



Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**



Grado en Matemáticas e Informática

Trabajo Fin de Grado

**Simulación del movimiento de la
población entre los medios rural y
urbano**

Autor: Gonzalo Losada Álvarez

Tutor(a): Juan A. Fdez del Pozo de Salamanca

Madrid, Enero 2025

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Matemáticas e Informática

Título: Simulación del movimiento de la población entre los medios rural y urbano

Enero 2025

Autor: Gonzalo Losada Álvarez

Tutor: Juan A. Fdez del Pozo de Salamanca
Departamento de Inteligencia Artificial
Escuela Técnica Superior de Ingenieros Informáticos
Universidad Politécnica de Madrid

Resumen

Este trabajo aborda la evolución del movimiento de la población entre medio rural y urbano en España, con un enfoque en la evolución de la proyección de sus poblaciones. La despoblación de las zonas rurales, reflejada en el concepto de “la España vacía”, es un problema de gran relevancia social, económica y territorial. El trabajo abordará este problema desde una perspectiva matemática.

El objetivo principal del estudio es desarrollar un modelo matemático que permita analizar y simular la evolución de la población, identificando las dinámicas que subyacen en las disparidades entre áreas rurales y urbanas. Para ello, se realiza un estudio exhaustivo del estado del arte sobre simulación y modelos predictivos aplicados a la demografía, prestando especial atención a los factores que influyen en las diferencias entre entornos rurales y urbanos.

El trabajo incluye el diseño y ajuste de un modelo matemático adaptado al caso de España, con capacidad para proyectar escenarios futuros de población. Este modelo se implementa computacionalmente con el objetivo de realizar simulaciones representativas de la realidad demográfica española.

Los resultados obtenidos se analizan en profundidad, evaluando su coherencia y comparándolos con otros estudios. Además, se reflexiona sobre las implicaciones del modelo en la planificación de recursos y en la formulación de políticas públicas, con un énfasis especial en los Objetivos de Desarrollo Sostenible (ODS).

Finalmente, se evalúa si los objetivos planteados se han alcanzado y se ofrece una reflexión personal sobre los aprendizajes del trabajo, junto con posibles líneas futuras de investigación.

Abstract

This work addresses the evolution of the population movement between rural and urban areas in Spain. It focuses on the evolution of population growth projections. The depopulation of rural areas, reflected in the concept of “la España vacía”, is a problem of great social, economic, and territorial relevance. This problem will be tackled from a mathematical perspective.

The study’s main objective is to develop a mathematical model that allows for the analysis and simulation of population growth, identifying the dynamics underlying the disparities between rural and urban areas. To achieve this, an extensive review of the state-of-the-art simulation and predictive models applied to demography is conducted, paying special attention to the factors that influence the differences between rural and urban environments.

This work also includes the design and adjustment of a mathematical model tailored to Spain’s case, capable of projecting future population scenarios. This model is implemented computationally to produce simulations that accurately reflect the demographic reality of Spain.

The results obtained are analyzed in depth, evaluating their consistency and comparing them with other studies. Additionally, the implications of the model in resource planning and public policy formulation are reflected upon, with a special emphasis on the Sustainable Development Goals (SDGs).

Finally, the extent to which the objectives set have been achieved is evaluated, and a personal reflection on the lessons learned from the work is provided, along with potential future research directions.

Tabla de contenidos

1. Introducción	1
1.1. Definición y motivación	1
1.2. Objetivos	1
1.3. Descripción del documento	2
2. Estado del arte	3
2.1. Población rural y urbana	3
2.1.1. Factores que influyen en la Evolución de la Población Rural y Urbana en España	4
2.2. Modelos matemáticos	5
2.2.1. Modelo exponencial de Malthus	5
2.2.2. Modelo logístico	6
2.3. Modelo de componentes y modelos de cambio de cohortes	6
2.4. Modelo de simulación basados en agentes	8
2.5. Modelos de simulación basados en análisis de series temporales . .	10
2.5.1. Modelos ARIMA	10
2.5.2. Modelo de caminata aleatoria con deriva	13
2.5.3. Modelo de suavizado exponencial estocástico	14
2.5.4. Modelo de Suavizado Exponencial con Tendencia (Holt) . . .	14
3. Desarrollo del modelo	17
3.1. Proyección de los nacimientos	19
3.2. Proyección de defunciones	21
3.3. Proyección inmigración y emigración exterior	24
3.4. Proyección de saldo migratorio rural a urbano	26
4. Modelo computacional	29
4.1. Paquetes utilizados	29
4.2. Población en el primer año	30
4.3. Proyección tasa fecundidad	31
4.4. Proyección tasa mortalidad	31
4.5. Proyección saldo migratorio exterior	32
4.6. Proyección de la migración rural-urbana	32
4.7. Proyección final de población	32
5. Análisis de resultados	35

TABLA DE CONTENIDOS

5.1. Evolución tasa fecundidad	35
5.2. Tasas mortalidad	37
5.3. Análisis crecimiento vegetativo	38
5.4. Análisis saldo migratorio exterior	39
5.5. Análisis poblacional	41
5.6. Análisis saldo migratorio rural/urbana	45
5.7. Comparación con otros estudios	46
6. Conclusiones e impacto	49
6.1. Evaluación objetivos del trabajo	49
6.2. Posibles líneas futuras de investigación	49
6.3. Evaluación personal del proceso	50
6.4. Impacto	51
Bibliografía	53
Anexos	59
A. Datos: tablas de tasas y poblaciones anuales	59
B. Scripts de R: implementación de la simulación y del análisis	63
B.1. Población en el primer año	63
B.2. Proyección de fecundidad	64
B.3. Proyección de mortalidad	66
B.4. Proyección de saldo migratorio exterior	68
B.5. Proyección de la migración rural-urbana	69
B.6. Proyección de población	70
B.7. Código usado para las gráficas	72

Índice de figuras

5.1. Gráfico de la evolución de la tasa de fecundidad	37
5.2. Gráfico de la evolución de la tasa de mortalidad	38
5.3. Gráfico de la evolución del crecimiento vegetativo	39
5.4. Gráfico de la evolución de población España	41
5.5. Gráfico de la evolución de población urbana	42
5.6. Gráfico de la evolución de población rural	43
5.7. Pirámide poblacional Rural/Urbana	44
5.8. Gráfico de la evolución del saldo migratorio de rural a urbano . . .	46
5.9. Gráfico de la evolución de Saldo Vegetativo del INE	47
5.10. Gráfico de la evolución de la migración exterior del INE	47
5.11. Gráfico de la evolución poblacional INE	48

Índice de cuadros

3.1. Número asignado a cohortes de edad en intervalos de 5 años. . . .	18
3.2. Elaboración propia a partir de valores de tasa de fecundidad del INE	20
3.3. Tasa de mortalidad por 1000 habitantes según grupo de edad en años. Elaboración propia a partir de datos del INE)	23
3.4. Modelos ARIMA óptimos para cada cohorte de edad	25
4.1. Elaboración propia a partir de datos del INE	30
4.2. Distribución porcentual de población rural y urbana por grupo de edad.	31
5.1. Tabla de fecundidades por grupo de edad y cada 10 años	36
5.2. Saldo migratorio exterior por cohorte de edad y año	40
A.2. Tabla de tasa de mortalidad por año, 2022-2012	60
A.4. Tabla de tasa de mortalidad por año, 2011-2002	61
A.6. Tabla de saldo migratorio exterior por año, 2023-2008	62

Capítulo 1

Introducción

1.1. Definición y motivación

La evolución de la población y la división entre áreas rurales y urbanas es un tema de gran relevancia en la actualidad. Conceptos como “la España vacía” resuenan tanto en las noticias como en las conversaciones cotidianas, reflejando una creciente preocupación por la despoblación en muchas zonas rurales. Esta disparidad en el crecimiento poblacional entre zonas rurales y urbanas es un fenómeno global, pero este trabajo se centrará específicamente en el caso de España, donde las implicaciones sociales, económicas y territoriales son particularmente significativas.

En este contexto, las matemáticas ofrecen herramientas poderosas para analizar y modelar problemas reales como este. A través de modelos predictivos, es posible simular escenarios de crecimiento o decrecimiento poblacional, permitiendo no solo comprender mejor los factores que los determinan, sino también facilitar la planificación de recursos y políticas públicas de manera informada.

1.2. Objetivos

El objetivo principal de este trabajo es abordar el problema de la evolución poblacional con división rural y urbana desde una perspectiva matemática, proporcionando herramientas para su análisis y comprensión. En este marco, los objetivos específicos del trabajo son los siguientes:

- Realizar un estudio exhaustivo del estado del arte sobre simulación y modelos matemáticos aplicados a la predicción de poblaciones. Además, analizar las diferencias entre áreas rurales y urbanas, considerando los factores que influyen en estas dinámicas.
- Diseñar un modelo matemático adaptado al caso de España, capaz de generar simulaciones que proyecten escenarios futuros.
- Implementar computacionalmente el modelo desarrollado.

Capítulo 1. Introducción

- Evaluar y analizar los resultados obtenidos analizando su coherencia y contrastándolos con estudios similares para validar la fiabilidad del modelo.

1.3. Descripción del documento

Siguiendo estos objetivos se ha dividido el trabajo en una serie de capítulos:

Estado del arte: Se revisarán los principales modelos matemáticos utilizados en la simulación y predicción de poblaciones, analizando sus aplicaciones, ventajas e inconvenientes. También se analizarán los conceptos de población rural y urbana, viendo los factores que las determinan. Este capítulo sentará las bases teóricas del trabajo.

Desarrollo del modelo: A partir de la revisión teórica, se seleccionará y ajustará el modelo más adecuado para el caso de España. En este capítulo, se detallará el diseño del modelo, incluyendo los factores considerados y los métodos matemáticos empleados.

Implementación computacional: Se describirá la implementación práctica del modelo seleccionado y qué herramientas se han utilizado. Este capítulo incluirá detalles sobre el proceso de ajuste de parámetros y la generación de simulaciones.

Análisis de resultados: Se presentarán y discutirán los resultados obtenidos de las simulaciones, comparando con otros estudios cuando sea posible.

Conclusión e impacto: Finalmente, se analizará si se han alcanzado los objetivos del trabajo. También se añadirá una reflexión personal del trabajo y posibles líneas futuras de investigación. Para finalizar, se analizarán las implicaciones del trabajo y los resultados centrados en los ODS (Objetivos de Desarrollo Sostenible).

Capítulo 2

Estado del arte

Un modelo de evolución de población es un modelo matemático que simula cómo una población cambia con el tiempo, considerando diversos factores como nacimiento, muerte y migración. Estos modelos son esenciales para el estudio de la demografía de un país. Nos permiten entender y predecir cambios demográficos de una población como su crecimiento o envejecimiento. Estos modelos son esenciales para la sociedad, ya que permiten hacer una planificación de recursos y servicios para prevenir las necesidades futuras de determinadas poblaciones. Al proyectar crecimiento o decrecimiento, o envejecimiento las políticas públicas se pueden adaptar para ajustarse a estas previsiones.

Los modelos matemáticos de población han evolucionado considerablemente a medida que han mejorado tanto las herramientas matemáticas como la disponibilidad de datos y la comprensión de los factores que influyen en la dinámica poblacional. Como estos factores modifican la estructura de los modelos, en este trabajo, es fundamental entender las diferencias entre poblaciones rurales y urbanas, ya que cada una presenta características demográficas particulares y esto afectará a los modelos que se utilicen.

2.1. Población rural y urbana

Por población urbana nos referimos a las personas que viven en áreas con alta densidad de población y desarrollo de infraestructura, como ciudades y grandes pueblos. Por otro lado, la población rural se refiere a las personas que residen en áreas menos densamente pobladas, como pueblos pequeños y zonas agrícolas.

La clasificación de las poblaciones en rural y urbana varía significativamente según el lugar. Algunos países emplean criterios cualitativos basados en parámetros como: el acceso a infraestructuras o servicios, tenencia de electricidad, etc. Pero la forma más común de caracterizar áreas rurales y urbanas es de manera cuantitativa, teniendo en cuenta el número de habitantes. Cada país define distintos umbrales de población y aplica su propia clasificación. Por ejemplo, la ONU aprobó un grado de urbanización donde las ciudades (zona urbana) son aquellas que tienen más de 50.000 habitantes en una determinada área que

Capítulo 2. Estado del arte

cumple al menos 1.500 habitantes por kilómetro cuadrado.[1]

De acuerdo con el INE en España se consideran municipios rurales aquellos de menos de 30.000 habitantes. Distingue entre municipios rurales de pequeño tamaño(menos de 5.000 habitantes) y aquellos de gran tamaño(entre 5.000 y 25.000 habitantes), mientras que considera urbanos aquellos de más de 30.000.[2]

Según datos del 2020 en España hay una población de 47.450.795 personas la cual está dividida en población rural y urbana. La población rural es de 7.538.929 lo que hace un 15,9% del total. Pero los municipios rurales ocupan un 84% de la superficie de España. Esto crea zonas con una muy baja densidad de población. Además, este fenómeno ha ido agravándose con el tiempo. La población española aumentó un 0,6% entre 2011 y 2020. Los municipios urbanos aumentaron su población en un 2,1% pero los municipios rurales han sufrido un descenso del 7,1% en ese mismo periodo de tiempo. [2] Para cualquier modelo de evolución de población es importantísimo tener en cuenta estas diferencias y entender qué factores afectan en estos cambios.

2.1.1. Factores que influyen en la Evolución de la Población Rural y Urbana en España

Para poder estudiar la evolución de la población en rural/urbano primero debemos entender por qué existe esta disparidad demográfica.

La principal característica que marca las diferencias entre las poblaciones es la distribución de la pirámide poblacional envejecida de las áreas rurales. Esto influye enormemente en los patrones de natalidad y mortalidad. Las zonas rurales en España presentan tasas de natalidad más bajas y una población más envejecida lo que contribuye al descenso de la población. Por otro lado, las áreas urbanas en España también presentan un envejecimiento poblacional, pero tienen tasas de crecimiento más sostenidas debido a una gran afluencia de población migrante. [3]

La migración interna y el éxodo rural es una de las principales razones por las que vemos esta disparidad en la población rural/urbana. Desde mediados del siglo XX España ha presentado un desplazamiento masivo de población rural a urbana, impulsado por la industrialización. En las zonas urbanas la diversificación de la economía y el crecimiento de sectores como el turístico, servicios o tecnológico ha atraído a una población joven y activa. Por otro lado, las áreas rurales dependen de sectores más clásicos como agricultura y ganadería que no siempre ofrecen las mismas oportunidades que otro tipo de sectores. Este desequilibrio económico ha fomentado un movimiento migratorio que ha contribuido a la despoblación de muchas zonas rurales, en especial en el interior de la península. [3]

Pero esta migración no solo se debe a factores económicos, sino que también influyen otros factores como es el acceso a infraestructuras y servicios. En áreas rurales, el acceso a servicios de salud, educación o transporte (entre otros) suele ser más costoso de mantener debido a la dispersión de la población que unido a una falta de inversión pública deja unos servicios deficientes. Esto no solo

dificulta la vida cotidiana de los habitantes rurales, sino que también reduce la capacidad de atraer a nuevos residentes, especialmente jóvenes profesionales y familias.[4]

En contraste, las zonas urbanas suelen contar con una estructura más eficiente y consolidada. Estas suelen ofrecer una mayor variedad de opciones educativas, desde escuelas hasta universidades, y un acceso más amplio a servicios de salud especializados. Estos aspectos resultan atractivos para quienes buscan oportunidades de desarrollo personal y profesional, lo que aumenta aún más la migración hacia áreas urbanas. [4]

2.2. Modelos matemáticos

La simulación de la evolución de la población se apoya en modelos matemáticos que permiten comprender y proyectar los cambios demográficos, teniendo en cuenta variables como la natalidad, mortalidad, migración, estructura de edad, etc. A continuación veremos una serie de modelos matemáticos comunes en simulación de evolución de poblaciones para posteriormente elegir un modelo sobre el que basar el trabajo.

2.2.1. Modelo exponencial de Malthus

El modelo exponencial de Malthus es un modelo matemático de evolución de población más antiguos. Este modelo fue propuesto por Thomas Malthus en su ensayo de 1798 “An Essay on the Principle of Population” [5] en el que plantea que la población tiende a aumentar de manera exponencial si no existen restricciones. Este modelo se basa en la idea de que la tasa de crecimiento de la población es proporcional al tamaño de esta; es decir, a medida que la población aumenta, la tasa de crecimiento también lo hace, generando un crecimiento exponencial. [6]

Este modelo sigue una ecuación diferencial básica

$$\frac{dP}{dt} = rP \quad (2.1)$$

Donde P nos representa el tamaño de población, r la tasa de crecimiento y t el tiempo. Para resolver esta ecuación diferencial se integra obteniendo

$$P(t) = P_0 e^{rt} \quad (2.2)$$

Donde P_0 es el tamaño de la población en el momento inicial (suponemos un t inicial de 0). Este tipo de ecuación nos dará una evolución exponencial de la población a medida que pasa el tiempo llegando a números desorbitados. Obviamente este modelo de población es en situaciones idílicas ya que no considera factores que limitan el crecimiento como la falta de recursos, enfermedades, migración, etc. Por tanto este modelo puede ser usado en estudios demográficos en

poblaciones humanas en periodos específicos de la historia o en escenarios de crecimiento rápido pero no se ajusta a las realidades de las poblaciones reales.

2.2.2. Modelo logístico

Otro modelo es el modelo logístico que es una evolución natural del modelo exponencial de Malthus el cual sí tiene en mente el límite máximo de población que puede darse en una localización.[5]

Este modelo pone en consideración el estado limitado de los recursos y por tanto la tasa de crecimiento se va reduciendo a medida que aumenta la población. El modelo logístico sigue también una ecuación diferencial

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K} \right)$$

Donde P representa el tamaño de la población, r la tasa de crecimiento, K el tamaño máximo de la población que puede ser sostenida y t , el tiempo. Este modelo presenta un crecimiento muy similar a la exponencial cuando el tamaño de la población respecto al máximo es bajo pero cuando la población se va acercando a K el término $\left(1 - \frac{P}{K} \right)$ va acercándose a 0 lo que reduce el crecimiento. La solución a esta ecuación diferencial es

$$P(t) = \frac{K}{\left(1 + \left(\frac{K-P_0}{P_0} \right) e^{-rt} \right)}$$

Donde P_0 es el tamaño de la población en el momento inicial (suponemos un t inicial de 0). Este modelo es más útil para representar poblaciones que se enfrentan a unos recursos límites y es usado especialmente en ecología. Aunque este modelo puede ser más útil en demografía en especial en áreas urbanas sigue sin tener en cuenta muchos factores como suponer que la K es constante o no tener en cuenta aspectos básicos como la tasa de nacimiento y muerte, la migración, etc.

2.3. Modelo de componentes y modelos de cambio de cohortes

El análisis de la evolución de poblaciones se suele realizar con modelos demográficos. Dentro de estos modelos, están los modelos de componentes y los modelos de cambios de cohortes, que aunque abordan la dinámica poblacional desde enfoques distintos, ambos se basan en componentes fundamentales como la natalidad, mortalidad y migración.

En el modelo de componentes, se utiliza la ecuación compensadora, que es la fórmula fundamental del análisis demográfico. Esta ecuación se basa en las

2.3. Modelo de componentes y modelos de cambio de cohortes

componentes principales que determinan el crecimiento de una población: natalidad (nacimientos), mortalidad (defunciones) y migración (inmigración y emigración) [7].

Esta ecuación se expresa generalmente como:

$$P_{t+n} = P_t + N_{t,t+n} - D_{t,t+n} + I_{t+n} - E_{t+n}$$

- P_t es la población en un tiempo t .
- P_{t+n} es la población proyectada en un tiempo futuro $t + n$.
- $N_{t,t+n}$ es el número de nacimientos durante el período de t a $t + n$.
- $D_{t,t+n}$ son las defunciones entre t y $t + n$.
- I_{t+n} es la inmigración acumulada hasta el tiempo $t + n$.
- E_{t+n} es la emigración acumulada hasta el tiempo $t + n$.

Esta ecuación sirve para estimar la población en función de estos componentes, pero no especifica cómo se distribuye la población por edad o sexo.

Uno de los mayores desafíos del modelo de ecuación compensadora, es la disponibilidad de datos fiables sobre nacimientos, defunciones y migración. Aunque el uso de este modelo sea generalizado, la falta de información en las estadísticas demográficas dificultan su implementación, especialmente cuando se aplica a subpoblaciones. Además, se intensifica en el caso de las migraciones, donde es frecuente la ausencia de datos [8].

Volviendo al modelo de componentes, es uno de los enfoques más utilizados en el análisis de la evolución de la población. Utiliza la ecuación compensadora, pero desagrega las estimaciones, pudiendo aplicarse a diferentes subpoblaciones, lo cual es una de sus principales ventajas. Al aplicar esta ecuación a diferentes regiones o cohortes, se puede obtener una visión más detallada de cómo cada uno de los componentes influyen en la evolución de la población en cada subpoblación. [9]

Con este modelo, sigue presente la dificultad mencionada anteriormente de que su precisión depende de la calidad de los datos disponibles y la falta de datos fiables o completos presenta un desafío.

El modelo de cambio de cohortes es un método alternativo y más sencillo para las proyecciones de población [8]. En este caso, el análisis se realiza mediante el seguimiento de grupos de individuos nacidos en un mismo periodo, y aunque también se consideran factores como la natalidad, mortalidad y migración, no se utiliza la ecuación compensadora. Se centra en la comparación de tasas de cambio dentro de las cohortes a lo largo del tiempo, con el objetivo de entender mejor los efectos de los factores mencionado anteriormente.

Este modelo es conocido por su capacidad para identificar patrones y tendencias a nivel de cohortes individuales, lo cual es esencial para analizar poblaciones heterogéneas. Al analizar los cambios en las cohortes a lo largo del tiempo,

los investigadores pueden hacer predicciones más precisas sobre cómo se desarrollará la estructura poblacional en el futuro. Swanson y Tedrow [10] también destacan que el uso de las relaciones de cambio de cohortes presenta ventajas en contextos donde los datos migratorios son escasos o difíciles de obtener.

Este modelo, aunque tiene una formulación sencilla, necesita unos datos de entrada específicos, como la distribución por edades de la población, normalmente obtenidos de censos sucesivos. El hecho de que estos censos se suelen encontrar en intervalos regulares, facilita la aplicación de este modelo para proyecciones a largo plazo. El procedimiento suele realizarse de la siguiente manera: primero, calcular las tasas de campo de cohorte entre los puntos temporales y luego aplicar estas tasas a la población en el segundo tiempo para proyectarla en períodos posteriores. [8]

A pesar de su sencillez, el modelo de cambio de cohorte tiene limitaciones. Una de sus desventajas es que, en contextos de rápido crecimiento o disminución de la población, las proyecciones pueden resultar irrazonablemente altas o bajas, respectivamente [11, 12]. Además, el método no permite incorporar diferentes escenarios basados en supuestos sobre fecundidad, mortalidad o migración futura, lo que lo limita frente a otros métodos más complejos como el método de componentes [8].

2.4. Modelo de simulación basados en agentes

Los modelos de simulación basados en agentes permiten reproducir el comportamiento complejo de fenómenos sociales, como en este caso, la evolución de la población rural y urbana, de una manera simplificada, simulando las acciones e interacciones de individuos dentro de un entorno. Son métodos computacionales que permiten crear, analizar y experimentar con modelos compuestos por agentes que interactúan dentro de un entorno. Se formulan como programas informáticos en los que hay unos agentes a los cuales se les asignan reglas de comportamiento y se les permite interactuar entre sí. [13]. Una de las ventajas del modelado computacional es que obliga a ser preciso ya que debe estar completamente y exactamente especificado para poder ejecutarse. Además, otra de las ventajas de estos modelos es que elimina los problemas del aislamiento y las dificultades éticas asociadas con los experimentos sociales, ya que se pueden experimentar con modelos virtuales de manera repetida, ajustando parámetros y permitiendo variaciones aleatorias. Aunque los resultados serán buenos si el modelo refleja adecuadamente el comportamiento del sistema real, lo que no se puede garantizar siempre.

El origen de la idea de este modelo se remonta a la década de 1940 y en el 1969 surgió uno de los primeros modelos basados en agentes, el modelo de segregación de Schelling [14]. No obstante, no es hasta la década de los 90, que empiezan a generalizarse y expandirse dentro de las ciencias sociales, a diferencia de las ciencias físicas y naturales, que ya utilizaban estos modelos anteriormente [13].

Según Epstein [15] los modelos de simulación basados en agentes permiten un

2.4. Modelo de simulación basados en agentes

enfoque diferenciador y característico de las ciencias sociales, que puede describirse como “generativo”, refiriéndose a la capacidad para crear o generar comportamientos y resultados complejos a partir de reglas simples de los agentes dentro del modelo. Además, Epstein habla sobre cómo este enfoque se diferencia de las ciencias inductivas y deductivas. Argumenta que aunque la ciencia social generativa, es empírica, porque se basa en observación y experimentación y no es inductiva, porque se centra en generar un fenómeno macroscópico a partir de la interacción de agentes en un entorno computacional en vez de basarse directamente en la recolección de datos para hacer generalizaciones. Por otro lado, también lo diferencia de los enfoques deductivos, aunque en este caso hay más relación. Lo diferencia porque no se limita a deducir fenómenos macroscópicos a partir de principios generales, sino que muestran cómo se generan estos fenómenos a partir de las interacciones individuales.

En este tipo de modelos entran en juego dos componentes, los agentes y el entorno. Los agentes virtuales representan a los actores sociales, como personas o comunidades, dentro de un entorno virtual. Estos agentes son los que interactúan entre sí a través de mensajes, simulando cómo se difunden las ideas en una red social, además de actuar en base a lo que aprenden de estos mensajes [13]. Y por otro lado, el entorno, es el mundo virtual en el que los agentes actúan y normalmente representan espacios geográficos. Aunque esta parte del modelo suele ser más sencilla de representar, es muy importante ya que la influencia del entorno en el mundo real es de gran importancia.

La simulación de cómo interactúan los agentes y el entorno puede implantarse de diferentes maneras en un ordenador. Se pueden distinguir tres enfoques distintos. Por un lado están los sistemas de reglas de producción, que son bastante claros y fáciles de interpretar. Utilizan conjuntos de reglas condicionales que determinan el comportamiento de los agentes basándose en su estado actual, aunque no tienen la capacidad de aprender. Por otro lado, están las redes neuronales artificiales que consisten en unidades distribuidas en capas, donde cada unidad procesa señales y genera respuestas. En este caso, sí que tienen la capacidad de aprender, pero son menos comprensibles. Por último, están los lenguajes y entornos de programación. Los lenguajes orientados a objetos son los más flexibles y rápidos, pero a su vez, los más complejos; y los entornos de programación son más fáciles de aprender, pero son más lentos y limitados. [16].

Para concluir, la simulación basada en agentes es una herramienta muy potente en la investigación de ciencias sociales, pero tiene algunos problemas debido a la verificación y validación de sus modelos. La validación empírica es crucial para aceptar o rechazar un modelo basado en agentes y esto implica comparar los resultados de los modelos con datos reales para verificar su correspondencia con el mundo real. Pero este proceso de validación enfrenta retos en la calibración y validación, ya que muchos modelos basados en agentes son complejos y no siempre se pueden analizar fácilmente con técnicas tradicionales. Esto hace que sea difícil asegurar que el modelo refleja con precisión la realidad, lo que limita su uso en algunos casos. [17].

2.5. Modelos de simulación basados en análisis de series temporales

Una serie temporal es una secuencia ordenada de observaciones cada una de las cuales está asociada a un momento de tiempo. Estas series pueden entenderse y modelarse a través de la teoría de procesos estocásticos. Este enfoque ayuda a identificar patrones en los datos y a hacer predicciones sobre su comportamiento futuro. [18]

Según la manera en la que evoluciona una serie temporal hay que tener en cuenta algunos conceptos, como la estacionariedad o no estacionariedad.

Los procesos estacionarios son aquellos cuyo comportamiento no cambia con el tiempo. Es decir, los parámetros estadísticos de la serie, como la media, la varianza y la covarianza, deben ser constantes a lo largo del tiempo [19]. Por lo que tiene las siguientes características clave:

- **Media constante:** La media de la serie temporal no cambia con el tiempo.
- **Varianza constante:** La dispersión de los valores de la serie es constante, es decir, los valores se distribuyen con una amplitud fija a lo largo del tiempo.
- **Covarianza constante:** La relación entre los valores de la serie en diferentes momentos no depende del tiempo absoluto, sino de la distancia temporal entre ellos.

Por el contrario, un proceso no estacionario es aquel cuya media, varianza o covarianza cambian con el tiempo [19]. Las causas más comunes de no estacionariedad son [20]:

- **Tendencia:** cuando la serie tiene un patrón de crecimiento o disminución continuo
- **Estacionalidad:** Cuando la serie muestra patrones que se repiten en intervalos regulares.

Para verificar si una serie temporal es estacionaria o no, es común aplicar pruebas estadísticas. Una de las más utilizadas es la prueba de Dickey-Fuller, que está diseñada para comprobar la presencia de una raíz unitaria en una serie temporal. La raíz unitaria es una característica que indica no estacionariedad en una serie [21].

2.5.1. Modelos ARIMA

A continuación, vamos a describir los diferentes modelos ARIMA (Autoregressive Integrated Moving Average). Esta es una técnica de modelado estadístico ampliamente utilizada para analizar y predecir series temporales, ya sean estacionarias o no. Los modelos ARIMA permiten capturar las dependencias temporales en los datos y hacer predicciones a corto y largo plazo. [18]

El modelo ARIMA (p,d,q) combina tres elementos clave que permiten analizar y predecir series temporales:

2.5. Modelos de simulación basados en análisis de series temporales

- AR: Procesos autorregresivos. Representa la relación lineal entre el valor actual de la serie y sus valores pasados. Se basa en el supuesto de que los valores previos tienen un impacto directo en los valores actuales.
- MA: Procesos de medias móviles. Este componente describe la dependencia del valor actual de la serie con los errores residuales de observaciones pasadas.
- I: Integrado. Indica el número de diferencias aplicadas a la serie para convertirla en estacionaria. Esto significa que eliminamos tendencias no estacionarias para poder aplicar los componentes AR y MA

Los parámetros p , d y q son:

- p : Orden del componente autorregresivo (AR). Representa cuántos valores pasados de la serie se utilizan para predecir el valor actual.
- d : Grado de diferenciación. Indica el número de diferencias necesarias para estabilizar la media de la serie temporal.
- q : Orden del componente de medias móviles (MA). Representa cuántos errores residuales previos se usan en la predicción del valor actual.

Los modelos ARIMA, son útiles tanto para series estacionarias como no estacionarias. Para las series estacionarias se suele aplicar directamente un modelo ARMA (p,q), que es un subconjunto del ARIMA sin la parte "Integrada", ARIMA($p,0,q$). Para series no estacionarias, el modelo ARIMA aplica un preprocesamiento conocido como diferenciación (representado por el parámetro d) para transformar la serie en estacionaria. Una vez que se logra la estacionariedad, se aplican los componentes autorregresivos (AR) y de medias móviles (MA).

Esto ocurre debido a que los modelos ARIMA incluyen casos específicos que reflejan diferentes comportamientos en las series temporales. Estos incluyen el modelo ARIMA(0,0,0) que representa ruido blanco, el ARIMA(1,0,0) que es autorregresivo de primer orden, el ARIMA(0,1,0) conocido como paseo aleatorio, y el ARIMA(0,0,1) que incorpora un componente de media móvil de primer orden. [20] A continuación, se presentan sus características y formulaciones.

1. Modelo ARIMA(0,0,0): Ruido Blanco

El modelo ARIMA(0,0,0) representa un proceso de ruido blanco, donde los valores de la serie son independientes e idénticamente distribuidos). Este modelo no tiene componentes autorregresivos, integrados ni de medias móviles, y se formula como:

$$Y_t = \mu + \epsilon_t$$

Donde:

- μ : es la media constante de la serie.
- ϵ_t : es un término de ruido blanco, un error aleatorio con media cero y varianza constante ($\epsilon_t \sim N(0, \sigma^2)$).

Este modelo se aplica a series que no muestran ninguna estructura temporal discernible y cuya única característica es la aleatoriedad de sus observaciones.

2. Modelo ARIMA(1,0,0): Autorregresivo de orden 1 (AR(1))

El modelo ARIMA(1,0,0) incorpora un término autorregresivo de primer orden. En este caso, el valor actual de la serie depende linealmente de su valor inmediatamente anterior, además de un término de ruido blanco:

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t$$

Donde:

- ϕ_1 : es el coeficiente autorregresivo que mide el impacto del valor pasado Y_{t-1} en Y_t .
- ϵ_t : es un término de ruido blanco.

Este modelo captura la dependencia temporal de las observaciones, común en series con inercia o persistencia en sus valores.

3. Modelo ARIMA(0,1,0): Paseo aleatorio

El modelo ARIMA(0,1,0) conocido como paseo aleatorio, es un proceso no estacionario donde el valor actual de la serie es igual al valor anterior más un término aleatorio:

$$Y_t = Y_{t-1} + \epsilon_t$$

Donde:

- Y_{t-1} : es el valor en el instante anterior.
- ϵ_t : es un término de ruido blanco.

Este modelo no tiene memoria a largo plazo ni tendencia definida, y su varianza crece con el tiempo. Es típico en series que presentan comportamiento acumulativo, como precios de acciones.

4. Modelo ARIMA(0,0,1): Media Móvil de orden 1 (MA(1))

El modelo ARIMA(0,0,1) incluye únicamente un componente de media móvil de primer orden. Esto significa que el valor actual de la serie depende de un término de ruido blanco y de un error aleatorio en el periodo anterior:

$$Y_t = \mu + \epsilon_t + \theta \epsilon_{t-1}$$

Donde:

2.5. Modelos de simulación basados en análisis de series temporales

- μ : es la media constante de la serie.
- ϵ_t : es el término de ruido blanco.
- θ_1 : es el coeficiente de media móvil que mide la influencia del error anterior ϵ_{t-1} en el valor actual.

Este modelo es útil cuando las observaciones muestran dependencia de corto plazo entre errores residuales.

2.5.2. Modelo de caminata aleatoria con deriva

El modelo de caminata aleatoria (MCA) es un ejemplo clásico de serie temporal no estacionaria. Este modelo puede clasificarse en caminata aleatoria con o sin deriva y es un ejemplo de lo que se conoce como proceso de raíz unitaria. Este término describe series temporales que no son estacionarias debido a la presencia de un coeficiente autorregresivo igual a 1.

El modelo de caminata aleatoria sin deriva es equivalente a un modelo ARIMA(0,1,0), también conocido como modelo de paso aleatorio. Aunque este modelo es ampliamente utilizado para representar series temporales no estacionarias, una variante común y útil es la caminata aleatoria con deriva, que incorpora un término adicional para capturar tendencias sistemáticas en los datos.

El modelo de caminata aleatoria con deriva se define de la siguiente manera:

$$Y_t = \delta + Y_{t-1} + u_t$$

Donde:

- Y_t : es el valor de la serie temporal en el periodo t .
- δ : es el parámetro de deriva o intercepto, que introduce una tendencia sistemática hacia el alza ($\delta > 0$) o hacia la baja ($\delta < 0$).
- Y_{t-1} : es el valor de la serie temporal en el periodo anterior ($t - 1$).
- u_t : es un término de error aleatorio, con media cero y varianza constante $u_t \sim WN(0, \sigma^2)$

Al reescribir esta ecuación en términos de la primera diferencia ($\Delta Y_t = Y_t - Y_{t-1}$), obtenemos:

$$\Delta Y_t = \delta + u_t$$

Esto implica que ΔY_t (los cambios en Y_t) son impulsados por el parámetro de deriva y las fluctuaciones aleatorias. Así, Y_t se desvía hacia arriba o hacia abajo dependiendo del signo de δ .

En cuanto a sus propiedades, al igual que la caminata aleatoria sin deriva, el modelo de caminata aleatoria con deriva es no estacionario, ya que su media y

varianza dependen del tiempo (t). Además, la varianza aumenta conforme avanza el tiempo, lo que implica que las observaciones se dispersan cada vez más. Por otro lado, el modelo de caminata aleatoria con deriva puede considerarse un caso particular de un modelo autorregresivo de orden 1 ($AR(1)$), en el que el coeficiente autorregresivo es igual a 1.

2.5.3. Modelo de suavizado exponencial estocástico

El modelo de Suavizado Exponencial Simple (SES) se define como[22] :

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$$

Donde:

- y_t : Valor observado en el tiempo t .
- ℓ_t : Nivel estimado o suavizado en el tiempo t .
- α : Parámetro de suavizado, $0 < \alpha < 1$. Los valores recientes tienen mayor peso si α es cercano a 1, y menor peso si es cercano a 0.

El valor de α se optimiza minimizando el Error Cuadrático Medio (MSE):

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_{t|t-1})^2$$

Sin embargo, el SES no captura tendencias ni patrones estacionales. Por lo tanto, en presencia de una tendencia, el modelo puede generar pronósticos poco confiables.

2.5.4. Modelo de Suavizado Exponencial con Tendencia (Holt)

El modelo de Holt extiende el SES al incluir una componente de tendencia, siendo adecuado para series temporales con tendencia lineal.[22]

- Nivel:

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + T_{t-1})$$

- Tendencia:

$$T_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)T_{t-1}$$

- Pronóstico:

$$\hat{y}_{t+h|t} = \ell_t + hT_t$$

Donde:

- T_t : Componente de tendencia en el tiempo t .
- α : Parámetro de suavizado para la tendencia, $0 < \alpha < 1$.
- β : Parámetro de suavizado para la tendencia, $0 < \beta < 1$.

2.5. Modelos de simulación basados en análisis de series temporales

- $\hat{y}_{t+h|t}$: Pronóstico para el tiempo $t + h$ basado en la información disponible en t .

El modelo de suavizado con tendencia nos permite ajustar el valor de importancia tanto del nivel como de la tendencia. De la misma manera que en modelo de suavizado exponencial se puede averiguar α y β usando el Error Cuadrático Medio(MSE).

Capítulo 3

Desarrollo del modelo

El modelo matemático seleccionado será una combinación del modelo de componentes y cohortes, aplicando distintos procesos basados principalmente en el análisis de series temporales con componentes estocásticos para simular los cambios en las diversas variables.

Esta elección no solo responde a que este es el modelo más utilizado para la proyección demográfica, sino también a que, en el caso de España, el acceso a datos históricos proporcionados por el INE hace que el uso de series temporales sea una opción muy adecuada. Además, a pesar de la investigación teórica realizada, otros modelos que analizan los factores que influyen en las distintas variables requieren un mayor nivel de especialización en temas demográficos.

En cuanto a la división rural/urbana seguiremos la clasificación que usa España, entendiendo como rurales aquellas zonas con población menor a 30.000 personas.

El modelo de población seguirá el modelo de componentes actualizando la población urbana y rural año tras año. De esta manera nuestro modelo se basará en la siguiente fórmula:

$$P_{t+1} = P_t + N_t - D_t + I_t - E_t \pm RaU_{t,t+1}$$

- P_t es la población en un tiempo t .
- P_{t+1} es la población proyectada en el año $t + 1$.
- N_t es el número de nacimientos durante el año t .
- D_t son las defunciones en el año t .
- I_t es la inmigración acumulada en el año t .
- E_t es la emigración acumulada en el año t .
- RaU_t es el saldo migratorio del cambio de población rural a urbana en el año t . Para el calculo de la población urbana esta variable se sumará mientras que se restará a la población rural.

Capítulo 3. Desarrollo del modelo

Sin embargo, como se mencionó previamente, este tipo de modelo no tiene en cuenta la distribución por edad de la población. Para desarrollar un modelo que especifique la diferencia entre población rural y urbana, es crucial considerar esta variabilidad, ya que la población rural presenta un marcado envejecimiento en comparación con la urbana. Este factor tiene un impacto significativo en la simulación de la población. Para abordar esto, se crearon cohortes de edad de 5 años, las cuales se actualizan anualmente. De este modo, nuestra ecuación basada en componentes se ajusta según la cohorte en cuestión. Así, cada componente de la ecuación tendrá un valor específico dependiendo de la cohorte que se esté analizando.

Si asignamos valores a cada cohorte siguiendo esta tabla:

n	Cohortes de edad
1	0 - 4
2	5 - 9
3	10 - 14
4	15 - 19
5	20 - 24
6	25 - 29
7	30 - 34
8	35 - 39
9	40 - 44
10	45 - 49
11	50 - 54
12	55 - 59
13	60 - 64
14	65 - 69
15	70 - 74
16	75 - 79
17	80 - 84
18	85+

Cuadro 3.1: Número asignado a cohortes de edad en intervalos de 5 años.

Con estos cohortes y usando la ecuación que hemos visto antes distinguimos entre tres casos para cada año de simulación:

Si $n = 1$ (primer cohorte de edad):

$$P_{1,t+1} = P_{1,t} \cdot 0,8 + N_t - D_{1,t} + I_{1,t} - E_{1,t} \pm RaU_{1,t}$$

Si $2 \leq n \leq 17$ (siguientes cohortes de edad sin tener en cuenta el último):

$$P_{n,t+1} = P_{n,t} \cdot 0,8 + P_{n-1,t} \cdot 0,2 - D_{n,t} + I_{n,t} - E_{n,t} \pm RaU_{n,(t)}$$

Si $n = 18$ (último cohorte de edad):

3.1. Proyección de los nacimientos

$$P_{18,t+1} = P_{18,t} + P_{17,t} \cdot 0,2 - D_{18,t} + I_t - E_t \pm RaU_t$$

- $P_{n,t}$: Población del cohorte de edad n en el tiempo t .
- $P_{n,t+1}$: Población proyectada del cohorte de edad n en el tiempo $t + 1$.
- N_t : Número de nacimientos en el año t (solo relevante para el primer cohorte, $n = 1$).
- $D_{n,t}$: Defunciones del cohorte de edad n en el año t .
- $I_{n,t}$: Inmigración acumulada en el cohorte de edad n para el tiempo t .
- $E_{n,t}$: Emigración acumulada del cohorte de edad n para el tiempo t .
- $RaU_{n,t}$: Saldo migratorio rural-urbano para el cohorte de edad n en el año t .

La base del sistema de cohortes es que la población de un cohorte va aumentando en edad y por tanto pasando a otro cohorte a medida que pasan los años. Para eso suponemos que la población de cada cohorte de edad se divide de manera uniforme en los años. De esta manera, un quinto de la población de cada cohorte pasa al siguiente cohorte cada año, al haber aumentado su edad.

El primer cohorte y el último son diferentes ya que el primer cohorte recibe la totalidad de los nacimientos de cada año (las personas no nacen con edad) mientras que el último cohorte no pierde un quinto de su población a ningún otro cohorte ya que todas las edades mayores de 85 se encuentran en ese cohorte.

Por último para el cálculo de la población total al cabo de un año realizamos el sumatorio de las edades de cada población:

$$P_{t+1} = \sum_{n=1}^{18} P_{n,t+1}$$

Estos procesos se realizan tanto para la población rural como para la población urbana, es decir, cada año obtenemos dos cohortes de población y dos resultados de población, la rural y la urbana. Nuestro modelo quiere simular la diferencia de crecimiento entre estas poblaciones pero sirve también para simular la población de España, simplemente sumando los dos valores.

3.1. Proyección de los nacimientos

En nuestro modelo no vamos a diferenciar tasas de natalidad entre rural y urbano ya que según la investigación realizada, la descompensación de la tasa de natalidad entre estas zonas es solo efecto de la estructura de edades de la población, la cual ya tenemos en cuenta en el apartado anterior.[23]

Para la proyección de nacimientos se distinguen indicadores de natalidad, aquellos referentes a los nacimientos ocurridos en una población e indicadores de

Capítulo 3. Desarrollo del modelo

fecundidad, referentes a los nacimientos en relación con la población en la que tienen lugar, es decir, las mujeres en edad fértil. [24]

Como nuestro modelo tiene en cuenta la edad de las poblaciones el cálculo de natalidad en un año es:

$$N_t = \sum_{n=1}^{18} (TGF_n * M_{n,t})$$

- N_t es el número de nacimientos proyectados en el año t .
- TGF_n es la tasa global de fecundidad de la cohorte n .
- $M_{n,t}$ es la población de mujeres en la cohorte n en el año t .

Es decir, el número de nacimientos será el sumatorio de la multiplicación de la tasa global de fecundidad por cohorte de edad, multiplicado por el número de mujeres en cada cohorte de población.

Para simular la evolución de la TGF se ha decidido usar un modelo de deriva con caminata aleatoria con deriva para la simulación a partir de un análisis de datos históricos. Se ha tomado esta decisión ya que gracias al INE se poseen los datos de evolución histórica sobre la tasa de fecundidad y podemos observar una clara tendencia de los datos. Usando los datos del INE desde 2002, véase el Cuadro 3.2:

Año	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54
2023	0.06	4.31	20.19	43.48	75.49	60.85	16.85	1.60	0.60
2022	0.06	4.61	20.45	44.72	78.32	63.42	17.27	1.74	0.67
2021	0.05	4.61	20.40	45.67	81.97	63.97	17.10	1.57	0.59
2020	0.05	5.36	22.76	47.69	79.87	61.75	16.30	1.43	0.47
2019	0.07	5.99	24.08	49.75	83.61	63.73	16.64	1.49	0.50
2018	0.06	6.26	24.24	51.69	86.41	64.23	16.93	1.49	0.50
2017	0.10	7.07	25.26	54.68	90.42	65.73	16.79	1.42	0.44
2016	0.10	7.36	25.71	57.01	92.96	66.33	16.46	1.27	0.37
2015	0.09	7.74	26.17	57.41	93.15	65.25	15.91	1.10	0.33
2014	0.12	8.09	26.72	57.80	92.04	63.59	15.19	1.05	0.34
2013	0.13	8.38	27.07	55.82	88.20	60.94	14.43	0.91	0.24
2012	0.13	8.99	29.33	57.89	91.65	62.69	14.05	0.90	0.37
2011	0.14	9.51	30.81	59.51	93.53	62.96	13.08	0.87	0.31
2010	0.12	10.61	33.85	61.68	94.48	61.73	12.30	0.87	0.33
2009	0.12	11.83	34.90	62.95	94.84	60.37	11.61	0.86	0.32
2008	0.17	13.26	37.91	66.99	99.66	60.57	11.33	0.73	0.31
2007	0.14	12.82	34.68	63.95	97.07	57.07	10.33	0.72	0.30
2006	0.16	12.01	32.56	64.37	99.29	55.40	9.88	0.66	0.29
2005	0.15	11.35	30.59	63.03	99.29	52.98	9.29	0.48	0.28
2004	0.14	10.75	30.39	63.72	99.06	51.16	8.73	0.43	0.27
2003	0.16	10.55	28.79	64.32	98.27	49.44	8.23	0.42	0.26
2002	0.12	9.86	27.70	63.57	95.51	46.45	7.56	0.39	0.25

Cuadro 3.2: Elaboración propia a partir de valores de tasa de fecundidad del INE

3.2. Proyección de defunciones

Por tanto y aplicando el modelo de caminata con deriva aleatoria cada tasa de fecundidad para el año t se calcula como:

$$f_{i,t} = f_{i,t-1} + \mu_i + \epsilon_{i,t}$$

Donde:

- $f_{i,t}$ es la tasa de fecundidad del año t para la cohorte i ,
- μ_i es la deriva para la cohorte i . La deriva (μ_i) es el promedio de las diferencias interanuales para cada cohorte, es decir la media de la tendencia de la tasa de fecundidad. Para calcularla:

$$\mu_i = \frac{1}{T-1} \sum_{t=2002}^{2023} \Delta f_{i,t}$$

Donde T es el número total de años (en este caso, desde 2002 hasta 2023) y $\Delta f_{i,t}$ son las diferencias interanuales de tasa de fecundidad en los datos.

- $\epsilon_{i,t}$ es el ruido estocástico para el año t , el cual sigue una distribución normal con media 0 y desviación estándar σ_i , es decir:

$$\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_i)$$

Donde calculamos σ_i como:

$$\sigma_i = \sqrt{\frac{1}{T-1} \sum_{t=2002}^{2023} (\Delta f_{i,t} - \mu_i)^2}$$

Este ruido estocástico simulará las fluctuaciones en torno a la tendencia calculada.

El nuevo valor de fecundidad se calculará para cada cohorte de edad. Es importante tener en cuenta que el valor de la tasa de fecundidad no puede ser negativo, por lo que realizamos un último ajuste:

$$f_{i,t} = \max(f_{i,t}, 0)$$

3.2. Proyección de defunciones

De la misma manera que con la tasa de natalidad, en nuestro modelo no vamos a diferenciar tasas de mortalidad entre rural y urbano, ya que, según la investigación realizada la descompensación del porcentaje de defunciones respecto a la población entre estas zonas es solo efecto de la estructura de edades de la población, la cual ya tenemos en cuenta en el apartado de población.[23]

Para la proyección de defunciones se usan dos indicadores principalmente: indicadores de mortalidad que hacen referencia a las defunciones ocurridas en una

Capítulo 3. Desarrollo del modelo

población, e indicadores de mortalidad infantil que son aquellas defunciones de personas de menos de un año.[24] Nuestro modelo de población por cohortes hace que los indicadores de mortalidad y tasas de mortalidad vayan por grupo de población, incluyendo por tanto la tasa de mortalidad infantil en nuestra primera cohorte. De esta manera y como vemos en la próxima tabla 3.2, la mortalidad es alta en el primer grupo debido a esta mortalidad infantil, que suele ser superior. En concreto en España la tasa de mortalidad infantil (menores de un año) es de 2,586 (por 1000). Esta tasa es superior a todas las edades hasta llegar a los 53 años (tasa de 2,612 por 1000). A medida que avanzamos en los grupos de edad, observamos una disminución significativa de la tasa de mortalidad, ya que los individuos que superan la etapa neonatal y la infancia temprana suelen presentar menores riesgos de fallecimiento.

Sin embargo, esta tendencia a la baja se mantiene solo hasta llegar a los grupos de edad avanzada, donde la mortalidad comienza a incrementarse nuevamente debido al envejecimiento natural y a la mayor prevalencia de enfermedades crónicas asociadas a la edad. Este fenómeno es visible en nuestra tabla, donde se evidencia cómo las tasas de mortalidad infantil influyen notablemente en el primer grupo de cohortes, mientras que en los grupos intermedios las tasas son considerablemente más bajas antes de repuntar en las edades más avanzadas.

De esta manera, aun sin variar las tasas de mortalidad, el número de muertos como porcentaje de la población será mayor en nuestra población rural debido a su estructura de edad envejecida.

A continuación muestro la tabla de tasa de mortalidad por cohortes para el año 2023, véase el Cuadro 3.3. En el Anexo A tenemos las tablas con datos desde 2002, los cuales vamos a usar.

3.2. Proyección de defunciones

Cohortes de Edad	Tasa de Mortalidad
0-4	0,6384
5-9	0,0744
10-14	0,0724
15-19	0,1628
20-24	0,2668
25-29	0,3008
30-34	0,3960
35-39	0,5409
40-44	0,8357
45-49	1,3313
50-54	2,3426
55-59	4,0824
60-64	6,4954
65-69	9,9054
70-74	14,9650
75-79	25,2209
80-84	45,3566
85+	137,874

Cuadro 3.3: Tasa de mortalidad por 1000 habitantes según grupo de edad en años. Elaboración propia a partir de datos del INE)

La tasa de mortalidad, al igual que la tasa de natalidad, cuenta con una extensa serie temporal en la que se puede identificar una tendencia clara (ver Tabla Anexo 1). Sin embargo, a diferencia del enfoque utilizado para la tasa de natalidad, he optado por no emplear un modelo de caminata aleatoria con deriva, ya que este se centra exclusivamente en la evolución constante de la tasa bajo el supuesto de estabilidad en los patrones de mortalidad. Dado que eventos imprevistos, como la pandemia de COVID-19 en 2020, y avances médicos significativos pueden provocar variaciones notables en la mortalidad, considero fundamental utilizar un modelo más flexible y adaptativo.

En este contexto, he seleccionado un modelo de suavizado exponencial con tendencia y componente estocástico. Como se analizó en el estado del arte, este enfoque incorpora el valor observado en el año anterior, la tendencia estimada y un componente estocástico que captura la incertidumbre. Además, al actualizarse anualmente, el modelo se ajusta dinámicamente a los cambios en las variables. Así, la ecuación que utilizaremos para simular la tasa de mortalidad en el año t es:

$$y_t = \alpha y_{t-1} + (1 - \alpha)(y_{t-1} + T_t) + \epsilon_t$$

Donde:

- y_t : valor esperado de la tasa de mortalidad para el año t ,
- y_{t-1} : valor real de la tasa de mortalidad para el año anterior $t - 1$,

Capítulo 3. Desarrollo del modelo

- α : Parámetro de suavizado para la tendencia, $0 < \alpha < 1$.
- T_t : componente de tendencia estimado en el año t ,
- $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$: error aleatorio con media cero y varianza σ^2 calculada a partir de la varianza de los datos de tasa de mortalidad desde el año 2002.

La tendencia seguirá la siguiente ecuación:

$$T_t = \beta(y_{t-1} - y_{t-2}) + (1 - \beta)T_{t-1}$$

Donde:

- $y_{t-1} - y_{t-2}$: diferencia entre los valores observados en los años anteriores, indicando el cambio en la serie temporal.
- β : parámetro de suavizado para la tendencia, $0 < \beta < 1$.

Para un valor alto de α (cercano a 1) tendríamos en cuenta más los resultados recientes y en menor medida la tendencia general. De la misma manera, un valor cercano a 1 de β nos indicará una mayor repercusión de la diferencia anual que la tendencia general. Para no decidir arbitrariamente usaremos la técnica de minimización del MSE usando como datos la tabla.

Como hemos comentado anteriormente, el trabajo se basa en la simulación por cohortes; de esta manera, cada cohorte tendrá su propia ecuación con unos valores α y β distintos dependiendo de la evolución de la tasa de mortalidad en cada cohorte.

3.3. Proyección inmigración y emigración exterior

Para proyectar la inmigración y la emigración exterior, analizamos los datos históricos del saldo migratorio exterior en España. Tal y como se observa en la tabla del Anexo A, no identificamos una tendencia clara, a diferencia de los casos de fecundidad y mortalidad previamente analizados. Además, se evidencia una marcada diferencia entre los años posteriores a la crisis económica y los datos más recientes, lo que refuerza la necesidad de un enfoque estocástico para las proyecciones.

Para simular entonces utilizamos un modelo basado en series temporales, específicamente el modelo ARIMA. Cada cohorte de edad se declarará como su serie temporal que se descompone en componentes de tendencia, estacionalidad (si fuera aplicable) y ruido aleatorio, ajustándose a los datos históricos.

En cuanto a la elección de qué modelo ARIMA escoger tenemos herramientas que seleccionan automáticamente el mejor modelo ARIMA para cada cohorte. Estos métodos optimizan los (p,d,q) que corresponden a los componentes autorregresivos, de diferenciación y de media móvil, respectivamente, minimizando el error cuadrático medio (MSE).

3.3. Proyección inmigración y emigración exterior

Una vez ajustado el modelo para cada cohorte, tendremos un modelo determinista de la proyección cuyo resultado será Y_t . Para representar la variabilidad e incertidumbre de la vida real introduciremos componentes estocásticos de manera que nos queda la ecuación:

$$S_{t+1} = Y_{t+1} + \varepsilon_{t+1}$$

Donde ε_{t+1} es un término aleatorio que se calcula a partir del intervalo de confianza de las predicciones deterministas. Estas tendrán un intervalo de confianza y restando el límite superior menos el inferior y aplicándole una multiplicación de un número aleatorio entre -0.5 y 0.5 nos dará que el error:

$$\varepsilon_{t+1} = (IC_{\text{sup}} - IC_{\text{inf}}) \cdot U$$

De esta manera cada cohorte de edad seguirá su modelo ARIMA calculado automáticamente y tendrá un factor estocástico dependiente de su propia evolución temporal.

Aunque ha sido calculado posteriormente, los modelos ARIMA para cada cohorte son, véase el Cuadro 3.4:

Cohorte de Edad	Modelo ARIMA
0-4	(0,1,0)
5-9	(0,1,0)
10-14	(0,1,0)
15-19	(0,1,0)
20-24	(0,1,0)
25-29	(0,1,0)
30-34	(0,1,0)
35-39	(0,1,0)
40-44	(0,1,0)
45-49	(0,1,0)
50-54	(0,1,0)
55-59	(0,1,0)
60-64	(0,1,0)
65-69	(2,1,0)
70-74	(0,1,1)
75-79	(0,0,0)
80-84	(0,0,0)
85+	(0,1,0)

Cuadro 3.4: Modelos ARIMA óptimos para cada cohorte de edad

Como podemos observar, la mayoría de los modelos óptimos corresponden al modelo de paseo aleatorio o ruido blanco, los cuales fueron explicados previamente en el estado del arte. Este modelo es útil cuando no se observan patrones

Capítulo 3. Desarrollo del modelo

evidentes de tendencia o estacionalidad en los datos, ya que asume que las variaciones en la serie son completamente aleatorias. Sin embargo, se identifican dos modelos que no han sido descritos explícitamente antes, pero que podrían ofrecer una mejor descripción de las dinámicas de las cohortes migratorias.

El modelo ARIMA(2,1,0) es un modelo autorregresivo de orden 2 con una componente de término aleatorio (ruido blanco). En este modelo, el valor de la serie en el tiempo t depende no solo de su valor inmediatamente anterior (como en un modelo autorregresivo de primer orden), sino también de su valor en el tiempo $t - 2$. Esta estructura permite capturar dependencias a más largo plazo. Específicamente, el modelo ARIMA(2,1,0) sigue la siguiente ecuación:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

Donde ϕ_1 y ϕ_2 son los coeficientes autorregresivos que determinan la influencia de los valores pasados en la predicción futura, y ϵ_t es el término de error aleatorio, que representa las fluctuaciones no explicadas por el modelo.

Por otro lado, el modelo ARIMA(0,1,1) es un modelo de media móvil de orden 1 con una componente de término aleatorio. El componente de media móvil se refiere a la relación entre el valor de la serie y los errores de predicción de periodos anteriores. Aunque este modelo fue mencionado anteriormente en el estado del arte, en este caso se incorpora un término aleatorio adicional, representando las fluctuaciones no explicadas por las variables pasadas en la serie. El modelo ARIMA(0,1,1) se expresa como:

$$Y_t = Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1}$$

Donde θ_1 es el coeficiente del componente de media móvil y ϵ_t es el error aleatorio.

3.4. Proyección de saldo migratorio rural a urbano

En este apartado, no hemos podido aplicar el análisis de series temporales como en casos anteriores. Tal y como se mencionó en el estado del arte, la definición de lo que se considera rural y urbano varía significativamente a lo largo del tiempo y entre regiones, lo que dificulta la obtención de datos históricos uniformes y fiables para España. La fuente más cercana disponible es el registro de migraciones interiores entre municipios según su tamaño, proporcionado por el Instituto Nacional de Estadística (INE). Sin embargo, este conjunto de datos carece de una categoría que contemple específicamente municipios con hasta 30,000 habitantes, lo que imposibilita esta opción.

Debido a esto para la proyección del saldo migratorio rural a urbano se ha decidido usar el modelo usado por naciones unidas para adaptar la evolución del porcentaje de rural/urbano en un país. A partir de este dato podremos observar la diferencia en población y extrapolar el saldo migratorio entre rural y urbano.

3.4. Proyección de saldo migratorio rural a urbano

El modelo se basa[25] en el cálculo de la proporción urbana utilizando URR_t , definida como el cociente entre la población urbana U_t y la población rural R_t .

$$URR_t = \frac{U_t}{R_t}$$

donde obtenemos la población urbana con

$$PU_t = \frac{URR_t}{1 + URR_t}$$

El crecimiento de la relación urbano-rural $rur_{t,n}$ mide cómo evoluciona el URR a lo largo del tiempo. Está definido como:

$$rur_{t,n} = \frac{1}{n} \ln \left(\frac{URR_{t+n}}{URR_t} \right)$$

En los casos sin datos directos, se asume que rur es constante en el intervalo, y se calcula el URR_T mediante:

$$URR_T = URR_t \cdot e^{rur_t(T-t)}$$

Pero usando esto basarían que la diferencia de crecimiento urbano-rural permanece constante en el tiempo. Sin embargo, la evidencia muestra que la diferencia de crecimiento urbano-rural disminuye con el aumento de la proporción urbana. Para reflejar esto, desarrollaron un modelo en 2018 que ajusta la diferencia hacia una norma global ($hrur$) basada en tendencias históricas. En su modelo distinguen entre poblaciones de más de 2 millones y de menos, por lo que en nuestro caso usaremos la norma global para poblaciones de más de 2 millones.

$$hrur_t = 0,032561 - 0,0153693 \cdot PU_t$$

Además realizaron revisiones al método ajustando las fórmulas para predicciones para la transición centrándose en dos aspectos. En primer lugar no se asumió que la convergencia del diferencial proyectado de crecimiento urbano-rural (rur) hacia la norma global fuera lineal. Para lograr una transición más suave implementaron un cálculo recursivo. Además, después de 25 años, el valor de rur no se mantiene constante en el nivel de la norma global, sino que sigue la evolución del valor de $hrur$ correspondiente al nivel proyectado de urbanización. Estas modificaciones les permitieron un modelo más realista y dinámico que refleja de manera precisa las tendencias de urbanización en función del tiempo.

De esta manera tenemos si $t < t_0 + 25$ donde t_0 es el año del último valor observado de la proporción urbana:

$$rur_{t+1} = rur_t + \frac{hrur_t(PU_t) - rur_t}{25 - (t - t_0)}$$

Capítulo 3. Desarrollo del modelo

Después de 25 años ($t > t_0 + 25$), el diferencial proyectado sigue la evolución de la norma global:

$$rur_t = hrur(PU_t),$$

Finalmente, el ratio urbano-rural proyectado (URR_t) se actualiza utilizando el valor de rur :

$$URR_{t+1} = URR_t \cdot e^{rur_t}$$

Pero lo que queremos conocer en nuestro modelo es RaU el cambio migratorio de la población rural a urbana en España. Para esto lo que haremos será calcular la población esperada en P_{t+1} tanto en el aspecto rural como urbano sin tener en cuenta las migraciones de rural a urbano. Después calculamos el porcentaje de población urbana PU_{t+1} y lo comparamos con el porcentaje esperado usando la simulación de naciones unidas. De esta manera calculado el diferencial de porcentajes, se lo podemos aplicar a la población para, basándonos en el porcentaje esperado, obtener el total de personas que migran de rural a urbano. Sea PU_t^* el porcentaje esperado y PU_t la población sin tener en cuenta el movimiento rural-urbano tenemos la siguiente fórmula:

$$RaU_t = PU_t^* \cdot P_t - PU_t$$

Un valor positivo de RaU_t nos indicará un movimiento migratorio desde las áreas rurales a urbanas, mientras que si es negativo nos indicará movimientos desde las zonas urbanas a las rurales.

Capítulo 4

Modelo computacional

Se ha elegido implementar el modelo en R. R es una herramienta muy poderosa y flexible para realizar simulaciones de población debido a su amplia gama de paquetes estadísticos y de modelado, que permiten realizar simulaciones complejas de manera eficiente. Además, R es especialmente útil cuando se trabaja con grandes volúmenes de datos o cuando es necesario realizar análisis estadísticos avanzados.

Partiendo de datos reales de 2023, he implementado un modelo para proyectar la evolución poblacional desde 2024 hasta 2073 (50 años). Este enfoque permite incorporar componentes clave como la fecundidad, mortalidad, migración interna y externa, además de realizar simulaciones estocásticas que capturan la incertidumbre inherente en los datos. El modelo ajusta año a año la población rural y urbana, integrando los resultados de simulaciones previas y generando datos listos para análisis posteriores.

Con esta implementación, no solo he intentado reproducir el modelo de la manera más fiel posible, sino también un código legible y reproducible, que con pequeños cambios puede analizar más situaciones que nuestra situación inicial.

4.1. Paquetes utilizados

forecast[26]

El paquete `forecast` es ampliamente utilizado para la predicción y el análisis de series temporales. En el contexto de simulación, se utiliza para la modelización de las tendencias de población (u otras variables) a través de métodos estadísticos como el suavizado exponencial o modelos ARIMA.

Algunas funciones clave de este paquete que hemos usado en el código son:

- `ets()`: Permite ajustar un modelo de suavizado exponencial (como Holt-Winters) con componentes de tendencia y estacionalidad. Es ideal para predecir series temporales con componentes de tendencia.

Capítulo 4. Modelo computacional

- `auto.arima()`: Selecciona automáticamente el mejor modelo ARIMA (Auto-Regressive Integrated Moving Average) para una serie temporal.
- `forecast()`: Genera predicciones futuras basadas en el modelo ajustado.

purrr(tidyverse)[27]

El paquete `purrr` es una herramienta extremadamente útil para realizar tareas repetitivas y trabajar con estructuras de datos de manera funcional y eficiente. En el contexto del trabajo lo he empleado principalmente para generar múltiples simulaciones, lo cual es uno de sus usos más destacados.

ggplot2[28]

`ggplot2` es un paquete de R diseñado para la creación de gráficos basados en la gramática de gráficos, lo que permite construir visualizaciones de datos. El el trabajo lo hemos usado en el código del Anexo B para crear gráficas de nuestros resultados.

4.2. Población en el primer año

Los datos de población de 2023 los podemos ver en la siguiente tabla:

Cohorte de Edad	Población (2023)
0-4	1.786.406
5-9	2.221.807
10-14	2.529.002
15-19	2.562.806
20-24	2.502.990
25-29	2.573.472
30-34	2.799.479
35-39	3.085.598
40-44	3.739.321
45-49	4.055.364
50-54	3.800.108
55-59	3.583.098
60-64	3.158.134
65-69	2.631.651
70-74	2.271.259
75-79	1.913.012
80-84	1.268.777
85+	1.588.417

Cuadro 4.1: Elaboración propia a partir de datos del INE

Debido a la falta de datos más recientes tenemos que basarnos en datos del 2020 para encontrar el porcentaje de población rural y urbana.[2] Además, he encontrado un problema adicional, no me ha sido posible encontrar datos oficiales desglosados por cohorte de edad según la diferencia rural/urbana. Para

4.3. Proyección tasa fecundidad

poder hacer una estimación de la población he usado este estudio en el que nos da:

Grupo de edad	Población Rural (%)	Población Urbana (%)
Jóvenes (menos de 15 años)	13.9	16.4
Adultos (15 a 65 años)	62.6	66.6
Mayores (más de 65 años)	23.5	17.0

Cuadro 4.2: Distribución porcentual de población rural y urbana por grupo de edad.

Hemos usado estos datos hemos calculado la población en 2023. Para calcular las poblaciones rurales y urbanas desglosadas por cohorte de edad, primero distribuimos la población total de España en zonas rurales y urbanas utilizando los porcentajes predefinidos (15.9% rural y 84.1% urbana). Luego, dividimos estas poblaciones en tres grupos etarios: jóvenes, adultos y mayores, basándonos en los porcentajes específicos de cada grupo para cada tipo de zona (rural y urbana). Posteriormente, usamos las proporciones de cada cohorte de edad dentro de la población total para asignar el número de personas en cada cohorte (jóvenes, adultos, mayores) a las zonas rurales y urbanas de manera proporcional. Nuestros resultados finales son *poblacion_rural* y *poblacion_urbana* que serán la base para empezar las simulaciones. El código completo se encuentra en el Anexo B.1.

Esta es la base sobre la que realizaremos el proceso de simulación final. Cabe destacar que la simulación de los nacimientos, defunciones y migraciones exteriores no depende directamente de la población total, ya que estas se basan en las tasas de fecundidad y mortalidad así como la evolución del saldo migratorio exterior y del porcentaje de población urbana. Por ello, estos cálculos se realizan de forma independiente y se almacenan por separado para su uso posterior.

4.3. Proyección tasa fecundidad

Para la proyección de la tasa de fecundidad hemos seguido el modelo planteado en el anterior apartado. El archivo resultante de esta simulación es *media_simulaciones_fecundidad.RData* que nos dará la simulación de las tasas de fecundidad por cohortes para cada año. He usado la librería *purr* para realizar 1000 simulaciones ya que tenemos en cuenta el proceso estocástico del error. El código se muestra en el Anexo B.2:

4.4. Proyección tasa mortalidad

Para la proyección de la tasa de mortalidad, he seguido el modelo descrito en el apartado anterior. En esta proyección he utilizado la librería *forecast*, ya que nos permite aplicar la función *ets()*, que ajusta un modelo de suavizamiento exponencial basado en el enfoque ETS (Error, Trend, Seasonality). Al especificar el modelo como "AAN", indicamos que el modelo tendrá un error aditivo, una

Capítulo 4. Modelo computacional

tendencia aditiva y no incluirá estacionalidad. A partir de este modelo, podemos obtener los valores óptimos de los parámetros alpha y beta para nuestra simulación.

Además, he empleado la librería `purrr` para realizar 1000 simulaciones, dado que contamos con un componente estocástico en el modelo. Los resultados de estas simulaciones se almacenan en el archivo `media_simulaciones_mortalidad.RData`, que utilizaremos más adelante. Como en los apartados anteriores podemos observar el código en el Anexo B.3.

4.5. Proyección saldo migratorio exterior

Para la proyección del saldo migratorio exterior, utilizamos la librería `forecast`, que nos permite realizar las proyecciones mediante distintos modelos ARIMA. Además, emplearemos la librería `purrr` para llevar a cabo 1000 simulaciones. El resultado será el saldo migratorio exterior desglosado por cohorte de edad y año que guardaremos en `media_simulaciones_migracion_externa.RData` para su posterior uso.

Es importante resaltar la utilidad del paquete `forecast`, ya que no solo permite simular la evolución de la serie temporal, sino que también selecciona automáticamente el modelo ARIMA más adecuado para cada serie, optimizando así la precisión de las proyecciones y simplificando la implementación computacional.

El código se puede observar en el Anexo B.4.

.

4.6. Proyección de la migración rural-urbana

Para la simulación de la evolución del porcentaje rural-urbano seguimos al pie de la letra el modelo matemático de la ONU presentado en el apartado anterior. Guardamos los resultados en `proporcion_urbana.RData` y podemos observar el código en el Anexo B.5.

4.7. Proyección final de población

Para llevar a cabo la proyección de la población, comienzo cargando los archivos necesarios que contienen las simulaciones previas de fecundidad, mortalidad y migración externa, así como las proporciones de la población urbana. A continuación, defino las tasas de fecundidad, mortalidad y saldo migratorio, las cuales se utilizan en el modelo para calcular los efectos sobre la población.

En la simulación, comienzo con las cohortes de edad tanto en la población rural como en la urbana y calculo las proyecciones de nacimientos y defunciones para cada año. El siguiente paso consiste en incorporar la migración externa, que se divide según la proporción de población urbana y rural. Aunque es cierto que

4.7. Proyección final de población

este dato no se ha probado, no hay suficiente información sobre el destino y origen de los movimientos migratorios externos. Por ello, se ha tomado la hipótesis de que se dirigen proporcionalmente a las poblaciones rural y urbana.

Nuestro cálculo de migración interna se encuentra en una situación similar. Para poder hacer una estimación se ha investigado qué porcentaje poblacional es más propenso a moverse en cualquier migración interna y se ha supuesto que cumple las mismas proporciones para la migración rural-urbana.

En cuanto a la evolución de edad de la población a medida que avanza la simulación, actualizo las cohortes de edad, desplazando las cohortes hacia los grupos de edad superiores y añadiendo los nacimientos en el grupo de 0-4 años. Esto refleja cómo se mueve la población a través del tiempo, considerando tanto los flujos migratorios como los cambios naturales en la población.

Finalmente, al finalizar la proyección, guardo los resultados obtenidos. Estos incluyen la población por área (urbana y rural), el crecimiento vegetativo, las migraciones internas y las cohortes de edad actualizadas. Estos datos son los que utilizaré para realizar los análisis y generar las gráficas y tablas necesarias para el siguiente apartado. El código completo de la simulación se puede encontrar en el Anexo B.6.

Capítulo 5

Análisis de resultados

Las simulaciones y modelos predictivos, aunque poderosas, no están exentas de limitaciones. A través de la utilización de herramientas estadísticas y matemáticas, proyectamos tendencias y anticipamos futuros, pero estos dependen de supuestos que pueden cambiar con el tiempo. El análisis de resultados es, en esencia, una interpretación de datos esperando que sigan la evolución propuesta y no una predicción infalible. Un claro ejemplo de esto es la teoría de Malthus, del cual hemos hablado en el estado del arte, quien en su momento predijo que el crecimiento poblacional descontrolado conduciría al colapso de la humanidad debido a la falta de recursos[29]. Sin embargo, su modelo no pudo anticipar los avances tecnológicos, los cambios sociales y las políticas públicas que alteraron esas proyecciones.

Usando esta historia como moraleja voy a proceder a realizar un análisis de los resultados suponiendo su validez, pero teniendo en cuenta siempre que el futuro es impredecible.

Las gráficas que observamos a continuación han sido realizadas en R , el código está íntegro en el Anexo B.7.

5.1. Evolución tasa fecundidad

La evolución de las tasas de fecundidad presenta patrones diferentes según las cohortes de edad. Los datos muestran un desplazamiento significativo desde cohortes más jóvenes hacia las de mayor edad. Esto es coherente con tendencias observadas en muchas sociedades desarrolladas, donde los cambios culturales y económicos han modificado las edades preferidas para la maternidad.

Para la cohorte de 10-14 y 15-19 las tasas de fecundidad son extremadamente bajas a lo largo de todo el período tendiendo a 0. Esto subraya la ausencia de maternidad en adolescentes, lo cual es un indicador positivo de progreso en términos de salud pública y bienestar infantil. Aunque existen ligeros picos aislados en algunos años, estos son mínimos y no tienen un impacto relevante en la evolución general de la fecundidad.

Capítulo 5. Análisis de resultados

Grupo de Edad	2023	2033	2043	2053	2063	2073
10-14	0.06	0.02	0.00	0.02	0.04	0.00
15-19	4.31	0.56	1.54	0.00	0.16	0.00
20-24	20.19	16.69	10.84	14.18	13.56	6.26
25-29	43.48	38.80	33.41	28.51	18.39	9.36
30-34	75.49	61.84	51.59	39.17	41.61	35.19
35-39	60.85	71.92	75.77	85.97	92.58	103.07
40-44	16.85	22.99	28.80	33.64	38.33	41.97
45-49	1.60	2.10	2.97	3.32	3.46	3.94
50-54	0.60	0.76	0.52	0.67	1.31	1.54
Total	24.83	24.07	22.83	22.83	23.27	22.37

Cuadro 5.1: Tabla de fecundidades por grupo de edad y cada 10 años

En el caso de la cohorte de 25-29 años, se observa una clara disminución en las tasas de fecundidad. Este descenso refleja el aplazamiento de la maternidad hacia edades mayores, probablemente debido a factores como la inserción laboral de las mujeres y los cambios en las prioridades de vida.

Por otro lado, la cohorte de 35-39 años muestra un comportamiento opuesto al de las cohortes más jóvenes, con un aumento progresivo de las tasas de fecundidad a lo largo del tiempo. Este grupo se consolida como la edad principal en la que se tienen hijos. El siguiente grupo con mayor tasas de fecundidad pasa a ser el grupo de 40-44 años.

Esta situación puede derivar en problemas. Existe un consenso médico donde considera embarazo de riesgo a toda personas mayor de 35 años. Como vemos en los datos la mayoría de personas se encuentra en este grupo lo que puede derivar en problemas de salud.

Para paliar esto puede haber soluciones. Según una encuesta del INE[30], el 25% de las mujeres no se decide a ser madre por razones económicas, falta de políticas de conciliación o por inestabilidad laboral. Por tanto y viendo estos resultados es importante que se ajusten políticas públicas para que si se desea no haya tantas dificultades económicas para tener hijos.

En conjunto, estos resultados revelan una transición demográfica clara, donde la fecundidad se desplaza de las edades tempranas hacia las más avanzadas, acompañando las transformaciones culturales y económicas de las sociedades modernas. También cabe destacar que hay una leve disminución de la tasa de fecundidad general en edades fértiles, pero no parece ser el foco principal del problema.

En la Figura 5.1 podemos observar un gráfico de la evolución que refleja visualmente el cambio en los cohortes de edad predominantes.

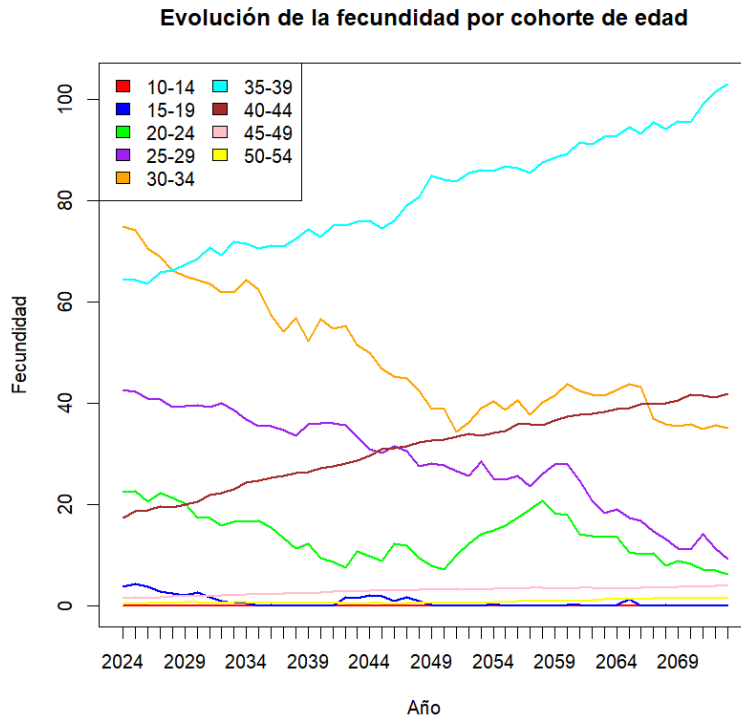


Figura 5.1: Gráfico de la evolución de la tasa de fecundidad

5.2. Tasas mortalidad

En cuanto a la evolución de las tasas de mortalidad, se observa una tendencia a la baja o a mantenerse relativamente estables hasta el año 2040, a partir del cual se registra un incremento en todas las cohortes.

Este comportamiento no era el esperado, dado que los datos históricos mostraban una tendencia a la baja en las tasas de mortalidad. Esto podría explicarse por el aumento registrado en 2020 y 2021 debido al impacto del COVID-19, o bien, podría ser una consecuencia de la influencia de los procesos estocásticos que han elevado las tasas de mortalidad a niveles más altos.

En el gráfico 5.2, representamos las cohortes de edad que presentan tasas de mortalidad significativas, ya que las tasas de los otros cohortes son muy cercanas a cero.

Este aumento proyectado de las tasas de mortalidad, sumado al envejecimiento de la población, podría resultar en un incremento significativo en el número de defunciones en el futuro.

Capítulo 5. Análisis de resultados

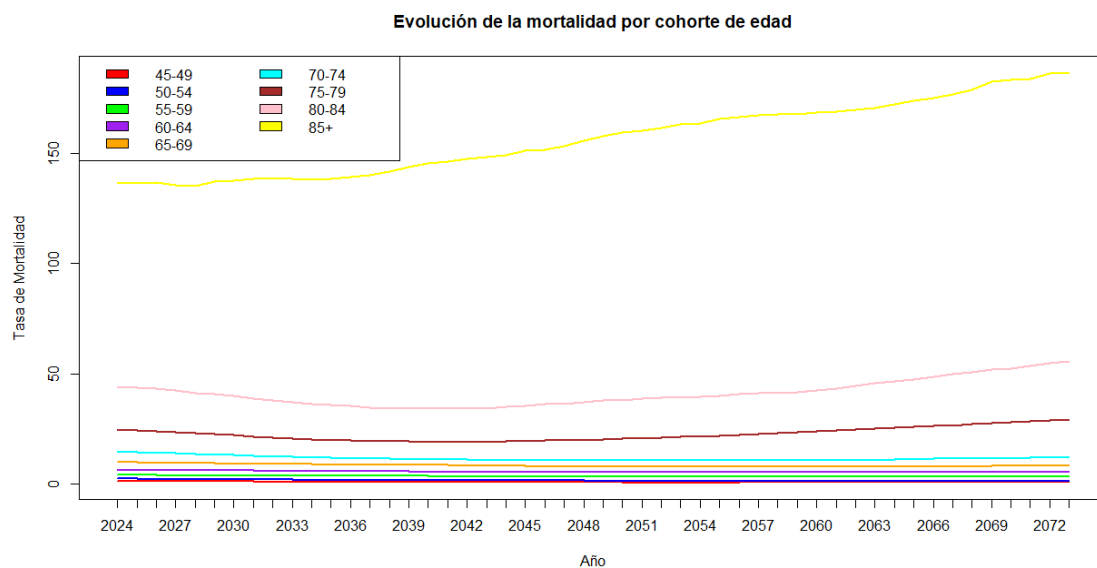


Figura 5.2: Gráfico de la evolución de la tasa de mortalidad

5.3. Análisis crecimiento vegetativo

El crecimiento o saldo vegetativo indica la diferencia entre nacimientos y defunciones en un periodo de tiempo. Como podemos observar en el gráfico 5.3, el crecimiento vegetativo muestra una clara tendencia decreciente a medida que avanza el tiempo. Comenzamos en 2024 con un valor negativo de -77,498 y alcanzamos una cifra aún más baja de -778,588 para 2073. Este patrón de crecimiento vegetativo negativo sugiere que las tasas de natalidad no serán suficientes para compensar las muertes, lo cual es característico de poblaciones envejecidas.

Este decrecimiento tiene una serie de implicaciones para la estructura social y económica de España. En primer lugar, indica que, aunque la población pueda experimentar fluctuaciones debido a la migración, la dinámica natural será de decrecimiento. Además, este fenómeno señala una necesidad urgente de fomentar la natalidad o promover la inmigración de personas jóvenes para contrarrestar el saldo natural de la población que le llevará al envejecimiento.

5.4. Análisis saldo migratorio exterior

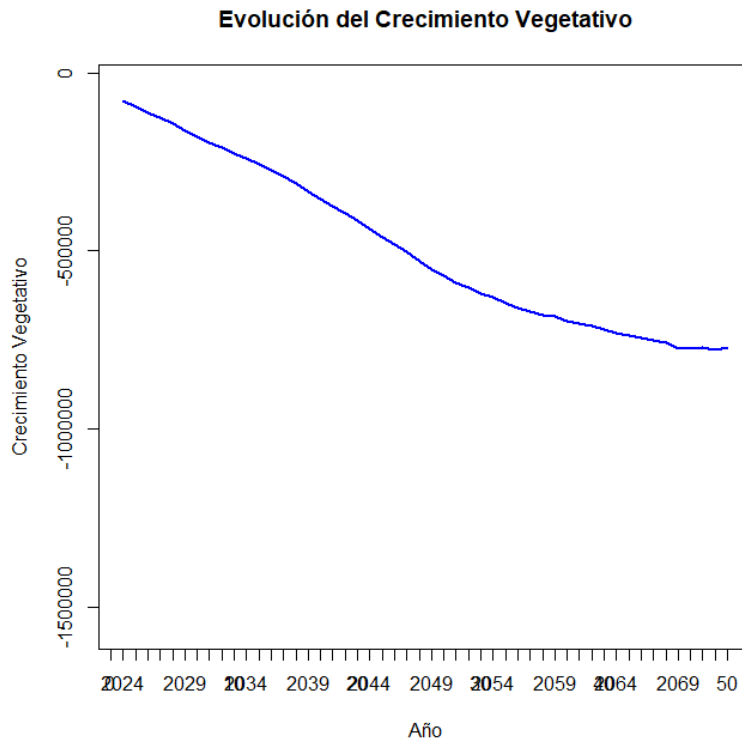


Figura 5.3: Gráfico de la evolución del crecimiento vegetativo

5.4. Análisis saldo migratorio exterior

El saldo migratorio exterior se mantiene más o menos constante a lo largo de los años. Esto se debe a que, en la mayoría de las cohortes, se modelizaron como series temporales ARIMA (0,1,0), lo que ha hecho que fluctúen en torno a unos valores de media.

En cuanto a su análisis, tras observar el crecimiento vegetativo, es positivo este saldo migratorio, que además está concentrado entre las edades de 20 a 34 años, lo que permite añadir personas jóvenes a una población envejecida. En la tabla podemos observar la evolución de los valores.

Capítulo 5. Análisis de resultados

Cohorte de edad	2024	2033	2043	2053	2063	2073
0-4	32.665	29.241	29.913	31.659	25.459	34.299
5-9	37.544	39.579	29.853	33.170	41.971	48.794
10-14	39.794	38.419	42.919	43.824	50.782	45.140
15-19	59.581	60.772	50.863	53.811	64.401	82.170
20-24	87.984	78.566	75.622	96.693	95.330	111.872
25-29	95.993	94.898	103.474	86.059	75.239	108.280
30-34	82.005	73.235	89.340	73.749	77.323	74.535
35-39	64.034	63.665	49.260	42.238	72.399	42.267
40-44	47.951	44.652	36.843	46.635	35.863	71.998
45-49	36.457	39.247	30.581	47.005	23.901	38.318
50-54	24.213	24.454	19.252	27.458	26.236	19.463
55-59	16.607	18.045	15.791	19.516	24.417	14.672
60-64	13.186	13.158	15.229	17.283	9.061	18.849
65-69	6.513	7.744	9.045	9.611	8.221	9.020
70-74	2.856	2.671	2.745	3.002	3.419	3.360
75-79	-161	351	18	-202	47	-260
80-84	-428	-451	-544	-511	-483	-479
85+	-2.285	-2.381	-2.267	-2.041	-2.721	-2.620
Total	644.509	625.865	597.937	628.959	630.865	719.678

Cuadro 5.2: Saldo migratorio exterior por cohorte de edad y año

5.5. Análisis poblacional

Dentro de nuestro análisis distinguiendo entre rural y urbano obtenemos también un modelo que nos muestra la evolución de la población general de España.

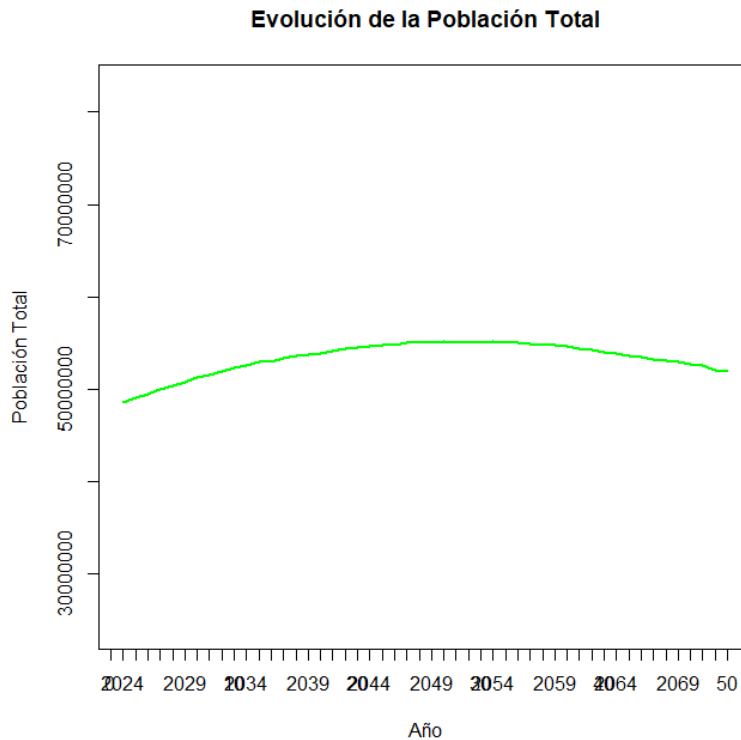


Figura 5.4: Gráfico de la evolución de población España

Como vemos en la gráfica 5.4, la población experimenta un aumento gradual hasta el año 2049, momento en el que comienza a descender por primera vez. A partir de este punto, se observa un período de aproximadamente 10 años de estabilidad, seguido por un descenso marcado en los años posteriores. Este fenómeno se debe a que, en ese momento, la población comienza a envejecer considerablemente y el crecimiento vegetativo negativo se vuelve tan significativo que, incluso, el saldo migratorio exterior no logra sostener la población.

En cuanto a la distribución entre población rural y urbana, al inicio del período, la población urbana representa un 84,1 % del total. Sin embargo, después de 50 años, las proyecciones indican que la población urbana alcanzará un 92,28%. Este cambio drástico genera una clara disparidad, lo que es evidente en las gráficas siguientes, donde la población urbana sigue una tendencia ascendente, mientras que la población rural experimenta una reducción considerable, prácticamente reduciéndose a la mitad en 50 años.

Capítulo 5. Análisis de resultados

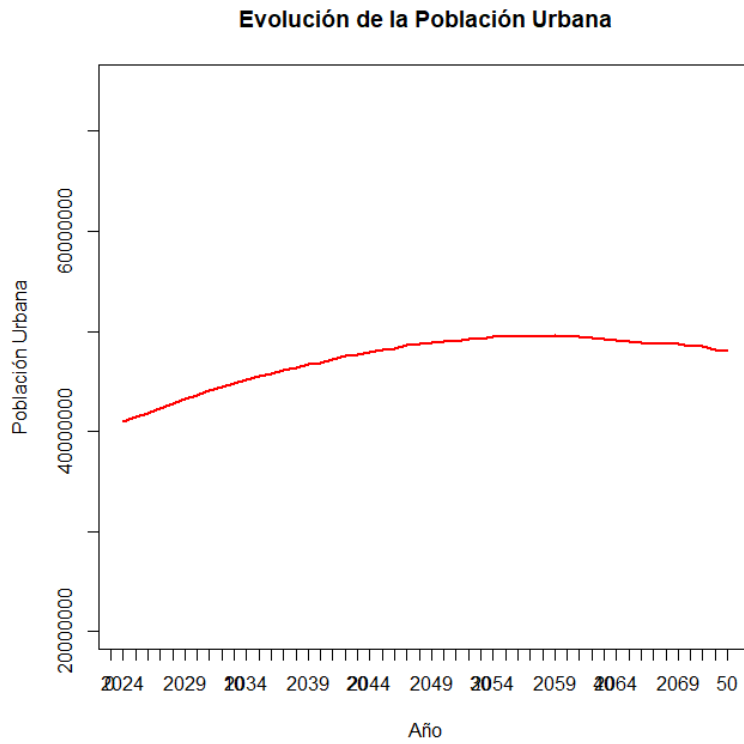


Figura 5.5: Gráfico de la evolución de población urbana

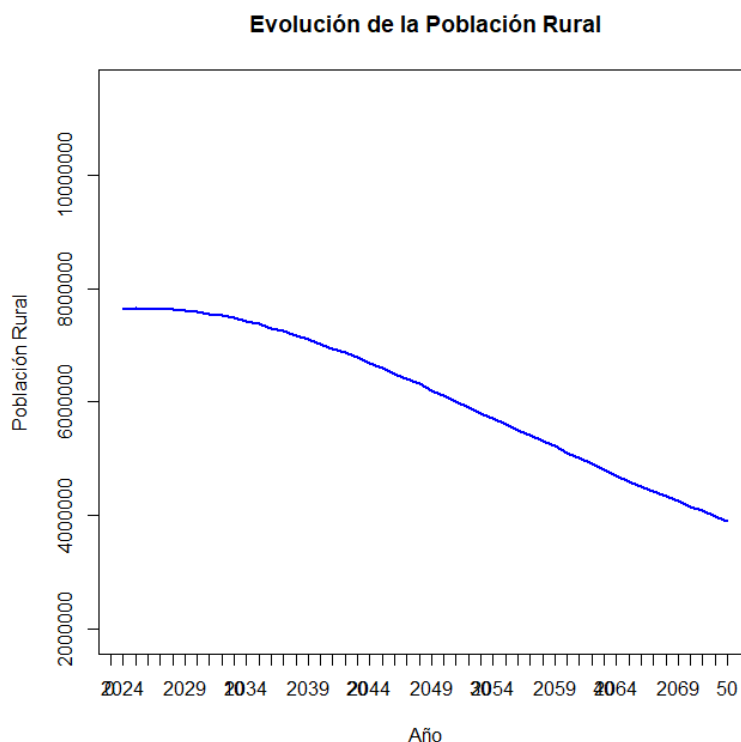


Figura 5.6: Gráfico de la evolución de población rural

Este cambio demográfico puede conllevar una serie de problemas, como la sobrepoblación y el colapso de infraestructuras en las áreas urbanas, mientras que las zonas rurales podrían enfrentar desafíos significativos en términos de desarrollo económico y recursos humanos. Además, la falta de habitantes en áreas rurales puede también resultar en la desaparición de servicios básicos como educación, atención médica y comercios, lo que dificulta la vida en estas zonas.[31]

Otro de los grandes problemas a los que nos enfrentaremos según la simulación es la estructura poblacional. Según nuestro modelo, en 2073 tendríamos esta distribución por cohortes de edad(Figura 5.7). Esta figura es muy relevante porque no solo muestra claramente la gran diferencia en la cantidad de personas entre rural y urbano, sino que muestra un futuro envejecido. Este avance en la edad de la población no ocurre solo para áreas rurales donde la población de mayores de 85 es la más predominante, sino también en áreas urbanas donde la población más predominante es de 65 a 69 años.

Capítulo 5. Análisis de resultados

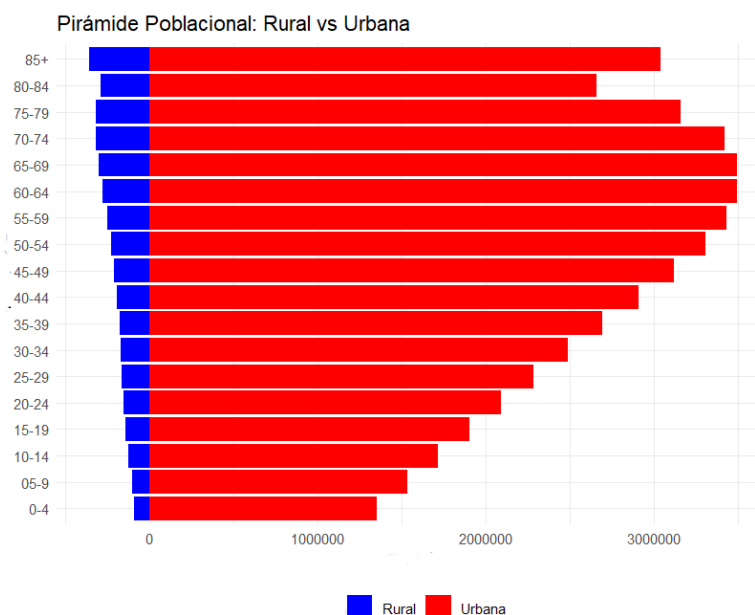


Figura 5.7: Pirámide poblacional Rural/Urbana

Las posibles repercusiones de este tipo de pirámide de población son variadas y de gran importancia, ya que pueden tener un impacto significativo en la estructura de la sociedad.

Para empezar, una de las repercusiones más directas es la presión sobre los sistemas de salud. Las personas mayores suelen tener más enfermedades crónicas, discapacidades y problemas relacionados con el envejecimiento, lo que aumenta significativamente la demanda de servicios de atención médica. Esto puede generar cuellos de botella en los hospitales, escasez de personal sanitario especializado en geriatría y mayores costos en el sistema de salud pública [32].

Otro de los aspectos más afectados es la economía. Las personas mayores tienden a consumir menos bienes y servicios que los jóvenes, lo que reduce la demanda en sectores clave. Además, las personas mayores tienen una mayor propensión a ahorrar y una menor a invertir, lo que puede afectar el crecimiento económico general [33].

Se enfrentan también a un mayor riesgo de exclusión social, ya que la vejez constituye un factor clave de vulnerabilidad. Esto se debe a la combinación de elementos como un nivel educativo bajo, la carencia de redes sociales, la dependencia de escasas ayudas sociales, el sufrimiento de patologías tanto físicas como psicológicas, las condiciones de vivienda desfavorables y la limitada implicación en actividades políticas y sociales [32].

Podríamos seguir analizando problemáticas de una población de este tipo, como los problemas con el sistema de pensiones, los desafíos en la infraestructura accesible para personas mayores, la disminución de la fuerza laboral, el crecimiento de la pobreza, etc. Lo que es claro es que nos dirigimos a este escenario.

5.6. Análisis saldo migratorio rural/urbana

Mi modelo muestra un envejecimiento muy marcado, pero todos los modelos que he investigado muestran esta tendencia hacia una población envejecida, en mayor o menor medida. Por ello, es esencial que se implementen políticas públicas y que se conciencie a la sociedad sobre un problema que, casi con total seguridad, enfrentaremos.

5.6. Análisis saldo migratorio rural/urbana

Para finalizar, analizamos el movimiento migratorio de rural a urbano proyectado.

Como vemos en la gráfica 5.8, el saldo migratorio rural a urbano proyectado muestra una tendencia positiva en los primeros años, alcanzando un pico de 78.121 personas en el año 2047. Sin embargo, a partir de ahí, el saldo migratorio comienza a disminuir de forma gradual hasta estabilizarse en torno a los 36.895 personas en el último año de la proyección.

La disminución progresiva del saldo migratorio refleja cambios en las dinámicas migratorias, influenciados principalmente por el envejecimiento de la población rural, lo que genera una creciente necesidad de repoblación. Aunque estos factores demográficos son determinantes, es importante subrayar que la migración hacia las zonas urbanas se mantiene positiva.

Esto nos indica que la despoblación rural no se debe únicamente al decrecimiento natural de la población, sino también a una migración continua hacia las ciudades, motivada por la búsqueda de mejores oportunidades laborales y condiciones de vida más favorables.

En este contexto, es urgente crear políticas públicas que fomenten el desarrollo en las zonas rurales, no solo mejorando las oportunidades de empleo y condiciones económicas, sino también invirtiendo en la mejora de infraestructuras, servicios básicos y la creación de un entorno positivo para la vida de las personas. El reto no solo es atraer a los jóvenes de nuevo a las zonas rurales, sino también garantizar que puedan acceder a las mismas oportunidades de desarrollo personal y profesional que ofrecen las ciudades.

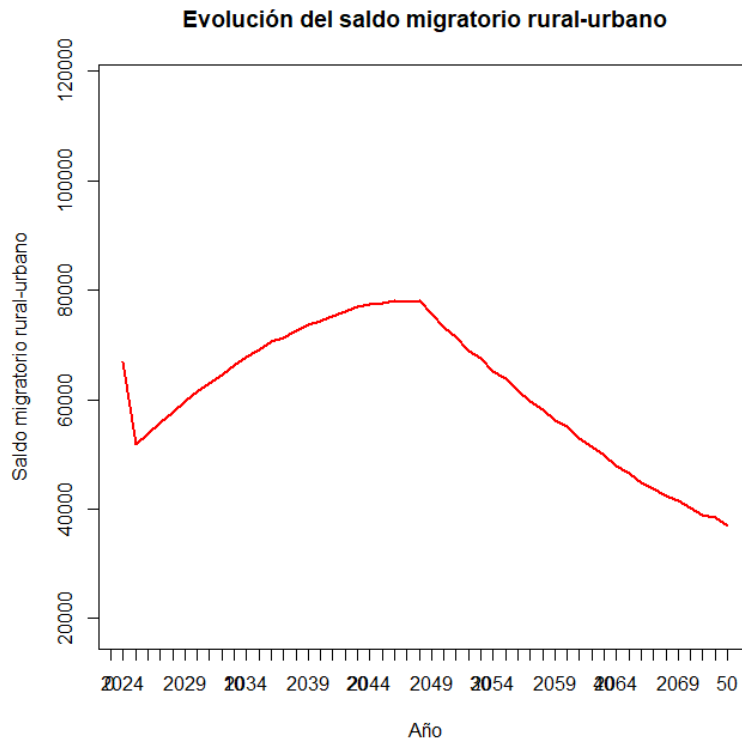


Figura 5.8: Gráfico de la evolución del saldo migratorio de rural a urbano

5.7. Comparación con otros estudios

Desgraciadamente, no ha sido posible encontrar otros estudios que proyecten la evolución con distinción rural-urbana para el concepto elegido en este trabajo. Por lo tanto, para realizar comparaciones, se utilizarán las proyecciones generales de población del INE[34], con el objetivo de analizar las diferencias en los aspectos simulados por ambos modelos. Dado que este estudio ofrece resultados para el año 2074, compararemos con nuestras estimaciones para 2073, siendo conscientes del margen de error.

5.7. Comparación con otros estudios

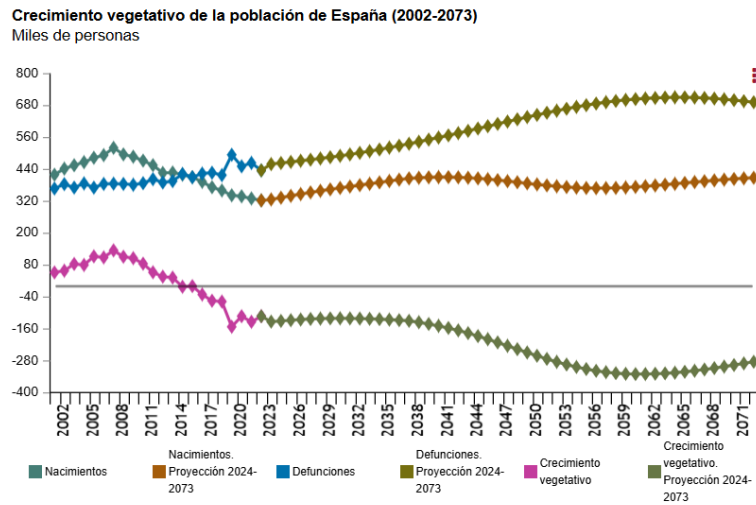


Figura 5.9: Gráfico de la evolución de Saldo Vegetativo del INE

Como puede observarse en la Figura 5.9, el INE también predice un saldo vegetativo negativo con un decremento significativo, que se estabiliza alrededor del año 2060. La mayor diferencia radica en que, mientras el INE anticipa un estancamiento del decremento del saldo vegetativo en los próximos 20 años, nuestro modelo muestra una caída considerable en ese mismo período.

En cuanto al saldo migratorio exterior, el INE proyecta una disminución hasta estabilizarse en valores cercanos a 280.000, mientras que nuestro modelo prevé una estabilidad en valores entre 600.000 y 700.000. Esto refleja que, mientras nuestro modelo asume estabilidad en la migración actual de España, el INE anticipa una disminución significativa en los próximos años.

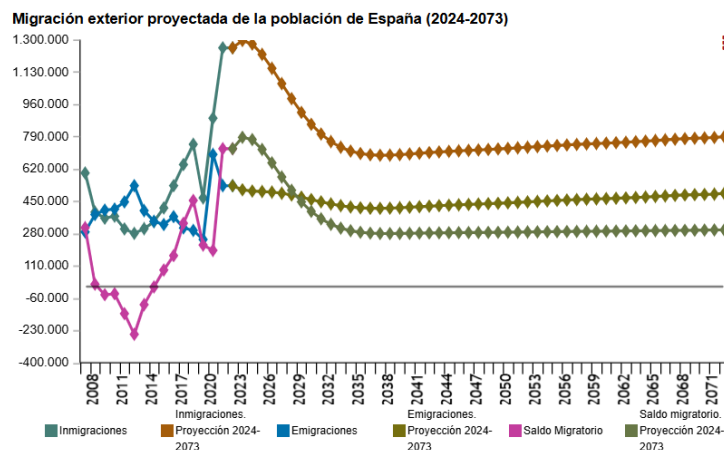


Figura 5.10: Gráfico de la evolución de la migración exterior del INE

En términos de población total, el INE proyecta que, en los próximos 15 años, España ganaría 5.137.447 habitantes, alcanzando 53,7 millones de personas en

Capítulo 5. Análisis de resultados

2039. Para 2074, la población total proyectada sería de 54,6 millones, con un incremento acumulado de 5,98 millones.

Por otro lado, nuestro modelo estima un crecimiento de 5.184.150 habitantes en 15 años, un valor muy cercano al proyectado por el INE. Sin embargo, para 2074, nuestro modelo proyecta 51,9 millones de habitantes, casi 3 millones menos que el INE. Esto sugiere que, mientras nuestro modelo es muy preciso a corto plazo, podría perder fiabilidad en horizontes más lejanos, validando las proyecciones del INE.

El INE analiza además posibles escenarios, como se muestra en la Figura 5.11. Al compararlo con nuestro modelo, el escenario de natalidad baja es el que más se asemeja, dado nuestro saldo vegetativo menor y un saldo migratorio mayor. Esto podría indicar que nuestro modelo subestima los nacimientos y sobreestima las muertes.

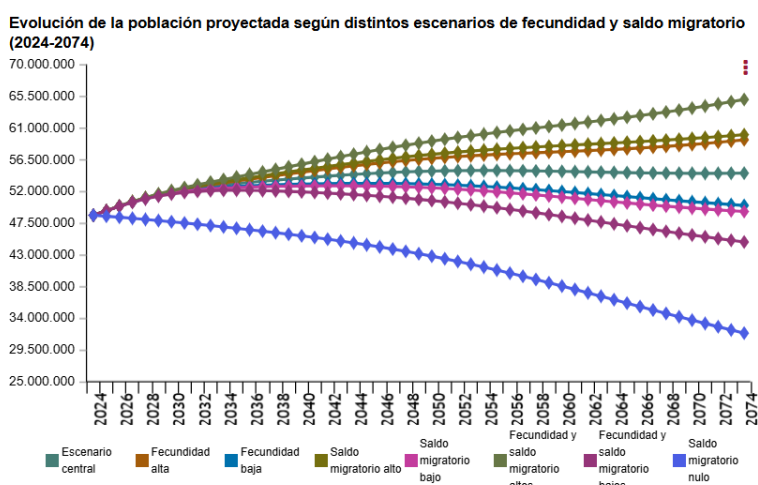


Figura 5.11: Gráfico de la evolución poblacional INE

En cuanto a la estructura por edades, el INE proyecta que la población entre 20 y 64 años representará el 54,2%, mientras que nuestro modelo estima un 53,2%. Para la población mayor de 65 años, el INE calcula un 30,3%, frente al 33,4% de nuestro modelo. Estos resultados son muy similares, aunque nuestro modelo proyecta una población ligeramente más envejecida.

Capítulo 6

Conclusiones e impacto

6.1. Evaluación objetivos del trabajo

En cuanto a los objetivos, creo que se han alcanzado de manera satisfactoria. En primer lugar, se llevó a cabo una investigación detallada sobre los modelos de simulación y proyección poblacional, lo que permitió una comprensión profunda de las metodologías existentes y su aplicabilidad al caso de España. Este paso fue clave para seleccionar las técnicas más adecuadas para el análisis.

Posteriormente, se diseñó un modelo matemático adaptado al contexto español, capaz de simular la evolución de las poblaciones rurales y urbanas, así como las migraciones entre ambas.

La implementación del modelo en R fue correcta, pudiendo realizar simulaciones que representaran escenarios futuros de la evolución poblacional.

Además, se analizaron los resultados obtenidos analizando sus características y consecuencias. También se realizó una comparación con la proyección presentada por el INE. La comparación del modelo no ha sido la más extensa debido a la falta de modelos que usen la misma concepción de rural/urbano.

6.2. Posibles líneas futuras de investigación

Aunque los objetivos iniciales se han logrado, considero que este trabajo puede servir como base para investigaciones futuras que introduzcan mejoras y amplíen su alcance. Entre las posibles líneas de desarrollo cabe destacar:

- **Adaptación a nivel de comunidades autónomas o provincias.** Hemos tratado en el trabajo a España con una división rural/urbana de todo el país pero es cierto que la situación en un municipio pequeño aislado de grandes poblaciones no es la misma que un municipio que tiene relativamente cerca acceso a poblaciones urbanas. Creo que se podría adaptar el trabajo a nivel de comunidades autónomas o incluso provincial.

- **Análisis multivariable de las proyecciones.** Aunque el trabajo se ha centrado en los aspectos básicos de la simulación, sería interesante ampliar el análisis incorporando factores económicos, sociales y de políticas públicas para evaluar cómo estos influyen en los resultados de la simulación.
- **Intervención de expertos demográficos.** En relación con el apartado anterior, aunque se ha realizado una amplia investigación, se decidió seguir modelos matemáticos basados principalmente en resultados y en el análisis de evoluciones temporales que en los factores que los causan. Sin embargo, con la intervención de expertos en demografía se podría reconfigurar el modelo y realizar estudios que tengan más en cuenta estas variables.
- **Análisis de escenarios futuros.** Aunque este no era un objetivo específico del trabajo, me parece muy interesante simular situaciones basadas en diversas hipótesis (por ejemplo, cambios en las políticas públicas, fluctuaciones en la natalidad, una pandemia como el COVID-19, o incluso conflictos bélicos) para evaluar el impacto de distintos escenarios futuros.

6.3. Evaluación personal del proceso

A nivel personal, el trabajo ha tenido un impacto significativo en mi desarrollo. Me he enfrentado a diversas dificultades a lo largo del proyecto. La gestión del tiempo fue uno de los mayores retos, ya que compaginar este proyecto con mi trabajo y otros estudios me ha resultado complicado. Esto derivó en que las fases del proyecto no siguieran los plazos estrictos de la planificación inicial.

Otro de los grandes retos ha sido la falta de consenso sobre la definición de población rural y urbana. Como se menciona en el estado del arte, cada país u organización utiliza parámetros diferentes, e incluso estudios del INE han variado con el tiempo sus criterios. Esto me llevó a invertir un tiempo considerable en encontrar datos adecuados o adaptar la información a los conceptos de población rural y urbana usados en este trabajo. También me ha generado dificultades la propia naturaleza del trabajo, que implica tomar decisiones críticas sobre qué modelo utilizar. Esto me ha generado bastantes momentos de presión e incertidumbre. He dudado si las elecciones realizadas eran las más adecuadas o si debería haber usado otras alternativas, lo que derivó en retrasos y cierta frustración.

Sumándole a esto, tuve un error en la proyección inicial de las tasas de fecundidad y mortalidad, lo que también me obligó a repetir parte del proceso. Además pasé mal unos datos, lo que generó un problema que no me di cuenta hasta que llegué al análisis de los resultados. Todo esto ha hecho que le dedique bastantes más horas de lo esperado a este trabajo.

Aun así creo que el trabajo me ha aportado consecuencias positivas. No solo he obtenido conocimientos sobre trabajos de investigación, factores que influyen en la población, modelos de simulación, conocimiento sobre R, etc sino también competencias como la gestión del tiempo, manejo del estrés o toma de decisiones. Personalmente, en mi futuro, creo que el haber realizado este trabajo me va

a ayudar mucho a nivel personal y profesional.

También quiero destacar el gran apoyo de mi tutor, quien me acompañó durante todo el proceso con reuniones cada dos semanas y mostró una gran disponibilidad para adaptarse a mi complicado horario (incluso reuniéndose en días festivos conmigo). Su orientación fue esencial y, sin ella, no habría sido posible alcanzar los objetivos planteados.

6.4. Impacto

En este trabajo se ha analizado el movimiento de la población entre los medios rural y urbano y las evoluciones de población rural/urbana, fenómenos que afectan a múltiples ámbitos. Al reflexionar sobre el potencial impacto respecto a los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 [35], reflexionamos sobre el impacto social, económico y ambiental del trabajo. Usando los ODS podemos identificar cuáles están más relacionados con los resultados obtenidos.

La migración de la población rural hacia las zonas urbanas puede tener un impacto complejo en la pobreza y por lo tanto estar relacionado con el **ODS 1**, "Fin de la pobreza". Por un lado, algunas personas se trasladan buscando mejores oportunidades de empleo y condiciones de vida, lo que puede contribuir a la reducción de la pobreza si logran acceder a trabajos dignos y servicios básicos. Sin embargo, este fenómeno también puede generar un aumento de la pobreza en las áreas urbanas, especialmente cuando la demanda de empleo supera la oferta, lo que lleva a una mayor competencia por trabajos y servicios limitados. Además, la sobrepoblación urbana puede resultar en la marginación de aquellos que no tienen los recursos o habilidades necesarias, lo que puede perpetuar la pobreza.

Este fenómeno demográfico también puede influir de manera significativa en el **ODS 8**, que busca promover el trabajo decente y crecimiento económico, de una manera muy similar al anterior ODS analizado. La migración de la población rural hacia las zonas urbanas está estrechamente vinculada con la búsqueda de mejores oportunidades laborales. Sin embargo, al haber un incremento de la población en las ciudades, se genera una mayor competencia por los empleos, lo que lleva a dificultades para encontrar trabajo digno y esto limita el crecimiento económico urbano. Además, en las zonas rurales, la pérdida de población activa puede llevar a una disminución en la actividad económica local, afectando aún más las condiciones laborales y de crecimiento económico.

El **ODS 10** de "Reducir la desigualdad en y entre los países" también se ve afectado. La desigualdad entre zonas rurales y urbanas afecta al desarrollo personal y económico de las personas. La migración de personas en busca de mejores oportunidades en las ciudades puede incrementar la brecha entre los que viven en zonas urbanas y los que viven en zonas rurales.

Por último, dentro de los ODS que tienen una relación muy estrecha con el movimiento de la población rural a urbana, destacaría también el **ODS 11**, "Lograr que las ciudades sean más inclusivas, seguras, resilientes y sostenibles". El

Capítulo 6. Conclusiones e impacto

aumento de la población urbana, sin una adecuada planificación, puede generar problemas como la escasez de vivienda y la sobrecarga de servicios básicos. Además, el crecimiento descontrolado de las ciudades puede incrementar la contaminación y afectar la calidad de vida tanto de los habitantes.

El **ODS 3**, "Garantizar una vida sana y promover el bienestar para todos en todas las edades", se ve afectado por los problemas de salud derivados de la concentración de personas en las ciudades, como la contaminación y la falta de acceso a servicios sanitarios adecuados. Además, el envejecimiento poblacional puede derivar en sobresaturación de los servicios sanitarios.

El **ODS 4**, "Garantizar una educación inclusiva, equitativa y de calidad, y promover oportunidades de aprendizaje durante toda la vida para todos", también está vinculado, ya que el aumento de la población urbana puede generar presión sobre los sistemas educativos, dificultando el acceso y la calidad de la educación para todos. Además, la despoblación puede dejar a alumnos sin acceso cercano a centros de enseñanza en las zonas rurales.

El **ODS 9**, "Construir infraestructuras resilientes, promover la industrialización sostenible y fomentar la innovación", tiene relación con la necesidad de adaptar las infraestructuras urbanas a un mayor número de habitantes, lo cual requiere inversión en innovación y sostenibilidad.

El **ODS 13**, "Adoptar medidas urgentes para combatir el cambio climático y sus efectos", también se ve influenciado por el crecimiento descontrolado de las ciudades, que aumenta las emisiones de carbono y la huella ecológica.

Finalmente, el **ODS 15**, "Gestionar sosteniblemente los bosques, luchar contra la desertificación, detener e invertir la degradación de las tierras, detener la pérdida de biodiversidad", se ve afectado por el fenómeno de la migración, ya que la despoblación de las zonas rurales puede llevar al abandono de tierras agrícolas y contribuir a la degradación del medio ambiente, afectando la biodiversidad y los ecosistemas.

Bibliografía

- [1] ONU-Hábitat, “¿cómo definir ciudades, pueblos y áreas rurales?” 2021, consultado el 14 de enero de 2025. [Online]. Available: <https://onu-habitat.org/index.php/como-definir-ciudades-pueblos-y-areas-rurales>
- [2] Ministerio de Agricultura, Pesca y Alimentación (MAPA), “Demografía en la población rural 2020,” 2020, accedido: 30-dic-2024. [Online]. Available: https://www.mapa.gob.es/va/ministerio/servicios/analisis-y-prospectiva/ayp_demografiaenlapoblacionrural2020_tcm39-583987.pdf
- [3] Instituto Geográfico Nacional (IGN), “Sistema urbano y áreas metropolitanas,” 2024, accedido: 30-dic-2024. [Online]. Available: https://atlasnacional.ign.es/wane/Sistema_urbano_y_%C3%A1reas_metropolitanas
- [4] Defensor del Pueblo, “La situación demográfica en España. reflexiones sobre el modelo territorial,” 2019, accedido: 30-dic-2024. [Online]. Available: https://www.defensordelpueblo.es/wp-content/uploads/2019/06/Separata_situacion_demografica.pdf
- [5] Math. Publishing, “Mathematical modelling of human population dynamics,” n.d., recuperado el 30-dic-2024. [Online]. Available: <https://www.researchpublish.com/upload/book/Mathematical%20Modelling%20of%20Human-7794.pdf>
- [6] AMSE Modelling, “Production engineering and management organization: Human and social problems,” n.d., recuperado el 30-dic-2024. [Online]. Available: https://amsemodelling.com/publications/modelling_measurement_and_contr/Production_Engineering_and_Management_Organization_Human_and_Social_Problems/391/39.01_01.pdf
- [7] Programa de las Naciones Unidas para el Desarrollo (PNUD), “Guía rápida de población,” 2021, accedido: 30-dic-2024. [Online]. Available: https://www.undp.org/sites/g/files/zskgke326/files/2022-09/PADIT_Gu%C3%ADa%20r%C3%A1pida%20sobre%20la%20din%C3%A1mica%20de%20la%20poblaci%C3%B3n.pdf
- [8] M. D. Gómez. (2022) Trabajo fin de máster: Revisión de técnicas de análisis demográfico y aplicación a datos reales, poyecciones de población. [Onli-

- ne]. Available: https://masteres.ugr.es/estadistica-aplicada/sites/master/moea/public/inline-files/TFM_D%C3%ADazG%C3%B3mez_Marcos.pdf
- [9] Instituto Nacional de Estadística (INE). (2017) Técnicas y ejercicios de demografía. [Online]. Available: https://www.ine.es/ine/oposiciones/temario_2021/Tecnicas_ejercicios_demografia.pdf
- [10] D. A. Swanson and L. Tedrow, "Using cohort change ratios to estimate life expectancy in populations with negligible migration: A new approach," *Canadian Studies in Population*, vol. 39, no. 1-2, pp. 83–90, 2012.
- [11] S. Smith, J. Tayman, and D. A. Swanson, *State and local population projections: Methodology and analysis*. New York: Kluwer Academic Press, 2001.
- [12] D. A. Swanson, A. Schlottmann, and R. Schmidt, "Forecasting the population of census tracts by age and sex: An example of the hamilton-perry method in action," *Population Research and Policy Review*, vol. 29, no. 1, pp. 47–63, 2010.
- [13] N. Gilbert, *Agent-Based Models*, ser. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: SAGE Publications, Inc., 2008, published online in 2011.
- [14] T. C. Schelling, "Models of segregation," RAND Corporation, Research Memorandum RM-6014-RC, May 1969. [Online]. Available: http://www.rand.org/pubs/research_memoranda/RM6014.html
- [15] J. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press, 2007. [Online]. Available: <http://www.cs.unibo.it/babaoglu/courses/cas/papers/Epstein%20-%202006%20-%20Generative%20Social%20Science%20Studies%20in%20Agent-Based%20Computational%20Modeling.pdf>
- [16] J. I. G.-V. Medina, "La simulación basada en agentes: una nueva forma de explorar los fenómenos sociales," *REIS - Revista Española de Investigaciones Sociológicas*, vol. 136, pp. 91–113, 2011. [Online]. Available: <https://dialnet.unirioja.es/descarga/articulo/3748058.pdf>
- [17] Álvaro Vélez Torres, "Modelación y simulación basada en agentes en ciencias sociales: una aproximación al estado del arte," *Polis*, vol. 18, no. 53, p. 1392, 2019. [Online]. Available: <http://dx.doi.org/10.32735/s0718-6568/2019-n53-1392>
- [18] P. González, *Análisis de Series Temporales: Modelos ARIMA*. Departamento de Economía Aplicada III (Econometría y Estadística) Facultad de Ciencias Económicas y Empresariales Universidad del País Vasco (UPV-EHU), 2009. [Online]. Available: https://addi.ehu.es/bitstream/handle/10810/12492/04-09gon.pdfzGrIthhU9jtwavkNjoYw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- [19] G. T. y. R. T. D. Peña, *A Course in Time Series Analysis*, 3rd ed. John Wiley, 2005.

- [20] J. A. Jiménez, “Introducción a los modelos de series temporales,” 2007. [Online]. Available: <https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>
- [21] D. G. Ramos, “Análisis de series temporales,” 2020. [Online]. Available: https://oa.upm.es/70926/1/TFG_DAVID_GUTIERREZ_RAMOS.pdf
- [22] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. OTexts, 2021. [Online]. Available: <https://otexts.com/fpp3/>
- [23] P. y. A. Ministerio de Agricultura, “Características demográficas de la población rural,” 2025, accessed: 2025-01-04. [Online]. Available: https://www.mapa.gob.es/ministerio/pags/biblioteca/fondo/pdf/18405_6.pdf
- [24] Instituto Nacional de Estadística, “Metodología de la base de datos estadísticos,” 2021, consultado: enero 2025. [Online]. Available: https://www.ine.es/metodologia/t20/metodologia_idb.pdf
- [25] United Nations Department of Economic and Social Affairs, “World urbanization prospects: The 2018 revision, methodology,” 2018, accessed: 2025-01-10. [Online]. Available: <https://population.un.org/wup/assets/WUP2018-Methodology.pdf>
- [26] R. J. Hyndman and G. Athanasopoulos, *forecast: Forecasting functions for time series and linear models*, 2024, accessed: 2025-01-12. [Online]. Available: <https://cran.r-project.org/web/packages/forecast/forecast.pdf>
- [27] H. Wickham, R. Francois, L. Henry, and K. Müller, “purrr: Functional programming tools,” 2025, accessed: 2025-01-12. [Online]. Available: <https://purrr.tidyverse.org/>
- [28] H. Wickham, “ggplot2: Create elegant data visualisations using the grammar of graphics,” 2021. [Online]. Available: <https://ggplot2.tidyverse.org/>
- [29] M.-H. Education, “La profecía de malthus: ¿un futuro sin recursos?” 2021, accedido: 2025-01-12. [Online]. Available: <https://www.mheducation.es/blog/la-profeca-malthus>
- [30] C. SER, “Más mascotas que niños,” 2024, accedido: 2025-01-12. [Online]. Available: https://cadenaser.com/andalucia/2024/10/21/mas-mascotas-que-ninos-radio-jerez/?utm_source=chatgpt.com
- [31] Eurostat, “Rural development and the challenges of depopulation,” 2021, european Commission. [Online]. Available: <https://ec.europa.eu/eurostat/web/rural-development>
- [32] M. García, Jesús y Fernández, “El envejecimiento poblacional y sus efectos en la economía,” *Gaceta Sanitaria*, vol. 26, no. 4, pp. 283-286, 2012. [Online]. Available: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1134-928X2012000400002
- [33] D. E. Bloom, D. Canning, and J. Sevilla, “The impact of population aging on economic growth, the labor force and productivity,” National Bureau

BIBLIOGRAFÍA

- of Economic Research, Tech. Rep. w16705, 2011. [Online]. Available: https://www.nber.org/system/files/working_papers/w16705/w16705.pdf
- [34] Instituto Nacional de Estadística, “Proyecciones de población. años 2024-2074,” 2024, Último acceso: 13 de enero de 2025. [Online]. Available: <https://www.ine.es/dyngs/Prensa/PROP20242074.htm>
- [35] Naciones Unidas, “Objetivos de desarrollo sostenible,” 2015. [Online]. Available: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

Anexos

Apéndice A

Datos: tablas de tasas y poblaciones anuales

Este anexo se ha realizado para incluir tablas de datos que eran demasiado grandes y dificultaban la legibilidad del trabajo.

Capítulo A. Datos: tablas de tasas y poblaciones anuales

Grupo de Edad	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012
Total	9.67	9.48	10.40	8.83	9.10	9.07	8.79	9.06	8.48	8.34	8.58
De 0 a 4 años	0.61	0.55	0.54	0.58	0.59	0.60	0.62	0.59	0.64	0.60	0.69
De 5 a 9 años	0.08	0.07	0.06	0.06	0.07	0.07	0.07	0.08	0.07	0.09	0.09
De 10 a 14 años	0.10	0.08	0.07	0.08	0.08	0.08	0.07	0.08	0.09	0.08	0.09
De 15 a 19 años	0.18	0.17	0.15	0.16	0.17	0.17	0.17	0.17	0.16	0.17	0.19
De 20 a 24 años	0.28	0.27	0.25	0.23	0.25	0.25	0.23	0.24	0.27	0.24	0.27
De 25 a 29 años	0.29	0.30	0.31	0.29	0.31	0.31	0.28	0.30	0.27	0.29	0.31
De 30 a 34 años	0.42	0.42	0.43	0.37	0.39	0.37	0.39	0.36	0.39	0.38	0.42
De 35 a 39 años	0.56	0.58	0.58	0.53	0.54	0.56	0.53	0.55	0.52	0.58	0.62
De 40 a 44 años	0.89	0.87	0.90	0.79	0.85	0.86	0.89	0.93	0.98	0.99	1.08
De 45 a 49 años	1.39	1.46	1.53	1.48	1.52	1.59	1.62	1.69	1.80	1.85	1.92
De 50 a 54 años	2.56	2.64	2.80	2.67	2.73	2.80	2.90	2.96	2.95	3.08	3.09
De 55 a 59 años	4.32	4.46	4.64	4.31	4.46	4.48	4.50	4.71	4.60	4.67	4.73
De 60 a 64 años	6.90	7.17	7.36	6.65	6.84	6.71	6.84	6.89	6.85	6.96	7.10
De 65 a 69 años	10.48	10.86	11.14	9.76	10.02	10.13	10.09	10.22	10.06	10.19	10.42
De 70 a 74 años	16.08	16.34	17.42	15.05	15.44	15.63	15.52	15.85	15.77	15.92	16.66
De 75 a 79 años	26.80	26.95	29.52	24.67	25.94	26.37	27.11	28.87	28.05	28.27	29.75
De 80 a 84 años	48.53	50.19	57.21	48.01	50.26	50.55	49.84	53.64	51.03	51.97	55.62
De 85 y más años	140.58	133.15	151.41	127.52	133.53	135.47	131.03	139.55	129.85	130.69	140.83

Cuadro A.2: Tabla de tasa de mortalidad por año, 2022-2012

Grupo de Edad	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002
Total	8.26	8.17	8.26	8.36	8.47	8.33	8.82	8.62	9.06	8.84
De 0 a 4 años	0.72	0.75	0.79	0.84	0.87	0.89	0.95	1.01	1.04	1.06
De 5 a 9 años	0.08	0.08	0.09	0.11	0.10	0.11	0.11	0.11	0.12	0.14
De 10 a 14 años	0.11	0.10	0.11	0.11	0.13	0.12	0.13	0.14	0.16	0.15
De 15 a 19 años	0.22	0.22	0.26	0.31	0.33	0.33	0.38	0.38	0.42	0.38
De 20 a 24 años	0.29	0.30	0.33	0.37	0.39	0.43	0.47	0.48	0.53	0.54
De 25 a 29 años	0.32	0.33	0.40	0.41	0.44	0.45	0.47	0.53	0.56	0.58
De 30 a 34 años	0.43	0.44	0.47	0.50	0.54	0.59	0.66	0.68	0.78	0.80
De 35 a 39 años	0.65	0.68	0.73	0.82	0.88	0.95	1.02	1.08	1.11	1.14
De 40 a 44 años	1.14	1.21	1.26	1.38	1.47	1.48	1.53	1.56	1.70	1.67
De 45 a 49 años	1.96	2.04	2.15	2.26	2.25	2.28	2.36	2.49	2.53	2.55
De 50 a 54 años	3.25	3.28	3.34	3.47	3.52	3.57	3.61	3.68	3.69	3.74
De 55 a 59 años	4.86	4.95	4.97	5.04	5.15	5.20	5.31	5.30	5.55	5.46
De 60 a 64 años	7.04	7.04	7.23	7.43	7.69	7.62	7.81	7.82	8.28	8.42
De 65 a 69 años	10.39	10.47	11.03	11.08	11.70	11.81	12.60	12.68	13.13	13.18
De 70 a 74 años	17.07	17.58	18.22	18.63	19.33	19.19	20.64	20.46	21.70	21.75
De 75 a 79 años	29.55	30.24	31.51	32.56	33.59	33.73	36.43	36.15	38.28	38.45
De 80 a 84 años	55.21	55.87	58.51	60.64	62.67	61.70	67.63	65.82	69.61	68.77
De 85 y más años	134.67	136.73	142.08	147.08	150.66	149.58	164.72	158.75	170.79	163.20

Cuadro A.4: Tabla de tasa de mortalidad por año, 2011-2002

Capítulo A. Datos: tablas de tasas y poblaciones anuales

Cohorte de Edad	2023	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
De 0 a 4 años	32864	40785	16258	16692	31159	26733	22196	13855	7691	1530	-6565	-3788	1338	2508	4683	16199
De 5 a 9 años	38112	47304	14075	13006	25964	17472	8727	3966	-471	-3546	-7467	-3375	1786	2231	5726	19869
De 10 a 14 años	39158	46045	11436	11028	22737	16046	10492	7807	3209	88	-4270	-884	4116	4622	9167	24687
De 15 a 19 años	59953	64432	25275	22910	44224	35629	25099	18073	11271	6661	-392	6081	12856	13051	19872	39998
De 20 a 24 años	84815	94822	34236	37547	72447	61122	44321	26372	17360	8562	-5756	4868	16670	14013	17644	61133
De 25 a 29 años	94177	104090	27251	38241	74138	60923	37980	18942	3798	-11554	-38492	-22883	-6869	-9733	-1885	52854
De 30 a 34 años	79625	88591	13603	26790	53135	39836	15562	2971	-10515	-27811	-56702	-41456	-25354	-24880	-15758	29238
De 35 a 39 años	64146	71074	5932	17102	36621	24318	2706	-3876	-14225	-26646	-45701	-33416	-21631	-19322	-13569	19097
De 40 a 44 años	47397	50793	-1405	9036	23272	12458	-2411	-5113	-11923	-18736	-33281	-23752	-14548	-13712	-9154	13198
De 45 a 49 años	34064	35805	-1777	6543	18318	9935	-1631	-2317	-6421	-11312	-21137	-13753	-8275	-7958	-6161	9312
De 50 a 54 años	25054	27032	-9	5848	14833	8478	540	34	-3161	-7313	-13997	-8771	-3853	-4707	-3131	6095
De 55 a 59 años	17673	20796	3062	7013	13415	7892	1012	1631	-1133	-3904	-7873	-3896	-279	-1476	-374	5731
De 60 a 64 años	13167	16503	3212	5903	10967	8058	3341	3275	1451	269	-2746	1026	2994	2114	3023	6812
De 65 a 69 años	12370	10531	2696	4110	9179	6320	2669	2302	1006	642	-1287	1800	2979	1726	2503	4066
De 70 a 74 años	3430	6896	-422	364	3802	1645	-1491	-203	-198	-939	-2373	111	616	-110	618	1461
De 75 a 79 años	-368	2707	-1599	-950	793	-738	-2343	-523	-218	-848	-2081	-277	-115	-302	-47	602
De 80 a 84 años	-1019	820	-1960	-997	-344	-763	-1518	-121	214	-267	-873	-269	33	-418	-83	226
85+	-2322	-2021	-1794	-829	-448	-1206	-1979	349	544	147	-539	84	-162	-322	-228	65

Cuadro A.6: Tabla de saldo migratorio exterior por año, 2023-2008

Apéndice B

Scripts de R: implementación de la simulación y del análisis

Este anexo incluye el código usado para la implementación de la simulación y la representación de las gráficas para el análisis. Se ha decidido incluirlo en un anexo para facilitar la legibilidad y comprensión del trabajo.

B.1. Población en el primer año

```
#Poblacion Espana 2023
poblacion = c(1786406, 2221807, 2529002, 2562806,
             2502990, 2573472, 2799479, 3085598, 3739321,
             4055364, 3800108, 3583098, 3158134, 2631651,
             2271259, 1913012, 1268777, 1588417)
poblacion_porcentajes=rep(0,18)
for (i in 1:length(poblacion)){
  poblacion_porcentajes[i]=poblacion[i]*100/sum(poblacion)
}

#Datos iniciales
poblacion_rural=rep(0,18)
poblacion_rural_porc=c(13.9,62.6,23.5)
poblacion_urbana=rep(0,18)
poblacion_urbana_porc=c(16.4,66.6,17.0)
total_rural=round(sum(poblacion)*0.159)
total_urbana=round(sum(poblacion)*0.841)

# Población de jóvenes, adultos y mayores en rural
poblacion_rural_jovenes = total_rural * poblacion_rural_porc[1]/100
poblacion_rural_adultos = total_rural * poblacion_rural_porc[2]/100
poblacion_rural_mayores = total_rural * poblacion_rural_porc[3]/100

# Población de jóvenes, adultos y mayores en urbano
poblacion_urbana_jovenes = total_urbana * poblacion_urbana_porc[1]/100
poblacion_urbana_adultos = total_urbana * poblacion_urbana_porc[2]/100
poblacion_urbana_mayores = total_urbana * poblacion_urbana_porc[3]/100

# Porcentaje total de jóvenes (cohortes 1 a 3)
porcentaje_total_jovenes = sum(poblacion_porcentajes[1:3])
# Porcentaje total de adultos (cohortes 4 a 13)
porcentaje_total_adultos = sum(poblacion_porcentajes[4:13])
# Porcentaje total de mayores (cohortes 14 a 18)
porcentaje_total_mayores = sum(poblacion_porcentajes[14:18])
```

Capítulo B. Scripts de R: implementación de la simulación y del análisis

```
# Cálculo para población rural y urbana por cohorte, con distribución por porcentaje
for (i in 1:length(poblacion)) {
  # Para los jóvenes (cohortes 1 a 3)
  if (i <= 3) {
    poblacion_rural[i] = round(poblacion_rural_jovenes * poblacion_porcentajes[i] /
                              porcentaje_total_jovenes)
    poblacion_urbana[i] = round(poblacion_urbana_jovenes * poblacion_porcentajes[i] /
                                porcentaje_total_jovenes)
  }
  # Para los adultos (cohortes 4 a 13)
  else if (i >= 4 && i <= 13) {
    poblacion_rural[i] = round(poblacion_rural_adultos * poblacion_porcentajes[i] /
                              porcentaje_total_adultos)
    poblacion_urbana[i] = round(poblacion_urbana_adultos * poblacion_porcentajes[i] /
                                porcentaje_total_adultos)
  }
  # Para los mayores (cohortes 14 a 18)
  else if (i >= 14 && i <= 18) {
    poblacion_rural[i] = round(poblacion_rural_mayores * poblacion_porcentajes[i] /
                              porcentaje_total_mayores)
    poblacion_urbana[i] = round(poblacion_urbana_mayores * poblacion_porcentajes[i] /
                                porcentaje_total_mayores)
  }
}
```

B.2. Proyección de fecundidad

```
library(purrr)
```

```
fecundidad <- data.frame(
  "2002" = c(0, 0, 0.12, 9.86, 27.70, 63.57, 95.51, 46.45, 7.56, 0.39, 0.25, 0, 0, 0, 0,
            0, 0, 0),
  "2003" = c(0, 0, 0.16, 10.55, 28.79, 64.32, 98.27, 49.44, 8.23, 0.42, 0.26, 0, 0, 0,
            0, 0, 0),
  "2004" = c(0, 0, 0.14, 10.75, 30.39, 63.72, 99.06, 51.16, 8.73, 0.43, 0.27, 0, 0, 0,
            0, 0, 0),
  "2005" = c(0, 0, 0.15, 11.35, 30.59, 63.03, 99.29, 52.98, 9.29, 0.48, 0.28, 0, 0, 0,
            0, 0, 0),
  "2006" = c(0, 0, 0.16, 12.01, 32.56, 64.37, 99.29, 55.40, 9.88, 0.66, 0.29, 0, 0, 0,
            0, 0, 0),
  "2007" = c(0, 0, 0.14, 12.82, 34.68, 63.95, 97.07, 57.07, 10.33, 0.72, 0.30, 0, 0, 0,
            0, 0, 0),
  "2008" = c(0, 0, 0.17, 13.26, 37.91, 66.99, 99.66, 60.57, 11.33, 0.73, 0.31, 0, 0, 0,
            0, 0, 0),
  "2009" = c(0, 0, 0.12, 11.83, 34.90, 62.95, 94.84, 60.37, 11.61, 0.86, 0.32, 0, 0, 0,
            0, 0, 0),
  "2010" = c(0, 0, 0.12, 10.61, 33.85, 61.68, 94.48, 61.73, 12.30, 0.87, 0.33, 0, 0, 0,
            0, 0, 0),
  "2011" = c(0, 0, 0.14, 9.51, 30.81, 59.51, 93.53, 62.96, 13.08, 0.87, 0.31, 0, 0, 0,
            0, 0, 0),
  "2012" = c(0, 0, 0.13, 8.99, 29.33, 57.89, 91.65, 62.69, 14.05, 0.90, 0.37, 0, 0, 0,
            0, 0, 0),
  "2013" = c(0, 0, 0.13, 8.38, 27.07, 55.82, 88.20, 60.94, 14.43, 0.91, 0.24, 0, 0, 0,
            0, 0, 0),
  "2014" = c(0, 0, 0.12, 8.09, 26.72, 57.80, 92.04, 63.59, 15.19, 1.05, 0.34, 0, 0, 0,
            0, 0, 0),
  "2015" = c(0, 0, 0.09, 7.74, 26.17, 57.41, 93.15, 65.25, 15.91, 1.10, 0.33, 0, 0, 0,
            0, 0, 0),
  "2016" = c(0, 0, 0.10, 7.36, 25.71, 57.01, 92.96, 66.33, 16.46, 1.27, 0.37, 0, 0, 0,
            0, 0, 0),
  "2017" = c(0, 0, 0.10, 7.07, 25.26, 54.68, 90.42, 65.73, 16.79, 1.42, 0.44, 0, 0, 0,
            0, 0, 0),
  "2018" = c(0, 0, 0.06, 6.26, 24.24, 51.69, 86.41, 64.23, 16.93, 1.49, 0.50, 0, 0, 0,
            0, 0, 0),
```

B.2. Proyección de fecundidad

```
"2019" = c(0, 0, 0.07, 5.99, 24.08, 49.75, 83.61, 63.73, 16.64, 1.49, 0.50, 0, 0, 0,
0, 0, 0, 0),
"2020" = c(0, 0, 0.05, 5.36, 22.76, 47.69, 79.87, 61.75, 16.30, 1.43, 0.47, 0, 0, 0,
0, 0, 0, 0),
"2021" = c(0, 0, 0.05, 4.61, 20.40, 45.67, 81.97, 63.97, 17.10, 1.57, 0.59, 0, 0, 0,
0, 0, 0, 0),
"2022" = c(0, 0, 0.06, 4.61, 20.45, 44.72, 78.32, 63.42, 17.27, 1.74, 0.67, 0, 0, 0,
0, 0, 0, 0),
"2023" = c(0, 0, 0.06, 4.31, 20.19, 43.48, 75.49, 60.85, 16.85, 1.60, 0.60, 0, 0, 0,
0, 0, 0, 0)
)

# Diferencias año a año por cohorte
diferencias <- t(apply(fecundidad, 1, diff))

# La deriva como promedio de las diferencias
mu <- rowMeans(diferencias, na.rm = TRUE)

# La desviación estándar del ruido
sigma <- apply(diferencias, 1, sd, na.rm = TRUE)

# Número de años a simular
n_años <- 50

simular_fecundidad <- funcion(fecundidad, mu, sigma, n_años) {
  fecundidad_sim <- fecundidad

  for (t in 1:n_años) {

    # Generar ruido estocástico para cada cohorte
    ruido <- rnorm(nrow(fecundidad_sim), mean = 0, sd = sigma)

    # Sumar tendencia y ruido a los valores del año anterior
    nuevo_año <- fecundidad_sim[, ncol(fecundidad_sim)] + mu + ruido

    # Fijar valor mínimo posible como 0, no puede haber fecundidad negativa
    nuevo_año <- pmax(nuevo_año, 0)

    # Agregar el nuevo año al dataframe
    fecundidad_sim <- cbind(fecundidad_sim, nuevo_año)
    colnames(fecundidad_sim)[ncol(fecundidad_sim)] <- paste0("Sim_", 2023 + t)
  }
  return(fecundidad_sim)
}

# Número de simulaciones
n_simulaciones <- 1000

# Realización de simulaciones usando map de la librería purr
simulaciones <- map(1:n_simulaciones, ~simular_fecundidad(fecundidad, mu, sigma, n_años)
)

# Extraer los datos de las simulaciones (columnas correspondientes a los años simulados)
fecundidad_sim_resumido <- do.call(abind::abind, list(simulaciones, along = 3))

# Calcular la media de las simulaciones por cohorte y año
media_simulaciones <- apply(fecundidad_sim_resumido, c(1, 2), mean, na.rm = TRUE)

# Convertir la media a un data frame para guardar como tabla
tabla_media_simulaciones <- as.data.frame(media_simulaciones)

# Actualizar la tabla solo con los años simulados, dándole nombres a las filas y
columnas
tabla_media_simulaciones <- tabla_media_simulaciones[23:72]
rownames(tabla_media_simulaciones) <- rownames(fecundidad)
colnames(tabla_media_simulaciones) <- paste0(2024:(2024 + n_años - 1))
```

Capítulo B. Scripts de R: implementación de la simulación y del análisis

```
#Guardamos la tabla para usarla posteriormente
save(tabla_media_simulaciones, file = "media_simulaciones_fecundidad.RData")
```

B.3. Proyección de mortalidad

```
library(forecast)

# Crea el dataframe con tus datos
data <- data.frame(
  años = 2002:2023,
  cohort_0_4 = c(1.06, 1.04, 1.01, 0.95, 0.89, 0.87, 0.84, 0.79, 0.75, 0.72, 0.69, 0.60,
    0.64, 0.59, 0.62, 0.60, 0.59, 0.58, 0.54, 0.55, 0.61, 0.64),
  cohort_5_9 = c(0.14, 0.12, 0.11, 0.11, 0.11, 0.11, 0.10, 0.11, 0.11, 0.10, 0.09, 0.09, 0.08,
    0.08, 0.09, 0.08, 0.07, 0.07, 0.06, 0.06, 0.07, 0.08, 0.07),
  cohort_10_14 = c(0.15, 0.16, 0.14, 0.13, 0.12, 0.13, 0.11, 0.11, 0.10, 0.09, 0.08,
    0.08, 0.07, 0.08, 0.07, 0.08, 0.07, 0.08, 0.08, 0.07, 0.10, 0.07),
  cohort_15_19 = c(0.38, 0.42, 0.38, 0.33, 0.33, 0.31, 0.26, 0.22, 0.22, 0.19, 0.17,
    0.17, 0.17, 0.16, 0.17, 0.16, 0.17, 0.15, 0.17, 0.17, 0.18, 0.16),
  cohort_20_24 = c(0.54, 0.53, 0.48, 0.47, 0.43, 0.39, 0.37, 0.33, 0.30, 0.29, 0.27,
    0.24, 0.27, 0.24, 0.23, 0.25, 0.25, 0.23, 0.25, 0.27, 0.28, 0.27),
  cohort_25_29 = c(0.58, 0.56, 0.53, 0.47, 0.45, 0.44, 0.41, 0.40, 0.33, 0.32, 0.31,
    0.29, 0.28, 0.30, 0.31, 0.29, 0.31, 0.31, 0.29, 0.30, 0.29, 0.30),
  cohort_30_34 = c(0.80, 0.78, 0.68, 0.66, 0.59, 0.54, 0.50, 0.47, 0.44, 0.43, 0.42,
    0.38, 0.39, 0.36, 0.39, 0.37, 0.39, 0.42, 0.43, 0.42, 0.42, 0.40),
  cohort_35_39 = c(1.14, 1.11, 1.08, 1.02, 0.95, 0.88, 0.82, 0.73, 0.68, 0.65, 0.62,
    0.58, 0.58, 0.53, 0.56, 0.54, 0.53, 0.58, 0.56, 0.58, 0.56, 0.54),
  cohort_40_44 = c(1.67, 1.70, 1.56, 1.53, 1.48, 1.47, 1.38, 1.26, 1.21, 1.14, 1.08,
    0.99, 0.98, 0.93, 0.89, 0.86, 0.85, 0.79, 0.90, 0.87, 0.89, 0.84),
  cohort_45_49 = c(2.55, 2.53, 2.49, 2.36, 2.28, 2.25, 2.26, 2.15, 2.04, 1.96, 1.92,
    1.85, 1.80, 1.69, 1.62, 1.59, 1.52, 1.48, 1.53, 1.46, 1.39, 1.33),
  cohort_50_54 = c(3.74, 3.69, 3.68, 3.61, 3.57, 3.52, 3.47, 3.34, 3.28, 3.25, 3.09,
    3.08, 2.95, 2.96, 2.90, 2.80, 2.73, 2.67, 2.80, 2.64, 2.56, 2.34),
  cohort_55_59 = c(5.46, 5.55, 5.30, 5.31, 5.20, 5.15, 5.04, 4.97, 4.95, 4.86, 4.73,
    4.67, 4.60, 4.71, 4.50, 4.48, 4.46, 4.31, 4.64, 4.46, 4.32, 4.08),
  cohort_60_64 = c(8.42, 8.28, 7.82, 7.81, 7.62, 7.69, 7.43, 7.23, 7.04, 7.04, 7.10,
    6.96, 6.85, 6.89, 6.84, 6.71, 6.84, 6.65, 7.36, 7.17, 6.90, 6.49),
  cohort_65_69 = c(13.18, 13.13, 12.68, 12.60, 11.81, 11.70, 11.08, 11.03, 10.47, 10.39,
    10.42, 10.19, 10.06, 10.22, 10.09, 10.13, 10.02, 9.76, 11.14, 10.86, 10.48, 9.9),
  cohort_70_74 = c(21.75, 21.70, 20.46, 20.64, 19.19, 19.33, 18.63, 18.22, 17.58, 17.07,
    16.66, 15.92, 15.77, 15.85, 15.52, 15.63, 15.44, 15.05, 17.42, 16.34, 16.08,
    14.97),
  cohort_75_79 = c(38.45, 38.28, 36.15, 36.43, 33.73, 33.59, 32.56, 31.51, 30.24, 29.55,
    29.75, 28.27, 28.05, 28.87, 27.11, 26.37, 25.94, 24.67, 29.52, 26.95, 26.80,
    25.22),
  cohort_80_84 = c(68.77, 69.61, 65.82, 67.63, 61.70, 62.67, 60.64, 58.51, 55.87, 55.21,
    55.62, 51.97, 51.03, 53.64, 49.84, 50.55, 50.26, 48.01, 57.21, 50.19, 48.53,
    45.36),
  cohort_85_plus = c(163.20, 170.79, 158.75, 164.72, 149.58, 150.66, 147.08, 142.08,
    136.73, 134.67, 140.83, 130.69, 129.85, 139.55, 131.03, 135.47, 133.53, 127.52,
    151.41, 133.15, 140.58, 137.87)
)

#Datos necesarios para llevar a cabo el modelo
resultados <- data.frame(cohort = character(0), alpha = numeric(0), beta = numeric(0),
  varianza_diferencias = numeric(0), media_diferencias = numeric(0))

# Obtenemos cada columna como una serie temporal menos la columna de años
for (cohorte in names(data)[-1]) {
  serie_temporal <- ts(data[[cohorte]], start = 2002, frequency = 1)

  modelo <- ets(serie_temporal, model = "AAN")

  # Extraer los valores de alpha y beta
```

B.3. Proyección de mortalidad

```
alpha <- modelo$par["alpha"]
beta <- modelo$par["beta"]

# Calcular las diferencias, media de diferencias y varianza
cohort_diff <- diff(data[[cohort]])
cohort_mean_diff <- mean(cohort_diff)
varianza_diferencias <- var(cohort_diff)

resultados <- rbind(resultados, data.frame(cohort = cohort, alpha = alpha, beta =
  beta, varianza_diferencias = varianza_diferencias, media_diferencias =
  cohort_mean_diff))
}
años=50

#La función para poder simular
simular_mortalidad <- función(resultados, alpha, beta, años){

#Declaración de variables y datos iniciales
tendencia_mortalidad <- rep(0,18)
tasa_mortalidad <- matrix(0, nrow = 18, ncol = años)
tasa_mortalidad[,1] <- c(0.64, 0.07, 0.07, 0.16, 0.27, 0.30, 0.40, 0.54, 0.84, 1.332,
  2.34,
  4.08, 6.49, 9.904, 14.97, 25.22, 45.36, 137.87)
media_mortalidad <- resultados$media_diferencias
var_mort <- resultados$varianza_diferencias
tasa_mortalidad[,1] <- tasa_mortalidad[,1] + media_mortalidad

#Aplicamos el modelo de suavizado exponencial con componente estocástica
for (t in 2:años) {
  error_estocastico <- mapply(función(var) rnorm(1, mean = 0, sd = sqrt(var)), var_mort
  )
  if( t==2){
    #EL modelo pide información de dos años anteriores por lo que hasta t=3 realizamos:
    tendencia_mortalidad <- media_mortalidad
    tasa_mortalidad[, t] <- alpha*tasa_mortalidad[, t-1]+(1-alpha)*(tasa_mortalidad[, t-1]+
    tendencia_mortalidad)+error_estocastico
  }
  else{
    tendencia_mortalidad <- beta*(tasa_mortalidad[, t-1]-tasa_mortalidad[, t-2])+(1-beta)*
    tendencia_mortalidad
    tasa_mortalidad[, t] <- alpha*tasa_mortalidad[, t-1]+(1-alpha)*(tasa_mortalidad[, t-1]+
    tendencia_mortalidad)+error_estocastico
  }
  tasa_mortalidad <- pmax(tasa_mortalidad, 0)
}
return(tasa_mortalidad)
}

n_simulaciones <- 1000

# Realizamos simulaciones con purrr
simulaciones <- purrr::map(1:n_simulaciones, ~ simular_mortalidad(resultados, alpha,
  beta, años))

# Convertir las simulaciones a un formato adecuado y calculamos la media
simulaciones_array <- array(unlist(simulaciones), dim = c(18, años, n_simulaciones))
media_simulaciones <- apply(simulaciones_array, c(1, 2), mean)

media_simulaciones_df <- as.data.frame(media_simulaciones)
colnames(media_simulaciones_df) <- paste0("Año_", 2024:2073)
#Guardamos resultados para usarlos después
save(media_simulaciones, file = "media_simulaciones_mortalidad.RData")
```

B.4. Proyección de saldo migratorio exterior

```
library(forecast)
library(purrr)

library(forecast)
library(purrr)

data <- data.frame(
  años = 2008:2023,
  cohort_0_4 = c(16199, 4683, 2508, 1338, -3788, -6565, 1530, 7691, 13855, 22196, 26733,
    31159, 16692, 16258, 40785, 32864),
  cohort_5_9 = c(19869, 5726, 2231, 1786, -3375, -7467, -3546, -471, 3966, 8727, 17472,
    25964, 13006, 14075, 47304, 38112),
  cohort_10_14 = c(24687, 9167, 4622, 4116, -884, -4270, 88, 3209, 7807, 10492, 16046,
    22737, 11028, 11436, 46045, 39158),
  cohort_15_19 = c(39998, 19872, 13051, 12856, 6081, -392, 6661, 11271, 18073, 25099,
    35629, 44224, 22910, 25275, 64432, 59953),
  cohort_20_24 = c(61133, 17644, 14013, 16670, 4868, -5756, 8562, 17360, 26372, 44321,
    61122, 72447, 37547, 34236, 94822, 84815),
  cohort_25_29 = c(52854, -1885, -9733, -6869, -22883, -38492, -11554, 3758, 18942,
    37980, 60923, 74158, 38241, 27251, 104090, 94177),
  cohort_30_34 = c(29238, -15758, -24880, -25354, -41456, -56702, -27811, -10515, 2971,
    15562, 39836, 53135, 26790, 13603, 88591, 79625),
  cohort_35_39 = c(19097, -13569, -19322, -21631, -33416, -45701, -26646, -14225, -3876,
    2706, 24318, 36621, 17102, 5932, 71074, 64146),
  cohort_40_44 = c(13198, -9154, -13712, -14548, -23752, -33281, -18736, -11923, -5113,
    -2411, 12458, 23272, 9036, -1405, 50793, 47397),
  cohort_45_49 = c(9312, -6161, -7958, -8275, -13753, -21137, -11312, -6421, -2317,
    -1631, 9935, 18318, 6543, -1777, 35805, 34064),
  cohort_50_54 = c(6095, -3131, -4707, -3853, -8771, -13997, -7313, -3161, 34, 540,
    8478, 14833, 5848, -9, 27032, 25054),
  cohort_55_59 = c(5731, -374, -1476, -279, -3896, -7873, -3904, -1133, 1631, 1012,
    7892, 13415, 7013, 3062, 20796, 17673),
  cohort_60_64 = c(6812, 3023, 2114, 2994, 1026, -2746, 269, 1451, 3275, 3341, 8058,
    10967, 5903, 3212, 16503, 13167),
  cohort_65_69 = c(4066, 2503, 1726, 2979, 1800, -1287, 642, 1006, 2302, 2669, 6320,
    9179, 4110, 2696, 10531, 12370),
  cohort_70_74 = c(1461, 618, -110, 616, 111, -2373, -939, -198, -203, -1491, 1645,
    3802, 364, -422, 6896, 3430),
  cohort_75_79 = c(602, -47, -302, -115, -277, -2081, -848, -218, -523, -2343, -738,
    793, -950, -1599, 2707, -368),
  cohort_80_84 = c(226, -83, -418, 33, -269, -873, -267, 214, -121, -1518, -763, -344,
    -997, -1960, 820, -1019),
  cohort_85_plus = c(65, -228, -322, -162, 84, -539, 147, 544, 349, -1979, -1206, -448,
    -829, -1794, -2021, -2322)
)

resultados_arima <- list()
sm_ano <- rep(0,50)
predicciones_estocasticas <- list()

n_simulaciones <- 100 # Número de simulaciones a realizar

for (cohorte in names(data)[-1]) {

  # Creamos serie temporal para cada cohorte
  serie_temporal <- ts(data[[cohorte]], start = 2008, frequency = 1)

  # Ajustar el modelo ARIMA a la serie temporal
  modelo_arima <- auto.arima(serie_temporal, seasonal = FALSE, stepwise = FALSE,
    approximation = FALSE)

  # Realizamos la predicción para los próximos 50 años
  prediccion <- forecast(modelo_arima, h = 50)

  predicciones_simuladas <- map(1:n_simulaciones, function(simulacion) {
```

B.5. Proyección de la migración rural-urbana

```
for (t in 1:50) {  
  # Calculamos la predicción estocástica para este año  
  sm_año[t] <- predicción$mean[t] + (predicción$upper[t, 2] - predicción$lower[t,  
    2]) * (runif(1) - 0.5)  
}  
  
return(sm_año)  
})  
  
# Calcular la media de todas las simulaciones  
predicciones_estocasticas[[cohorta]] <- round(rowMeans(do.call(cbind,  
  predicciones_simuladas)))  
}  
  
# Convertimos el resultado en una matriz de predicciones estocásticas para su posterior  
  uso  
predicciones_estocasticas_matriz <- do.call(cbind, predicciones_estocasticas)  
  
save(predicciones_estocasticas_matriz, file = "media_simulaciones_migracion_externa.  
  RData")
```

B.5. Proyección de la migración rural-urbana

```
PU_0 <- 0.841  
PU_menos1 <- 0.83965  
  
# Calculo del ratio urbano-rural unicial  
URR_0 <- PU_0 / (1 - PU_0)  
  
# Calculo del diferencial urbano-rural unicial  
rur_0 <- log(PU_0 / (1 - PU_0)) - log(PU_menos1 / (1 - PU_menos1))  
  
# Función para calcular el valor hrur(PU_t)  
hrur <- function(PU_t) {  
  return(0.032561 - 0.0153693 * PU_t)  
}  
  
# Número de años a proyectar  
años <- 51  
# Vamos a poner que los años son 51 debido a que guardamos la información inicial en el  
  año 1.  
  
# Declaramos las variables  
rur <- numeric(años)  
URR <- numeric(años)  
PU <- numeric(años)  
  
# Asignamos valores iniciales  
rur[1] <- rur_0  
URR[1] <- URR_0  
PU[1] <- PU_0  
  
# Realizamos la proyección para los primeros 25 años  
for (t in 2:26) {  
  hrur_PU_t <- hrur(PU[t - 1])  
  rur[t] <- rur[t - 1] + (hrur_PU_t - rur[t - 1]) / (25 - (t - 2))  
  URR[t] <- URR[t - 1] * exp(rur[t])  
  PU[t] <- URR[t] / (1 + URR[t])  
}  
  
# Realizar la proyección después de t=25
```

Capítulo B. Scripts de R: implementación de la simulación y del análisis

```
for (t in 27:years) {  
  rur[t] <- hrur(PU[t - 1])  
  URR[t] <- URR[t - 1] * exp(rur[t])  
  PU[t] <- URR[t] / (1 + URR[t])  
}  
  
# Crear un dataframe con los resultados  
proporcion_urbana <- data.frame(  
  Year = 1:years,  
  proporcion_urbana = PU  
)  
#Guardamos el resultado para usarlo en la proyección de la población  
save(proporcion_urbana, file = "proporcion_urbana.RData")
```

B.6. Proyección de población

```
anios <- 50  
  
load("media_simulaciones_fecundidad.RData")  
load("media_simulaciones_mortalidad.RData")  
load("media_simulaciones_migracion_externa.RData")  
load("proporcion_urbana.RData")  
tasa_fecundidad <- tabla_media_simulaciones  
tasa_mortalidad <- media_simulaciones  
saldo_migratorio_ext <- predicciones_estocasticas_matriz  
proporcion_urban <- proporcion_urbana$proporcion_urbana  
porcentajes_migracion_interna <- c(5.03, 4.05, 4.03, 4.67, 7.97, 12.14, 12.37, 10.30,  
  8.98, 7.53,  
  5.88, 4.44, 3.43, 2.74, 1.89, 1.51, 1.12, 1.91)  
  
porcentaje_urbana=rep(0,50)  
saldo_migratorio <- c(32864, 38112, 39158, 59953, 84815, 94177,  
  79625, 64146, 47397, 34064, 25054, 17673,  
  13167, 12370, 3430, -368, -1019, -2322)  
migracion_interna_acum <- rep(0,50)  
nacidos_urbana <- rep(0,50)  
nacidos_rural <- rep(0,50)  
muertos_urbana <- rep(0,50)  
muertos_rural <- rep(0,50)  
muertos <- rep(0,50)  
  
# Almacenamiento de resultados de la población total y por cohortes  
resultados_rural <- numeric(anios + 1)  
resultados_urbana <- numeric(anios + 1)  
  
# Cohortes de edad (inicialmente vacías)  
cohortes_rural <- rep(0, 18)  
cohortes_urbana <- rep(0, 18)  
  
# Inicializamos las poblaciones totales en el primer año  
cohortes_rural <- poblacion_rural  
cohortes_urbana <- poblacion_urbana  
  
resultados_rural[1] <- sum(cohortes_rural)  
resultados_urbana[1] <- sum(cohortes_urbana)  
  
# Simulación por cada año  
for (t in 1:anios) {  
  if (t==18){  
    #browser()  
  }  
  poblacion_total <- resultados_rural[t] + resultados_urbana[t]
```

B.6. Proyección de población

```
porcentaje_urbana[t] <- (resultados_urbana[t] / poblacion_total)

mujeres_edad_fertil_rural <- round(cohortes_rural * 0.5)
mujeres_edad_fertil_urbana <- round(cohortes_urbana * 0.5)
nacimientos_rural <- round(tasa_fecundidad[t]/1000 * mujeres_edad_fertil_rural)
nacimientos_urbana <- round(tasa_fecundidad[t]/1000 * mujeres_edad_fertil_urbana)

muertes_rural <- round(cohortes_rural * (tasa_mortalidad[,t]/1000))
muertes_urbana <- round(cohortes_urbana * (tasa_mortalidad[,t]/1000))
nacidos_urbana[t] <- sum(nacimientos_urbana)
muertos_urbana[t] <- sum(muertes_urbana)
nacidos_rural[t] <- sum(nacimientos_rural)
muertos_rural[t] <- sum(muertes_rural)

# Migración externa por cohorte
migracion_externa_rural <- round(saldo_migratorio_ext[t]*((1 - porcentaje_urbana[t])))
migracion_externa_urbana <- round(saldo_migratorio_ext[t]*porcentaje_urbana[t])

#Migración interna, esto me dice cuanta gente debería ir a urbana o irse de urbana(de
  ser negativo)

migracion_interna_total <- proporcion_urban[t+1]*poblacion_total-resultados_urbana[t]
migracion_interna <- round(migracion_interna_total*porcentajes_migracion_interna/100)
migracion_interna_acum[t] <- sum(migracion_interna)

# Actualización de la población por cohorte
cohortes_rural <- cohortes_rural - muertes_rural + migracion_externa_rural -
  migracion_interna
cohortes_urbana <- cohortes_urbana - muertes_urbana + migracion_externa_urbana +
  migracion_interna

# Desplazamiento de cohortes (cada 5 años)
#if (t %% 5 == 0) {
  # El penúltimo grupo (grupo 16 en este caso) no sustituye al último, sino que suma
  su valor
  cohortes_rural[length(cohortes_rural)] <- cohortes_rural[length(cohortes_rural)] +
    round(cohortes_rural[length(cohortes_rural) - 1]*0.2)
  cohortes_urbana[length(cohortes_urbana)] <- cohortes_urbana[length(cohortes_urbana)]
    + round(cohortes_urbana[length(cohortes_urbana) - 1]*0.2)
  # Desplazar las cohortes (todos menos el último grupo)
  for (i in (length(cohortes_rural) - 1):2) {
    cohortes_rural[i] <- round(cohortes_rural[i] * 0.8)+ round(cohortes_rural[i
      -1]*0.2)
    cohortes_urbana[i] <- round(cohortes_urbana[i] * 0.8)+ round(cohortes_urbana[i
      -1]*0.2)
  }
  # Los nacimientos pasan al primer grupo (0-4 años)
  cohortes_rural[1] <- round(cohortes_rural[1] * 0.8) + sum(nacimientos_rural)
  cohortes_urbana[1] <- round(cohortes_urbana[1] * 0.8) + sum(nacimientos_urbana)

  resultados_rural[t+1] <- sum(cohortes_rural)
  resultados_urbana[t+1] <- sum(cohortes_urbana)

  #acum_nacimientos_rural <- 0
  #acum_nacimientos_urbana <- 0

#else{
# Guardar resultados por cohorte
resultados_rural[t+1] <- sum(cohortes_rural)
resultados_urbana[t+1] <- sum(cohortes_urbana)

#}
}
```

Capítulo B. Scripts de R: implementación de la simulación y del análisis

```
nacimientos <- nacidos_rural+nacidos_urbana
defunciones <- muertos_rural+muertos_urbana

Analisis <- data.frame(
  pob_urb <- resultados_urbana[2:51],
  pob_rur <- resultados_rural[2:51],
  total <- resultados_urbana[2:51] + resultados_rural[2:51],
  crecimiento_vegetativo <- nacimientos-defunciones,
  migracion_interna <- migracion_interna_acum,
  ult_urb <- c(cohortes_urbana, rep(0,32)),
  ult_rural <- c(cohortes_rural, rep(0,32))
)

save(Analisis, file = "Resultados.RData")
```

B.7. Código usado para las gráficas

```
library(ggplot2)
años <- 2024:2073
grupos_edad <- c("0-4", "05-9", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39",
  "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", "70-74", "75-79", "80-84",
  "85+")
options(scipen = 999)

load("fecundidad_sim_solo.RData")
fecundidades <- fecundidad_sim_solo[3:11, ]
colnames(fecundidades) <- años
rownames(fecundidades) <- grupos_edad[3:11]

dev.new()
colores <- c("red", "blue", "green", "purple", "orange", "cyan", "brown", "pink", "
  yellow")
matplot(t(fecundidades), type = "l", xlab = "Año", ylab = "Fecundidad",
  col = colores, lty = 1, lwd = 2,
  xaxt = "n", main = "Evolución de la fecundidad por cohorte de edad")
axis(1, at = 1:length(años), labels = años)
legend("topleft", legend = grupos_edad[3:11],
  fill = colores,
  ncol = 2)

load("media_simulaciones_mortalidad.RData")
mortalidad <- media_simulaciones[10:18,]
colnames(mortalidad) <- años
rownames(mortalidad) <- grupos_edad[10:18]

dev.new()
matplot(t(mortalidad), type = "l", xlab = "Año", ylab = "Tasa de Mortalidad",
  col = colores, lty = 1, lwd = 2,
  xaxt = "n", main = "Evolución de la mortalidad por cohorte de edad")
axis(1, at = 1:length(años), labels = años)
legend("topleft", legend = grupos_edad[10:18],
  fill = colores,
  ncol = 2)

load("Resultados.RData")
Analisis_m <- as.matrix(Analisis)
pob_urb <- Analisis[,1]
pob_rur <- Analisis[,2]
total <- Analisis[,3]
```

B.7. Código usado para las gráficas

```
crecimiento_vegetativo <- Analisis[,4]
migracion_interior <- Analisis[,5]
urbana <- Analisis[1:18,6]
rural <- Analisis[1:18,7]
# Gráfico para "pob_urb"
dev.new() # Nueva ventana
plot(pob_urb, type = "l", col = colores[1], lty = 1, lwd = 2,
      xlab = "Año", ylab = "Población Urbana", main = "Evolución de la Población Urbana",
      ylim = c(min(pob_urb) / 2, max(pob_urb) * 1.5))
axis(1, at = 1:length(amos), labels = amos)

# Gráfico para "pob_rur"
dev.new() # Nueva ventana
plot(pob_rur, type = "l", col = colores[2], lty = 1, lwd = 2,
      xlab = "Año", ylab = "Población Rural", main = "Evolución de la Población Rural",
      ylim = c(min(pob_rur) / 2, max(pob_rur) * 1.5))
axis(1, at = 1:length(amos), labels = amos)

# Gráfico para "total"
dev.new() # Nueva ventana
plot(total, type = "l", col = colores[3], lty = 1, lwd = 2,
      xlab = "Año", ylab = "Población Total", main = "Evolución de la Población Total",
      ylim = c(min(total) / 2, max(total) * 1.5))
axis(1, at = 1:length(amos), labels = amos)


dev.new() # Nueva ventana
plot(crecimiento_vegetativo, type = "l", col = colores[2], lty = 1, lwd = 2,
      xlab = "Año", ylab = "Crecimiento Vegetativo",
      main = "Evolución del Crecimiento Vegetativo",
      ylim = c(min(crecimiento_vegetativo) * 2, max(crecimiento_vegetativo) / 2))
axis(1, at = 1:length(amos), labels = amos)

dev.new() # Nueva ventana
plot(migracion_interior, type = "l", col = colores[1], lty = 1, lwd = 2,
      xlab = "Año", ylab = "Saldo migratorio rural-urbano",
      main = "Evolución del saldo migratorio rural-urbano",
      ylim = c(min(migracion_interior/2), max(migracion_interior*1.5)))
axis(1, at = 1:length(amos), labels = amos)

df_piramide <- data.frame(grupos_edad, rural, urbana)

ggplot() +
  geom_bar(data = df_piramide, aes(x = -rural, y = grupos_edad, fill = "Rural"), stat =
    "identity") +
  geom_bar(data = df_piramide, aes(x = urbana, y = grupos_edad, fill = "Urbana"), stat =
    "identity") +
  scale_x_continuous(labels = abs) + # Mostrar valores absolutos en el eje X
  labs(title = "Pirámide Poblacional: Rural vs Urbana".) +
  theme_minimal() +
  theme(legend.position = "bottom") +
  scale_fill_manual(values = c("Rural" = "blue", "Urbana" = "red"))
```

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
Fecha/Hora	Tue Jan 14 11:39:45 CET 2025
Emisor del Certificado	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
Numero de Serie	561
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)