 | -19808 |
| 2044 | 7314784 | 81,2 | 87,2 | 1,1 | 26,3 | -21865 | -19979 |
| 2045 | 7179468 | 81,4 | 87,3 | 1,1 | 26,3 | -21849 | -20114 |
| 2046 | 7041493 | 81,5 | 87,4 | 1,1 | 26,3 | -21826 | -20212 |
| 2047 | 6901446 | 81,7 | 87,5 | 1,1 | 26,2 | -21799 | -20277 |
| 2048 | 6759029 | 81,8 | 87,6 | 1,1 | 26,2 | -21771 | -20310 |
| 2049 | 6614166 | 81,9 | 87,8 | 1,1 | 26,2 | -21743 | -20317 |
| 2050 | 6467564 | 82,1 | 87,9 | 1,1 | 26,2 | -21718 | -20301 |