



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CAMPUS D'ALCOI

Estudio de la evolución del IPV en la Comunidad Valenciana

MEMORIA PRESENTADA POR:

Enrique Borrás Nieto

GRADO DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Tutoras: Pérez Molina Ana Isabel y García Cifuentes, Noemí

Convocatoria de defensa: Septiembre 2021

Resumen

El precio de la vivienda en España, y en particular el IPV ha experimentado un crecimiento muy grande durante estos años, posteriores a la crisis del 2007, en este trabajo se va a centrar en los índice de precio de la vivienda que se han ido registrado hasta la fecha en la Comunidad Valenciana. Este trabajo se enfoca en analizar la evolución del IPV junto con el análisis del estudio del arte para tratar de saber que variables apuntan a la existencia de un papel preponderante sobre la variable a estudiar, y así junto a estas variables tratar de estudiar su evolución en los últimos periodos de los trimestres de la IPV en la C.V.

Resume

The price of housing in Spain, and in particular the Housing Price Index has experienced a very large growth during these years, after the crisis of 2007, in this work it is going to focus on the home price index that has been to date in the Valencian Community. This work focuses on analyzing the evolution of the IPV together with the analysis of the study of art to try to know which variables point to the existence of a preponderant role on the variable to be studied, and thus together with these variables try to study its evolution in the last periods of the quarters of the IPV in the CV.

Resum

El preu de l'habitatge a Espanya, i en particular l'IPV ha experimentat un creixement molt gran durant aquests anys, posteriors a la crisi de l'any 2007, en aquest treball es centrarà en els índex de preus de l'habitatges que s'han anat registrat fins a la data en la Comunitat Valenciana. Aquest treball s'enfoca en analitzar l'evolució de l'IPV juntament amb l'anàlisi de l'estudi de l'art per intentar saber que variables apunten a l'existència d'un paper preponderant sobre la variable a estudiar, i així al costat d'aquestes variables tractar d'estudiar la seva evolució en els últims períodes dels trimestres de la IPV a la CV.

Palabras Claves

IPV; crisis inmobiliaria; mercado inmobiliario; covid-19; actividades inmobiliarias

KEYWORDS

Housing price index, real estate crisis, real-estate market, Covid-19, Real estate activities

Agradecimientos:

Quería dar las gracias, en primer lugar, a mi familia y a mi pareja por aguantarme todos los días y hacer que no tirara la toalla. En segundo lugar, a todos los profesores con los que me he ido cruzando durante los años, por haberme hecho más fácil mi estancia en la Universidad. También quería dar las gracias en especial a mis dos tutoras de TFG, por no rendirse y darme caña cuando realmente lo necesitaba. Por último, quería hacer una mención especial a David Hervás Marín por la ayuda recibida en los momentos más críticos del TFG.

ÍNDICE DE CONTENIDOS

Resumen.....	2
Resume.....	2
Resum.....	2
Palabras Claves.....	2
KEYWORDS.....	2
1.-Introducción.....	15
2.Objetivos.....	16
2.1 Objetivo principal.....	16
2.2 Objetivos secundarios.....	16
3.Metodología.....	16
4.-Hipótesis.....	17
5.-Revisión bibliográfica.....	18
5.1. Análisis de los factores determinantes del precio del activo vivienda.....	18
5.2 Análisis del precio de la vivienda en España.....	19
5.3 Análisis regional del mercado laboral y de la inflación.....	21
5.4 El precio de la vivienda y la inflación en España.....	22
5.5 Precios de la vivienda: sobrevaloración y burbuja.....	22
5.6 Análisis y pronóstico del precio de la vivienda en España: modelo econométrico desde una perspectiva conductual.....	23
5.7 Factores determinantes de la diferencia de precios en el mercado inmobiliario.....	24
5.8 Evolución de precios en el mercado inmobiliario (vivienda).....	26
5.9 Determinantes del precio de viviendas en la región metropolitana de Chile.....	27
5.10 Un análisis sobre el gasto en servicios de vivienda en España.....	28
5.11 Conclusión sobre la revisión bibliográfica.....	30
6.-Determinación de las variables sobre las que se va a realizar el estudio.....	30
6.1.-Variable endógena o dependiente (Y).....	31
6.2.-Variables explicativas o independientes (X).....	31
7.-Análisis univariable.....	34
7.1.- X_1 : Renta. Anexo I tabla renta.....	36
7.2.- X_{11} : Salarios. Anexo I tabla salarios.....	39
7.3.- X_2 : Población 30-44 años. Anexo I tabla pobl.30-44.....	41
7.4.- X_{22} : Nacimientos. Anexo I tabla nacimientos.....	43

7.5.-X ₃ : Euribor. Anexo I tabla euribor	45
7.6.-X ₄ : Parados. Anexo I Tabla nº parados.....	47
7.7.-X ₅ : Viviendas Iniciadas. Anexo I tabla viviendas iniciadas.....	49
7.8.-X ₆ : Densidad de Población. Anexo I tabla densidad población.....	51
7.9.-X ₇ : PIB. Anexo I tabla PIB.....	54
7.10.-X ₈ : Precio de M ² . Anexo I tabla precio m ²	56
7.11.-X ₉ : Empresas Activas. Anexo I tabla empresas activas.	58
8.-Análisis Bivariante	60
8.1.-Modelo de regresión simple: IPV en función de la variable Salarios.....	63
8.2.- Modelo de regresión simple: IPV en función de la variable Nacimientos	65
8.3.- Modelo de regresión simple: IPV en función de la variable Euribor.....	67
8.4.- Modelo de regresión simple: IPV en función de la variable Parados	70
8.5.- Modelo de regresión simple: IPV en función de la variable Viviendas Iniciadas	72
8.6.- Modelo de regresión simple: IPV en función de la variable Densidad Población.....	75
8.7.- Modelo de regresión simple: IPV en función de la variable PIB	77
8.8.- Modelo de regresión simple: IPV en función de la variable M ²	80
8.9.-Conclusiones sobre el análisis bivariante.	82
9.-Series de Tiempo.....	83
9.1.-Modelo serie temporal IPV	84
9.2.-Modelo serie temporal Salario	90
9.3.-Modelo serie temporal Nacimientos	94
9.4.-Modelo serie temporal Euribor	100
9.5.-Modelo serie temporal Parados	105
9.6.-Modelo serie temporal Viviendas Iniciadas.....	109
9.7.-Modelo serie temporal Densidad de Población.	116
9.8.-Modelo serie temporal PIB	120
9.9.-Modelo serie temporal M ²	123
10.-Análisis de la multicolinealidad.....	129
10.1.-Matriz de correlación 1	130
10.2.-Matriz de correlación inversa 1	131
10.3.-Índice de acondicionamiento 1.....	131
10.4.-Matriz correlación 2.....	132
10.5.- Matriz inversa correlación 2	132
10.6.-Índice de acondicionamiento 2.....	133

10.7.-Matriz correlación 3	134
10.8.-Matriz inversa correlación 3	134
10.9.- Índice acondicionamiento 3	135
11.-Realización de un modelo de regresión.....	136
11.1.-Modelo teórico y ajustado	136
11.2.-Análisis de la significatividad	137
11.2.1.-Contraste de significación global del modelo	137
11.3.-Detección gráfica de posibles problemas.	138
11.3.1.-Residuos Vs X	139
11.3.2.-Residuos VS Y predicho	141
11.4.-Estudio de normalidad de residuos	142
11.4.1.-Contraste de hipótesis	142
11.4.2.-Gráfico probabilístico.....	143
11.4.3.-Histograma de residuos	143
11.5.-Estudio autocorrelación.....	144
11.6.-Estudio de la heterocedasticidad.....	146
11.6.1.-Parados	146
11.6.2.-M ²	146
11.6.3.-Nº Viviendas Iniciadas	147
11.6.4.-Densidad Población.....	147
12.- Nuevo modelo de regresión penalizado basado en el método de Elastic Net.	147
12.1.-Método de regresión Ridge.	148
12.2.-Metodo de regresión Lasso.	149
12.3.-Metodo Elastic Net.	150
12.3.1.- Formulación teórica del modelo penalizado Elastic net	150
12.3.2.-Validación cruzada y su ejecución	152
12.3.3.- Validación modelo regresión penalización Elastic Net	154
13.- Nuevo modelo ARIMAX.....	156
13.1.-Concepto ARIMAX.....	157
13.2.- Modelo ARIMAX con regresores.....	157
13.2.1.-Validación modelo ARIMAX con los regresores: Viviendas, M ² , Parados, PIB y Euribor.....	158
13.2.2.-Análisis de las variables.....	158
13.2.3.-Validación modelo ARIMAX con los regresores: m ² , parados y PIB	160
13.2.4.-Análisis y validación de las variables.....	160

13.2.5.- Selección mejor modelo ARIMAX	161
13.2.6.-Prediccion modelo ARIMAX con AIC= 158,37	161
14.- Objetivo secundario: Comparativa de la evolución IPV entre comunidades.....	163
14.1-Estudio individual de las comunidades.....	163
14.1.1.-Andalucia	163
14.1.2.-Aragón.....	163
14.1.3.-Asturias	164
14.1.4.-Balears.....	165
14.1.5.-Canarias.....	165
14.1.6.-Cantabria.....	166
14.1.7.-Castilla y León	166
14.1.8.-Castilla -La Mancha	167
14.1.9.-Cataluña	167
14.1.10.-Comunidad Valenciana	168
14.1.11.-Extremadura	168
14.1.12.-Galicia.....	169
14.1.13.-Comunidad Madrid	169
14.1.14.-Región de Murcia.....	170
14.1.15.-Navarra.....	170
14.1.16.-País Vasco.....	171
14.1.17.-La Rioja.....	172
14.1.18.-Ceuta	172
14.1.19.-Melilla	173
14.2.-Estudio global de las comunidades.....	174
15.-Conclusiones finales.....	175
15.1. Conclusiones sobre las hipótesis planteadas	175
15.2. Conclusiones generales.....	176

ÍNDICE DE GRÁFICAS

Gráfica 1	Dispersión Renta	37
Gráfica 2	Caja y Bigotes Renta	38
Gráfica 3	Histograma Renta	38
Gráfica 4	Dispersión Salario	39
Gráfica 5	Caja y Bigotes Salario	40
Gráfica 6	Histograma Salario	40
Gráfica 7	Dispersión Población 30-44	41
Gráfica 8	Cajas y Bigotes Población 30-44	42
Gráfica 9	Histograma Población 30-44	42
Gráfica 10	Dispersión Nacimientos	44
Gráfica 11	Caja y Bigotes Nacimientos	44
Gráfica 12	Histograma Nacimientos	45
Gráfica 13	Dispersión Euríbor	46
Gráfica 14	Caja y Bigotes Euríbor	46
Gráfica 15	Histograma Euríbor	47
Gráfica 16	Dispersión Parados	48
Gráfica 17	Caja y Bigotes Parados	48
Gráfica 18	Histograma Parados	49
Gráfica 19	Dispersión Vivienda	50
Gráfica 20	Caja y Bigotes Vivienda	50
Gráfica 21	Histograma Vivienda	51
Gráfica 22	Dispersión Densidad Población	52
Gráfica 23	Caja y Bigotes Densidad Población	53
Gráfica 24	Histograma Densidad Población	53
Gráfica 25	Dispersión PIB	54
Gráfica 26	Caja y Bigotes PIB	55
Gráfica 27	Histograma PIB	55
Gráfica 28	Dispersión M^2	56
Gráfica 29	Caja y Bigotes M^2	57
Gráfica 30	Histograma M^2	57
Gráfica 31	Dispersión Empresas Activas	58
Gráfica 32	Caja y Bigotes Empresas Activas	59
Gráfica 33	Histograma Empresas Activas	59
Gráfica 34	Autocorrelación Residuos Salarios	65
Gráfica 35	Autocorrelaciones Residuos Nacimientos	67
Gráfica 36	Autocorrelación Residuos Euríbor	69
Gráfica 37	Autocorrelación Residuos Parados	72
Gráfica 38	Autocorrelación Residuo Vivienda	74
Gráfica 39	Autocorrelación Residuos Densidad Población	77
Gráfica 40	Autocorrelación Residuos PIB	79
Gráfica 41	Autocorrelaciones Estimadas Residuos M^2	82
Gráfica 42	Serie Tiempo IPV 1	84
Gráfica 43	FAS IPV	85
Gráfica 44	FAP IPV	85

Gráfica 45 Serie Tiempo Ajustada 2 IPV.....	86
Gráfica 46 FAS Ajustado IPV.....	86
Gráfica 47 FAP Ajustado IPV.....	87
Gráfica 48 Observación de residuos IPV.....	88
Gráfica 49 Histograma Residuos IPV.....	88
Gráfica 50 FAS Residuos IPV.....	89
Gráfica 51 Serie Tiempo Salario.....	90
Gráfica 52 FAS Salario.....	91
Gráfica 53 FAP Salario.....	91
Gráfica 54 Serie Tiempo Ajustada Salario.....	92
Gráfica 55 FAS Ajustado Salarios.....	92
Gráfica 56 FAP Ajustado Salarios.....	93
Gráfica 57 Serie Tiempo Nacimientos.....	94
Gráfica 58 FAS Nacimiento.....	95
Gráfica 59 FAP Nacimiento.....	95
Gráfica 60 Serie Tiempo Ajustada Nacimiento.....	96
Gráfica 61 FAS Ajustado Nacimiento.....	96
Gráfica 62 FAP Ajustado Nacimiento.....	97
Gráfica 63 Observación de residuos Nacimientos.....	98
Gráfica 64 Histograma Residuos Nacimiento.....	99
Gráfica 65 FAS Residuos Nacimientos.....	99
Gráfica 66 Series Tiempo Euribor.....	100
Gráfica 67 FAS Euribor.....	101
Gráfica 68 FAP Euribor.....	101
Gráfica 69 Serie Tiempo Ajustada Euribor.....	102
Gráfica 70 FAS Ajustado Euribor.....	102
Gráfica 71 FAP Ajustado Euribor.....	103
Gráfica 72 Series Tiempo Parados.....	104
Gráfica 73 Series Tiempo Parados.....	105
Gráfica 74 FAS Parados.....	105
Gráfica 75 FAP Parados.....	106
Gráfica 76 Serie Tiempo Ajustes Parados.....	106
Gráfica 77 FAS Ajuste Parados.....	107
Gráfica 78 FAP Ajuste Parados.....	107
Gráfica 79 Residuos Ajuste Parados.....	108
Gráfica 80 Histograma Residuos Parados.....	109
Gráfica 81 Serie Tiempo VivlNi.....	109
Gráfica 82 FAS VivlNi.....	110
Gráfica 83 FAP VivlNi.....	110
Gráfica 84 Serie Tiempo Ajustado VivlNi.....	111
Gráfica 85 FAS Ajustado VivlNi.....	111
Gráfica 86 FAP Ajustado VivlNi.....	112
Gráfica 87 Residuos Ajuste VivlNi.....	113
Gráfica 88 Histograma Residuos VivlNi.....	114
Gráfica 89 FAS Ajustado VivlNi.....	114
Gráfica 90 FAP Ajustado VivlNi.....	115

Gráfica 91 Serie Tiempo DensPobla	116
Gráfica 92 FAS DensPobla	116
Gráfica 93 FAP DensPobla	117
Gráfica 94 Serie Tiempo Ajustada DensPobla	117
Gráfica 95 FAS Ajustado DensPobla	118
Gráfica 96 FAP Ajustado DensPobla	118
Gráfica 97 Serie Tiempo PIB	120
Gráfica 98 FAS PIB	120
Gráfica 99 FAP PIB	121
Gráfica 100 Serie Tiempo Ajustada PIB	121
Gráfica 101 FAS Ajustado PIB	122
Gráfica 102 FAP Ajustado PIB	122
Gráfica 103 Serie Tiempo M ²	123
Gráfica 104 FAS M ²	124
Gráfica 105 FAP M ²	124
Gráfica 106 Serie Tiempo Ajustada M ²	125
Gráfica 107 FAS Ajustado M ²	125
Gráfica 108 FAP Ajustado M ²	126
Gráfica 109 Observación Residuos Ajuste M ²	127
Gráfica 110 Histograma Residuos M ²	128
Gráfica 111 FAS Residuos M ²	128
Gráfica 112 Residuos vs Parados Modelo Regresión	139
Gráfica 113 Residuos vs M ² Modelo Regresión	139
Gráfica 114 Residuos vs Vivlni Modelo Regresión	140
Gráfica 115 Residuos vs DensPobla Modelo Regresión	140
Gráfica 116 Residuos vs Salario Modelo Regresión	141
Gráfica 117 Residuos vs Predicho IPV Modelo Regresión	141
Gráfica 118 Probabilidad Normal Residuos Modelo Regresión	143
Gráfica 119 Histograma Modelo Regresión	143
Gráfica 120 FAS Modelo Regresión	144
Gráfica 121 FAP Modelo Regresión	144
Gráfica 122 Residuos VS IPV. Fuente: R Studio	155
Gráfica 123 Probabilístico normal de los residuos. Fuente: R Studio	155
Gráfica 124 Residuos número de filas. Fuente: R Studio	156
Gráfica 125 Representación predicción para 4 periodos. Fuente: R Studio	162
Gráfica 126 IPV Andalucía	163
Gráfica 127 IPV Aragón	163
Gráfica 128 IPV Asturias	164
Gráfica 129 IPV Baleares	165
Gráfica 130 IPV Canarias	165
Gráfica 131 IPV Cantabria	166
Gráfica 132 IPV Catilla y León	166
Gráfica 133 IPV Castilla-La Mancha	167
Gráfica 134 IPV Cataluña	167
Gráfica 135 IPV Comunidad Valenciana	168
Gráfica 136 IPV Extremadura	168

Gráfica 137 IPV Galicia	169
Gráfica 138 IPV Comunidad Madrid	169
Gráfica 139 IPV Región Murcia.....	170
Gráfica 140 IPV Principado Navarra	170
Gráfica 141 IPV País Vasco	171
Gráfica 142 IPV La Rioja.....	172
Gráfica 143 IPV Ceuta.....	172
Gráfica 144 IPV Melilla	173
Gráfica 145 Evolución IPV por Comunidades Autónomas	174

ÍNDICE DE TABLAS

Tabla 1 Resumen estadístico Renta.	36
Tabla 2 Resumen Estadístico para Salario.....	39
Tabla 3 Resumen Estadístico Población 30-44	41
Tabla 4 Resumen estadístico Nacimientos.....	43
Tabla 5 Resumen estadístico Euribor.	45
Tabla 6 Resumen estadístico Parados.....	47
Tabla 7 Resumen estadístico Vivini.....	49
Tabla 8 Resumen estadístico Densidad Población.....	51
Tabla 9 Resumen estadístico PIB.....	54
Tabla 10 Resumen estadístico M^2	56
Tabla 11 Resumen estadístico Empresas Activas.....	58
Tabla 12 Análisis Varianza Salarios.....	63
Tabla 13 Pruebas Normalidad Residuos Salarios.	64
Tabla 14 Autocorrelación Residuos Salario.....	64
Tabla 15 Análisis Varianza Nacimientos.....	65
Tabla 16 Pruebas Normalidad Residuos Nacimientos.	66
Tabla 17 Autocorrelación Residuos Nacimientos.....	67
Tabla 18 Análisis Varianza Euribor.	68
Tabla 19 Prueba Normalidad Residuos Euribor	68
Tabla 20 Autocorrelación Residuos Euribor	69
Tabla 21 Análisis Varianza Parados	70
Tabla 22 Pruebas Normalidad Residuos Parados.....	71
Tabla 23 Autocorrelación Residuos Parados.	71
Tabla 24 Análisis Varianza Vivini.	72
Tabla 25 Pruebas Normalidad Residuos Vivini.....	73
Tabla 26 Autocorrelación Residuos Vivini.....	74
Tabla 27 Análisis Varianza Densidad Población.	75
Tabla 28 Prueba Normalidad Residuos Densidad Población.	76
Tabla 29 Autocorrelación Residuos Densidad Población.....	76
Tabla 30 Análisis Varianza PIB.....	77
Tabla 31 Normalidad Residuos PIB.	78
Tabla 32 Autocorrelación Residuos PIB.....	79
Tabla 33 Análisis Varianza M^2	80
Tabla 34 Pruebas Normalidad Residuos M^2	81
Tabla 35 Autocorrelaciones Residuos M^2	81
Tabla 36 Resumen Modelo ARIMA IPV.	87
Tabla 37 Pruebas Normalidad Residuos IPV	88
Tabla 38 Pronósticos IPV.	89
Tabla 39 Resumen Modelo ARIMA Salarios.....	93
Tabla 40 Resumen Modelo ARIMA Nacimientos.	97
Tabla 41 Prueba Normalidad Residuos Nacimiento.....	98
Tabla 42 Pronósticos Nacimientos.	100
Tabla 43 Resumen ARIMA Euribor.	103
Tabla 44 Pruebas Normalidad Residuos Euribor.....	104

Tabla 45 Resumen Modelo ARIMA Parados.....	107
Tabla 46 Pruebas Normalidad Parados.....	108
Tabla 47 Resumen Modelo ARIMA Vivini.....	112
Tabla 48 Normalidad Residuos Vivini.....	113
Tabla 49 Pronósticos Vivini.....	115
Tabla 50 Modelo ARIMA DensPobla.....	119
Tabla 51 Modelo ARIMA PIB.....	122
Tabla 52 Resumen Modelo ARIMA M ²	126
Tabla 53 Prueba Normalidad Residuos M ² -.....	127
Tabla 54 Matriz correlación 1 estudio.....	130
Tabla 55 Matriz correlación inversa 1 estudio.....	131
Tabla 56 Índice acondicionamiento 1 estudio.....	131
Tabla 57 Matriz correlación 2 estudio.....	132
Tabla 58 Matriz inversa correlación 2 estudio.....	132
Tabla 59 Índice acondicionamiento 2 estudio.....	133
Tabla 60 Matriz correlación 3 estudio.....	134
Tabla 61 Matriz inversa correlación 3 estudio.....	134
Tabla 62 Índice acondicionamiento 3 estudio.....	135
Tabla 63 Análisis Varianza Modelo Regresión.....	137
Tabla 64 Análisis Varianza Individua Modelo Regresión.....	137
Tabla 65 Análisis Varianza Modelo Regresión 2.....	142
Tabla 66 Análisis Varianza Individual Modelo Regresión 2.....	142
Tabla 67 Normalidad Residuos Modelo Regresión.....	142
Tabla 68 Heterocedasticidad.....	146

ÍNDICE DE IMÁGENES

Imagen 1 Resumen Serie Temporal. Fuente:UPV	83
Imagen 2 Modelo de regresión penalizado Elastic Net. Fuente: R Studio	151
Imagen 3 Validación Cruzada. Fuente: R studio.....	153
Imagen 4 Resultado Modelo Regresión Penalizado. Fuente R Studio.	154
Imagen 5 Apuntes Econometría. Fuente: UPV EPSA.....	157
Imagen 6 Apuntes Econometría. Fuente: UPV EPSA.....	157
Imagen 7 ARIMAX con Regresores. Fuente: UPV EPSA.....	157
Imagen 8 ARIMAX. Fuente R Studio	158
Imagen 9 P-Valor Modelo ARIMAX. Fuente: R Studio.....	158
Imagen 10 ARIMAX 2. Fuente: R Studio.....	160
Imagen 11 P-Valor Modelo ARIMAX 2. Fuente: R Studio.....	160
Imagen 12 Predicción para 4 periodos. Fuente: R Studio	161

1.-Introducción.

El Índice de Precios de Vivienda (IPV) es un índice elaborado en España por el Instituto Nacional de Estadística (INE), este fue publicado por primera vez el 1 de octubre de 2008 con la finalidad de mostrar la evolución de los precios de la vivienda. Además, el IPV tiene como objetivo calcular el progreso de los precios de compraventa de las viviendas de precio libre, tanto para las nuevas como para las de segunda mano, a lo largo del tiempo.

España ha sido un país muy marcado por la burbuja inmobiliaria durante muchos años. La burbuja inmobiliaria comenzó en el año 2001 y duró hasta el 2007, el máximo histórico que se alcanzó respecto al pico de precios fue en el año 2007.

Seguidamente empezó la crisis en España, en el ámbito económico comenzó en 2008 y esta provocó la mayor crisis inmobiliaria en el país, lo que afectó mucho y de manera directa al Índice de Precio de la Vivienda. Esta crisis concluyó en el año 2014, pero la economía española y sobre todo los trabajos relacionados con la construcción y el inmueble tardaron muchos más años en recuperarse. Expertos afirman que la bajada de los precios iba a continuar varios años más allá de 2014 debido a que existía un gran parque inmobiliario vacío lo que conllevaba una ausencia de demanda por diferentes razones, entre estas estaban las demográficas y el alto desempleo.

El IPV incorpora al conjunto de personas físicas, tanto residentes como no residentes en España, que hayan comprado u obtenido una vivienda en el periodo de referencia en el territorio nacional. Cabe destacar que las compraventas efectuadas por personas jurídicas no forman parte del ámbito poblacional.

En marzo de 2020 aparece a nivel mundial la pandemia de COVID-19 lo que ha provocado un hundimiento en el precio de venta y dependiendo del sitio también ha afectado al alquiler de las viviendas. Además, los bajos intereses del Euribor siendo cercanos a 0 o en ocasiones en términos negativos, da lugar a que las viviendas sean el activo financiero por excelencia a nivel internacional, lo que provoca un impedimento de bajadas de precios, en el alquiler y venta, debido a que no objetan a la oferta y demanda real del uso.

La depresión de la crisis provocada por la COVID-19 presagia una larga crisis en el mercado inmobiliario y un nuevo ajuste de los precios a la baja tanto en el alquiler como en la compraventa de vivienda y en de locales comerciales.

Según el estudio de Rodríguez, O. (2021, 24 enero). De la crisis del ladrillo a la del COVID-19: el número de inmobiliarias crece un 58%. El Independiente. El Monitor Adecco de Ocupación II¹, explica que desde 2008 hasta 2020, el empleo en actividades inmobiliarias ha crecido un 33,5%. Es la rama donde más ha aumentado el consumo, por delante de la administración pública o las actividades relacionadas con el agua y la energía.

Además del Monitor de Ocupación II, según Alcover cree que el inmobiliario se ha mantenido como un valor refugio durante la crisis del coronavirus y confía en que todo el ahorro precautorio que han generado las familias tenga impacto en la reactivación de la compraventa de viviendas.

¹ Monitor Adecco de Ocupación II es la persona encargada de hacer el análisis de la creación de empleo en España desde el perfil demográfico de los nuevos ocupados en España.

Por último, según el Instituto Nacional de Estadística (INE) destaca que el número de empresas que se dedican a la actividad de los inmuebles crece en un 58% entre la crisis del 2008 y el año 2020. Además, ha aumentado el número de locales donde se realizan actividades inmobiliarias, casi en un 62%.

2.Objetivos

2.1 Objetivo principal.

El objetivo principal es la realización de un estudio estadístico y econométrico para conocer cuáles son los factores que afectan al Índice de Precios de las Viviendas en la Comunidad Valenciana. Para este estudio se escogerán después de un análisis del estado del arte, las variables que se estudiarán durante todo el trabajo.

2.2 Objetivos secundarios

Se ha establecido un objetivo secundario, basado en analizar las diferencias del IPV que pueden existir entre las diferentes Comunidades Autónomas de España.

Las diferencias del IPV que pueden existir entre las diferentes Comunidades Autónomas de España.

3.Metodología.

En primer lugar, se realiza la elaboración de antecedentes. Mediante la realización de una extensa revisión bibliográfica en líneas de trabajo relacionadas.

En segundo lugar, se realiza el establecimiento de los objetivos y se explicita la metodología de trabajo a utilizar.

En tercer lugar, se definirán las hipótesis y el planteamiento de las variables. Una vez las variables estén definidas, se realizará una búsqueda de los datos pertinentes para el estudio. Posteriormente se tratará la información, se analizarán los datos existentes mediante diferentes herramientas estadísticas y se tratará de obtener conclusiones sobre la evolución del IPV en la Comunidad Valenciana.

La sistemática para seguir para analizar los datos del presente trabajo es la siguiente:

- Realización de un estudio univariante con enfoque cuantitativo y cualitativo de los datos experimentales obtenidos para estudiar todas las variables de manera independiente. Para ello, se utilizarán las herramientas Statgraphcis versión 18 y la herramienta R Studio....
- Después, se realizará un estudio bivariante con enfoque cuantitativo y cualitativo de los datos experimentales obtenidos para observar qué relación tiene la variable endógena (Y) con cada una de las variables explicativas (X).

- Posteriormente, se realiza un análisis de series temporales, con el objetivo de predecir valores futuros u observar la variable, mediante el estudio de comportamientos pasados.
- Asimismo, se hará un análisis de la multicolinealidad para observar la relación de dependencia entre ciertas variables explicativas y eliminar alguna de estas debido a la fuerte relación que pueden llegar a tener las variables entre sí.
- Para finalizar, se va a realizar un modelo de regresión lineal, en el caso de que el análisis de erróneo cuando se concluya se pasará a realizar el modelo de regresión lineal penalizado con el programa R Studio, y si procede un modelo ARIMAX penalizado, junto con la predicción correspondiente. El modelo que se va a usar para el modelo penalizado será el método del Elástico Net. Este modelo comprende dos tipos de modelos diferentes el modelo Lasso y el modelo Ridge que se explicarán en el apartado correspondiente.

4.-Hipótesis

Se plantean X hipótesis iniciales, las cuales se tratarán de validar a lo largo del trabajo.

Hipótesis 1:

El IPV de la Comunidad Valenciana será inferior al de las comunidades de Cataluña y Madrid, pero cercano a estas, debido a que es una comunidad muy importante en España.

Hipótesis 2:

Se prevé que las variables que afectarán de forma más significativa en las variaciones del IPV, son los M^2 de los que dispone la vivienda, el nivel de salarios de las personas y el número de desempleados.

Hipótesis 3:

Se prevé la variable Euribor no tenga una relación directa con el IPV, debido a que en los últimos años se ha mantenido en valores cercanos a 0 o incluso negativos y se espera que no tenga una influencia significativa en el IPV.

5.-Revisión bibliográfica

En este apartado se realiza un análisis bibliográfico sobre los trabajos, artículos científicos, etc. relacionados con los precios de las viviendas o con otros aspectos relacionados con las viviendas. El objetivo de realizar este estudio es analizar los trabajos existentes y obtener información sobre las variables que otros autores han considerado importantes en cuanto a su afectación en el IPV. Para facilitar la comprensión del presente apartado, se facilita un resumen de los diferentes artículos analizados, así como se muestran las principales aportaciones relacionadas con el presente trabajo de cada uno de dichos artículos.

5.1. Análisis de los factores determinantes del precio del activo vivienda

En primer lugar, se analiza un artículo de Peña Cerezo, M. Á., Ruiz Herrán, V., & García Merino, J. D. (2004). Análisis de los factores determinantes del precio del activo vivienda. En este artículo, se diferencian tres tipos de factores, los cuales se van desarrollando para comprobar cuál de los tres tiene una mayor importancia y pueden llegar a ser buenos para la determinación del precio de la vivienda.

Estos factores se diferencian en 2 tipos:

- Factores determinantes de la demanda:

Por un lado, se puede encontrar los de largo plazo, como pueden ser los factores demográficos, estos son la población total, población perteneciente a un determinado corte de edad (entre 24 y 34 años, por ejemplo), índice de nupcialidad, volumen de divorcios, número de personas por vivienda. Estas variables pueden determinar la demanda potencial de la vivienda.

Por otro lado, existen los de corto plazo, estos pueden ser los factores económicos, como por ejemplo el nivel de renta per cápita/unidad familiar; los factores financieros como son los tipos de interés del mercado hipotecario, volumen disponible de crédito hipotecario, el ratio préstamo/valor de la vivienda, vencimiento del préstamo, etc. y los factores fiscales como por ejemplo minoraciones en el impuesto derivadas del pago de intereses y de la inversión en vivienda.

- Factores determinantes de la oferta:

* Disponibilidad de crédito: cuanto mayor sea el ahorro, mayor será la cantidad de dinero utilizable para la inversión en construcción.

* Variables que afectan a la demanda: población, renta, stock de viviendas existentes, etc.

* Costes de construcción: ésta puede ser la variable que más fuertemente influye en la fijación del precio, desde el punto de vista de la oferta constituyendo el límite inferior al precio de la vivienda.

* En general, la oferta se va a basar en la demanda por lo que casi todos los factores que afectan a la demanda también influyen en la oferta.

Una vez explicados los factores este autor procede a definir los dos tipos de variables (las variables endógenas y las exógenas) que se van a utilizar, estas son las siguientes:

Variables exógenas, son las variables que más relación tienen con la variable de estudio que se va a tratar, en este trabajo era la variable del índice de precios de la vivienda.

1. Incremento (general) de los precios al consumo excluyendo la vivienda en propiedad.
2. Incremento de los precios de los alquileres.
3. Incremento en el Producto Interior Bruto (PIB) per cápita.
4. Factores demográficos: población comprendida entre los 25 y los 34 años.
5. Interés de referencia del mercado hipotecario.
6. Volumen de crédito disponible.
7. Número de viviendas terminadas en un determinado periodo.
8. Coste de construcción.

Estas son las variables que el autor ha considerado que son las que más relación tiene con su variable endógena el índice de precios de las viviendas.

Por lo tanto, este autor ha decidido usar estos factores para el análisis del precio del activo de la vivienda.

5.2 Análisis del precio de la vivienda en España

En segundo lugar, se analiza el trabajo realizado por Martínez Pagés, J., & Maza Lasierra, L. A. (2003). Análisis del precio de la vivienda en España. *Documentos de trabajo/Banco de España, 0307* en el cual observan las variables que están adquiriendo y cuáles son los factores determinantes que afectan a las viviendas.

Según el documento diferencia varios tipos de factores.

Primeramente, habla de la vivienda, y comenta que es un componente fundamental de la riqueza de las familias y que, por tanto, se considera uno de los factores más importantes debido a que contribuyen a explicar sus decisiones de gasto. Además, existen unos factores diferenciales que son los más obvios en las viviendas como es el caso del tamaño, mediante la medición bien del precio medio por metro cuadrado de la vivienda, o bien, del precio medio de las viviendas de un determinado tamaño (número de dormitorios) y estructura, pero que esta relación no es directa al precio debido a que también existen factores de calidad que hacen que aumente o disminuya el precio de las viviendas.

Existen también diferentes factores relacionados con el turismo y con el peso de las grandes ciudades y en la decisión de compra de una vivienda para uso propio influyen de manera inseparable factores de demanda de inversión y de consumo.

Factores regulatorios como por ejemplo la comparación entre la evolución del precio de la vivienda en el Reino Unido y en España. Además de tener en cuenta factores como el efecto riqueza o los costes de ajustes que también tienen una relación fuerte con los factores regulatorios.

Entre los primeros, cabe destacar a los factores demográficos y los relacionados con la oferta de financiación a la vivienda, no tiene en cuenta cambios en la distribución de la población entre los distintos grupos de edad y, en particular, en relación con los tramos de edad en los que normalmente se constituyen nuevos hogares (entre los 20 y los 35 años), el volumen de inmigración procedente de otros países y los desplazamientos migratorios en general (por

ejemplo, del campo a la ciudad), así como otros factores que afectan a la formación de hogares como la tendencia a la formación de hogares unipersonales, el retraso en la edad de matrimonio.

En este documento cuenta con 18 variables que son aptas de haber contribuido a determinar la evolución del precio de las viviendas en España en el año 2003, todas estas se van a nombrar a continuación.

1. Renta considerada es la renta bruta disponible de las familias (la renta disponible por hogar, y tiene la ventaja sobre esta última de ser exógena, puesto que la población lo es, mientras que el número de hogares no)
2. Precio real de la vivienda
3. Renta bruta disponible real por habitante de más de 25 años ajustada del efecto de la devolución de impuestos en el año 1993.
4. PIB real por habitante de más de 25 años. Fuente: INE
5. Número de ocupados de la Contabilidad Nacional
6. Coste de uso de la vivienda, definido como el tipo de interés nominal (I) menos la tasa de variación esperada del precio de la vivienda.
7. Tipo de interés de los préstamos bancarios a los hogares para adquisición de vivienda.
8. Tipo de interés real, definido como el tipo de interés nominal menos una media móvil centrada de cinco años de la inflación.
9. Rentabilidad real de la Bolsa, calculada como la tasa de variación del índice general de la Bolsa de Madrid, deflactado por el IPC
10. Número de viviendas en oferta. Se calcula como el valor estimado del stock de viviendas terminadas más el número de viviendas iniciadas en los últimos 18 meses.
11. Número de habitantes de más de 25 años
12. Número de viviendas en oferta por habitante de más de 25 años
13. Porcentaje de la población entre 25 y 29 años sobre la población de más de 25 años
14. Riqueza financiera neta real por habitante de más de 25 años.
15. Riqueza inmobiliaria real por habitante de más de 25 años.
16. Deuda hipotecaria real (deflactada por el IPC) por habitante de más de 25 años.
17. Número de viviendas protegidas iniciadas en el año por cada 1000 hogares.
18. Índice de costes reales de la construcción.

Estas son las variables que el autor ha escogido debido a que tras un estudio ha considerado que son las que más relación tienen con el estudio del análisis de precios de la vivienda en España.

5.3 Análisis regional del mercado laboral y de la inflación

En tercer lugar, se analiza el artículo de Navarro, M. L., & Cisneros, C. U. C. (2006). Factores explicativos del precio de la vivienda en España. *Análisis regional del mercado laboral y de la inflación*, (7), 141-149 en el cual se diferencian dos tipos de factores que mayor importancia pueden tener en la determinación del precio de la vivienda en España.

- Factores explicativos del precio de la vivienda referidos a la oferta
 - * El número de viviendas terminadas de nueva construcción es un indicador de la evolución de la oferta de viviendas.
 - * El suelo y el precio del m²
 - * La evolución de los costes de construcción
- Factores explicativos del precio de la vivienda referentes a la demanda
 - * Factores demográficos como la observación de la evolución de la población entre 20 y 34 años
 - * La población extranjera, tanto con rentas bajas como con rentas altas
 - * El empleo debido a que los individuos demandan viviendas si tienen ingresos.
 - * La Tasa de paro, la falta de trabajo lleva consigo la falta de ingresos y por consiguiente afecta a la demanda.
 - * La Renta Nacional Real y el Producto Interior Bruto Real
 - * Los tipos de interés del mercado hipotecario, que se ve representado por los costes de financiación de las viviendas.
 - * La cotización en bolsa, como son los activos bursátiles como el IBEX 35, ya que se cree que, ante el aumento del riesgo y la reducción de la rentabilidad, gran parte del capital bursátil se refugia en la vivienda.
 - * La deuda pública lo que provoca con una rentabilidad cada vez menor o un riesgo mayor, ha hecho que la vivienda tenga un mayor interés por los inversores.

5.4 El precio de la vivienda y la inflación en España.

En cuarto lugar, se analiza las variables utilizadas en el trabajo realizado por el autor Bellod Redondo, José Francisco. (2009). El precio de la vivienda y la inflación en España. El trimestre económico, 76(302), 379-405. Epub 20 de noviembre de 2020.

Este trabajo realiza un estudio compuesto por 9 variables:

1. Tasa de interés hipotecaria
2. Subíndice de precios de “alquiler de vivienda”, IPC e IPCA
3. Índice de precios de la vivienda
4. Índice de precios al consumo armonizado
5. Ponderación oficial del gasto en las clases de bienes sobre la tasa de interés hipotecaria.
6. Ponderación oficial del gasto en vivienda
7. Precio de la vivienda/renta bruta por hogar
8. Proporción de viviendas principales en régimen de arrendamiento
9. Proporción de viviendas principales en régimen distinto del arrendamiento, fundamentalmente en propiedad.

Los factores que usa este trabajo están sacados de ciertos documentos para llegar a las variables como por ejemplo factores económicos como son, la renta disponible del hogar, la proporción de renta derivada de transferencias públicas o privadas, los precios tanto de compra como de alquiler de las viviendas, y la estabilidad laboral, además junto a los factores sociodemográficos son los que mayor influencia generan en las decisiones de formación de hogar.

Y por otro lado otros factores como son diferencias en gustos como son el sexo, la edad y el nivel de estudios del sustentador principal, el número de miembros del hogar, el tamaño del municipio de residencia y si en el hogar convive o no una pareja. Después de las estimaciones indicaban que los factores económicos son determinantes en todas las elecciones.

5.5 Precios de la vivienda: sobrevaloración y burbuja.

En quinto lugar, se procede hablar del documento realizado por el autor Sánchez, J. A. (2013). Precios de la vivienda: sobrevaloración y burbuja. *Universidad Complutense de Madrid.*, en el documento trata también de los factores y variables que ha utilizado

Los factores que afecta al aumento de la demanda y que ha usado son los siguientes, factores macroeconómicos, demográficos y financieros.

En el ámbito macroeconómico usa las variables de la renta nacional disponible y de tasa de desempleo.

En el ámbito demográfico usa las variables de población mayor de 25 años (var. Interanual) y población extranjera.

En el ámbito financiero utiliza la variable de tipo de interés nominal de las hipotecas.

Cabe destacar que existe una burbuja intrínseca que, a diferencia de otras burbujas racionales, esta no obedece a los factores exógenos ajenos a las variables que determinan el valor teórico, sino que es provocada por la sobre-reacción de los agentes a la variación de los fundamentales.

5.6 Análisis y pronóstico del precio de la vivienda en España: modelo econométrico desde una perspectiva conductual

En sexto lugar, se analiza el artículo de revista cuyo autor y título es: Soriano Llobera, J. M., Gras Alomà, R., & Roig Hernando, J. (2015). Análisis y pronóstico del precio de la vivienda en España: modelo econométrico desde una perspectiva conductual. *Revista de estudios empresariales*, (1), 145-166. En este trabajo se va a tratar de analizar y estudiar los factores y las variables que utiliza para dicho análisis.

Este artículo usa un planteamiento intrínseco cuya variable de referencia es PVN también conocido como precio promedio del metro cuadrado de vivienda nueva, y para el enfoque extrínseco se han analizado numerosas variables independientes que veremos más adelante.

Cabe destacar que no solo las variables globales macroeconómicas explican los retornos inmobiliarios, sino que variables locales como son los ciclos de economía local, causas demográficas condiciones financieras, políticas bancarias, razones geográficas, medidas regulatorias. La variable clave es la nombrada anteriormente PVN pero la usada por la mayoría de investigadores corresponde al PIB, pero la variable que establecen varios documentos de investigación como es Malpezzi y Wachter (2005) establece de variable el precio de la vivienda.

La base de datos creada integra 379 variables, entre macroeconómicas, regionales, demográficas, conductuales-psicológicas y específicas del sector inmobiliario. Las variables explicativas que esta revista científica identifica relacionadas con el precio de la vivienda nueva son la ocupación en España (EPA Ocupados), el número de Viviendas Iniciadas Total, el ratio Precio de la Vivienda / Renta Bruta por Hogar, la Estimación de la Cantidad de Viviendas, el Tipo de los Préstamos Libres para Adquisición de Viviendas, así como los Precios de la Vivienda Deflactados respectivos de los Estados Unidos y de Francia.

Este modelo ha combinado varios tipos de enfoques, entre ellos el extrínseco, intrínseco y conductual, éste último se ha aplicado con el objetivo detectar episodios de Exuberancia Irracional, es decir, para detectar las fases del ciclo económico donde el precio dista significativamente del valor fundamental, debido a causas psicológicas de tipo eufórico o depresivo.

5.7 Factores determinantes de la diferencia de precios en el mercado inmobiliario

En séptimo lugar se habla del trabajo nombrado realizado por Jiménez López, F. (2018). Factores determinantes de la diferencia de precios en el mercado inmobiliario. Caso aplicado a los municipios de Cataluña.

Existen muchas variables que han usado en este trabajo, en un primer lugar cabe destacar las principales variables que se van a tener en cuenta, estas son, la evolución del precio por metro cuadrado, el número de viviendas de obra nueva, el número de transacciones de compraventa, las condiciones de financiación y las políticas de viviendas que realizan las diferentes administraciones desde el punto de vista del número de viviendas de protección oficial iniciadas.

Además de todas estas variables principales el autor del trabajo comienza usando un trabajo de Aguayo y Expósito, López 1998 y usa estas variables con su respectiva explicación:

1. Precio por metro cuadrado. En este se mide el precio real de la vivienda por metro cuadrado en cada región.
2. Tipo de interés de los préstamos hipotecarios. Esta variable se multiplica por 0,8. Esto es debido a que los préstamos hipotecarios solo financian el 80% del coste de adquisición. Teniendo que hacer frente al resto el comprador con fondos propios.
3. Tipo de interés de los depósitos a plazo. Esta variable se multiplica por 0,2. Como consecuencia que el 20% del valor de compra de la vivienda, lo debe aportar el comprador.
4. Renta disponible de los hogares. Esta variable se mide a nivel regional en miles de millones de pesetas.
5. Stock de capital residencia. En miles de millones de pesetas y relativizado por el número de población.
6. Población entre 20 y 34 años Se ha escogido la población en esta franja de edad, con el objetivo de captar la posible demanda potencial de cada región.

De estas 6 variables la principal es la renta disponible per cápita, a continuación, el stock de capital residencial y finalmente los costes de financiación, por el contrario, la población de edad comprendida entre 20 y 34 años no es significativa y por tanto se eliminó del modelo.

Además, usa otro trabajo elaborado por Vicente Royuela, Esther Vayà y Jordi Suriñachi en el año 2002 en el que se centra más para su estudio debido a que es un trabajo realizado por Grupo de Análisis Cuantitativa Regional de la universidad de Barcelona.

Las variables que se revelaron como factores explicativos en este trabajo son las siguientes:

1. Distancia a Barcelona, que es la medida en kilómetros de carretera.
2. Dimensión del municipio. En este caso se utiliza el número de habitantes de cada municipio. El objetivo es captar el grado de información de los compradores, así como captar el tamaño del mercado. Si el municipio es más grande, los posibles compradores tienen más información debido a que pueden comparar entre diferentes precios.

3. Renta de las familias utilizando las medidas según la Renta Familiar Bruta Disponible.
4. Oferta de suelo que mide la medida en hectáreas rústicas disponibles.
5. Ratio de personas por hogar.
6. Menú fiscal de los municipios cuya medida es según el precio medio del recibo del IBI.

Una vez visto las variables de estos dos trabajos, el autor se centra a realizar sus variables y sus respectivas descripciones para su modelo.

Descripción de las variables:

1. Precio por metro cuadrado en municipios de más de 5.000 habitantes
2. IRPF Base Imponible General por declarante y PIB per cápita.
3. Porcentaje de Población entre 25 y 34 años
4. Porcentaje de parados sobre la población potencialmente activa
5. Densidad de Población
6. Población Total
7. ATM de Barcelona*Calidad del Transporte Público
8. Capital de Comarca
9. Población de más de 50.000 habitantes
10. Crecimiento de la población
11. Porcentaje de Superficie Urbanizable
12. Vivienda Iniciada por cada 1.000 habitantes
13. Vivienda Acabada por cada 1.000 habitantes
14. Número compraventas por cada 1000 habitantes
15. Lleida
16. Barcelona
17. Costa
18. AMB
19. Plazas turísticas por cada 10.000 habitante
20. Cuota Íntegra por recibo del IBI
21. Vivienda de Protección Oficial Iniciada por cada 1.000 habitantes
22. Corona del ATM
23. Distancia a Barcelona en línea recta
24. Tiempo de desplazamiento a Barcelona en coche
25. Tiempo de desplazamiento a Barcelona en Transporte Público
26. IPC y Euribor

Dentro de estas variables una vez realizado el estudio se considerarán más relevantes, el nivel de renta (medido a través de la Base Imponible General del IRPF), el número de viviendas iniciadas por cada mil habitantes, la variable ficticia AMB y Costa, el número de plazas turísticas de cada municipio, el tiempo que se tarda en llegar a la ciudad de Barcelona en transporte público, la Corona del ATM en la que se encuentra el municipio y por último el porcentaje de población con una edad comprendida entre 25 y 35 años.

En cambio, existen otras variables que no son relevantes, por ejemplo, el número de viviendas de protección oficial iniciadas, el porcentaje de superficie urbanizable, la cuota íntegra por recibo del IBI, el número de viviendas acabadas y el número de compraventa por cada mil habitantes, la variable que mide la calidad del transporte público en función de la corona del ATM en la que se encuentra el municipio y la distancia a Barcelona en línea recta.

5.8 Evolución de precios en el mercado inmobiliario (vivienda).

En octavo lugar, se analiza el trabajo de López Serrano, A. (2019). Evolución de precios en el mercado inmobiliario (vivienda). Este trabajo entra en el análisis en variables o factores y saber si van a ser explicativas o no en la evolución de los precios de la vivienda.

El primer factor determinante será la variable demográfica, esta se demostrará más adelante que según la evolución en la historia y en la economía, tenderá a ser un factor que cada vez tendrá menos peso en la demanda de viviendas. Esta deja de ser una variable relevante en cuanto a la explicación del incremento en la construcción de nuevas viviendas. Una variable que aparece por el crecimiento demográfico podría ser el espacio físico.

La variable renta es el segundo factor importante, considerando esta como el nivel de ingresos, procedentes del trabajo o la renta permanente.

La tercera variable serán los precios y la inflación, que pueden medirse mediante índices de mercado ya disponibles o bien a partir de indicadores construidos, como es el caso de los precios hedónicos.

El factor financiero es otra variable relevante para determinar la demanda de los bienes residenciales y es, si cabe, uno de los más importantes. Como es el caso de la disponibilidad de financiación que es una de las variables más importantes de este factor

Otro factor que se encuentran los factores subjetivos que determinan el beneficio de la propiedad sobre un tipo de vivienda u otra. Son variables como es el caso del coste de mantener la vivienda o el rendimiento que podría generar esta adquisición.

Y este autor pasa a estudiar las siguientes hipótesis:

1. Renta per cápita
2. Crecimiento demográfico
3. Crecimiento de la población inmigrante
4. Soporte financiero
5. Tipos de interés

6. Coyuntura económica
7. Compraventas de viviendas y llegada de turistas a Baleares
8. Precio alquiler de la vivienda
9. Accesibilidad de la vivienda

Para la renta per cápita los expertos y junto al estudio del autor del trabajo han determinado que es la variable o factor que más peso tiene en la determinación del precio del bien inmobiliario, tanto para el incremento como para su disminución.

En relación con el crecimiento demográfico, muchos de los autores los excluyen, sin embargo, tras su estudio se piensa que tiene una correlación alta y se concluye que si añades el crecimiento de la población inmigrante puede seguir siendo una variable importante para determinar el precio.

Para este autor, que debe elegir las cuatro variables que más relación tienen con su estudio escogería la renta per cápita, el crecimiento de la población extranjera, la coyuntura económica y el plazo medio de hipotecas concedidas debido a su análisis.

En el soporte financiero y la compraventa de vivienda en Baleares o el precio de alquiler de la vivienda como producto sustituto de la compra de vivienda, son variables que se pueden asociar también a las variaciones de precios medios de las viviendas.

Sin embargo, el tipo de interés no tiene una buena correlación y esto es debido a que el tipo de interés es fijado por el Banco Central Europeo, por lo que no se fija en la economía del país, en este caso España. Es decir, que, si existe una época de “boom inmobiliario”, puede que se encuentre una bajada masiva del tipo de interés en un año, y en el futuro tener un aumento del tipo de interés o por el contrario cuando se trata de una recesión se produce el efecto contrario, es decir una aumento del tipo de interés que irá disminuyendo con el paso del tiempo, por lo que el autor detalla que puede que se trate de una variable demasiado volátil, ya que es una variable externa del país que se está analizando en este caso España y comparando con variables que si son a nivel nacional.

Para finalizar la última variable que analiza es el ratio de préstamo/valor, este ratio lo que trata de comprender es cuanta parte del préstamo es financiado por los bancos según el valor que cueste el inmobiliario. Y resulta que el autor concluye que la correlación es demasiado baja, ya que podría deberse a la política de los bancos.

5.9 Determinantes del precio de viviendas en la región metropolitana de Chile

En el noveno lugar, se analiza el trabajo de Sagner T., Andrés. (2011). Determinantes del precio de viviendas en la región metropolitana de Chile. El trimestre económico, 78(312), 813-839.

Este trabajo es un tanto peculiar debido que está centrado en un país diferente a España, para analizar si todos los estudios enfocan el planteamiento de la misma manera, para ello en este trabajo se va a analizar las variables que emplea y el estudio que ha realizado de cada una de las variables, cuyo título es: Determinantes del precio de viviendas en la región metropolitana de Chile.

Para empezar a analizar las variables, se observa que el autor como muchos otros anteriormente analizados crea un tipo de variables ficticias, estas reciben un nombre específico “Dummy”. En específico para este trabajo el autor decide crear 36 variables “Dummy”

A pesar de las variables ficticias el autor analiza estas variables independientes son:

1. Precio de compra vivienda
2. Antigüedad
3. Superficie construida
4. Superficie terreno
5. Subsidio de renovación urbana
6. Distancia estación metro
7. Distancia área verde
8. Distancia a clínica
9. Distancia a hospital
10. Distancia a colegio
11. Ingresos del hogar
12. Crecimiento del PIB
13. Interés de crédito hipotecario
14. Índice de precios selectivos de acciones (IPSA)

De acuerdo con el autor, se tienen en cuenta que las variables superficie y antigüedad, ambas variables son determinantes estadísticamente significativas para la formación del precio de la propiedad.

Respecto a las variables macro financieras el resultado significativo es el ingreso real del hogar, debido a que un aumento en el ingreso del hogar provocaría un aumento en el precio de la propiedad. Por el contrario, el interés y el índice de precios de acciones no son estadísticamente significativos.

Otro de los valores que se consideran significativos son las variables de entorno, en concreto la variables de adquisición de viviendas en comunas y la variable barrios de mayores ingresos.

Se observa como el autor detalla cómo más importante y significativo el ingreso por hogar, debido a que es la variable con mayor relación con el precio de la vivienda, por esta razón es una variable que se tienen en cuenta para el estudio.

5.10 Un análisis sobre el gasto en servicios de vivienda en España

Por último, se analiza el trabajo de Colom, M. C., & Molés, M. C. (1999). Un análisis sobre el gasto en servicios de vivienda en España. Instituto Valenciano de Investigaciones Económicas. En el cual el trabajo se centra en analizar unas cuantas variables o factores que utiliza para el análisis del gasto.

Las variables que está usando este autor son las siguientes:

En primer lugar, el autor elige la variable del sexo como variable cualitativa es decir esta variable va a indicar el sexo del sustentador principal. Es una variable categórica que toma el valor 1 cuando el sustentador principal es varón y el valor 0 cuando es mujer, siendo esta última la categoría de referencia.

Como segunda variable el autor ha escogido la variable estudios, debido a que esta indica el nivel de estudios del sustentador principal. Es una variable categórica puede llegar a tener tres valores: nivel de estudios primarios, nivel de estudios secundarios y nivel de estudios universitarios.

La variable Estudio, al tener tres valores para diferenciar los tres valores de las variables, se introducen mediante las variables ficticias ESTUDIO1, que toma el valor 1 si los estudios que tiene el sustentador principal son como máximo primarios y el valor 0 si posee otro tipo de estudios ; ESTUDIO2, que toma los valores 1 y 0 para representar un nivel de estudios secundarios y otro tipo de estudios, respectivamente y ESTUDIO3, que identifica la categoría de estudios universitarios, pero que no se introduce en el modelo por considerarla como la categoría de referencia.

La tercera variable que el autor ha elegido es la edad, esta será la edad del sustentador principal, introducida en forma cuadrática. En análisis previos se ha ensayado con diferentes especificaciones para esta variable, obteniéndose que la especificación cuadrática era la más adecuada.

Como cuarta variable, nos encontramos con la profesión, esta indica la categoría socio-profesional del sustentador principal. Es una variable categórica que toma seis valores:

1. Trabajadores manuales de industria y servicios,
2. Trabajadores no manuales de industria y servicios
3. Autónomos de industria y servicios
4. Trabajadores de la agricultura
5. Jubilados
6. Otro tipo de profesiones.

En el análisis se han considerado cinco variables ficticias (PROFESIÓN1, PROFESIÓN2, PROFESIÓN3, PROFESIÓN4 y PROFESIÓN5) que indican las primeras categorías. La última categoría (otros tipos de profesión) se ha dejado como la categoría de referencia.

Como quinta variable, el autor utiliza la variable MIEMHOG, esta variable representa el número de miembros del hogar (incluyendo al sustentador principal).

Como sexta variable, está la variable MIGRA, la cual indica si el hogar ha cambiado su municipio de residencia después del 1 de abril de 1986. Sus dos valores son: 0 si el hogar no ha cambiado de municipio, y 1 si ha cambiado.

Como séptima variable, esta es variable, es la renta disponible del hogar. La variable renta se obtiene como el conjunto de ingresos monetarios y no monetarios percibidos por los miembros del hogar perceptores de ingresos, cualquiera que sea su naturaleza. La renta disponible es el

valor que resulta una vez deducidas las cantidades satisfechas en concepto de impuestos, cotizaciones a la Seguridad Social y otros pagos asimilados.

En este trabajo se ha considerado la renta disponible como una proxy de la renta permanente, puesto que no se dispone de la suficiente información para obtener esta última.

Por último, la variable que escoge es el precio, en concreto el precio de las viviendas. El precio de la vivienda no es homogéneo para todos los hogares, ya que depende de la localización geográfica de la vivienda. Se pueden observar unas grandes diferencias en el precio entre comunidades autónomas y entre el ámbito de residencia habitual del hogar.

5.11 Conclusión sobre la revisión bibliográfica

En este apartado se va a tratar de explicar que variables son las que más usan los compañeros para el estudio del Trabajo.

1. La variable por excelencia es Renta, muchos de los trabajos mencionan esta variable como una variable fundamental.
2. Otra de las variables fundamentales es el PIV.
3. Otra de las variables que se centran en el trabajo son superficie construida.
4. Además de las dos anteriores otra de las variables que usan en muchos de los trabajos son Tipo de interés
5. Otra variable que destaca mucho es el IPC
6. Esta destaca debido al sentido que le da el autor, esta variable es la de Viviendas Iniciadas
7. Por último, la variable de la que también se habla es de la tasa de paro

Guiando así a otros trabajos académicos en un futuro, y sirviendo de ayuda para la decisión de la toma de las variables escogidas para este trabajo.

6.-Determinación de las variables sobre las que se va a realizar el estudio

En este apartado se definirán las variables que se van a estudiar en el presente trabajo de fin de grado y que se pueden consultar en el **Anexo I**, dichas variables han sido extraídas de las páginas webs oficiales que se detallan en el Anexo I.

La periodicidad de las variables con la que se va a trabajar va a ser trimestral como consecuencia de las características congénitas al mercado de la vivienda. El uso de los datos con un fraccionamiento mayor no va a reportar la información adicional, ya que la demanda de activos residenciales (y su oferta) sigue factores que se desarrollan a corto, medio y largo plazo.

6.1.-Variable endógena o dependiente (Y).

La variable endógena o dependiente que se va a explicar es el índice en el precio de la vivienda (IPV). Los datos del IPV de la Comunidad Valenciana se han extraído de la web <https://pegv.gva.es/es/informacion-estadistica>. El propósito de escoger el IPV como variable independiente es conocer su evolución a lo largo del tiempo y conseguir hacer una predicción para el futuro.

6.2.-Variables explicativas o independientes (X)

Una vez definida la variable endógena (Y), se procede a seleccionar las variables que tienen una relación directa o afectan directamente a la variable IPV, estas variables se denominan variables explicativas o independientes (X).

A raíz del estudio bibliográfico de las variables que se han ido incorporando para el análisis de la vivienda durante la evolución por años, realizada desde el 2001 hasta el 2020, se ha decidido trabajar con las variables que se consideran más relevantes con respecto al índice de precio de las viviendas (IPV) cuya explicación se detalla a continuación.

En primer lugar, después de una larga búsqueda y un largo estudio de los documentos anteriores, la mayoría de los trabajos coinciden en que la variable Renta también conocida como el IRPF, en especial la renta por habitante es una variable imprescindible para el estudio que se quiere realizar del índice de precio de la vivienda, ya que, si la gente o las familias no generan ingresos no pueden permitirse la compra o el alquiler de una vivienda.

En segundo lugar, se ha decidido implementar para el análisis del índice de precio de la vivienda, es la variable Población Comprendida entre los rangos de edad de 25 a 34 años, el problema se encuentra al observar esta variable, ya que algunos de los trabajos que implementan esta variable son estudios y análisis realizados antes de la crisis de 2007, por consiguiente esta variable se encuentra desactualizada. Debido a la situación económica actual de España muchos jóvenes prorrogan la edad de independizarse de sus hogares familiares, es por esta razón que se ha decidido escoger la variable de población entre 30 a 44 años (Pobla30-44).

En tercer lugar, se va a tratar de incorporar la variable el Tipo de Interés, en especial el Euribor. En algunos documentos analizados, se habla del coste de uso de la vivienda relacionado con los tipos de interés, siendo los resultados significativos frente al IPV. Se debe recordar que el Euribor no es un tipo de interés fijado en España, sino que viene marcado por la Unión Europea.

Por otro lado, respecto al tipo de interés, uno de los motivos por los que se ha decidido escoger esta variable, ha sido por la reflexión de Taltavull en 2001, que nombra los tipos de interés como un factor decisivo para la subida o disminución de la demanda de viviendas.

En cuarto lugar, se ha escogido como variable el Número de Parados, esto ha sido debido a que se quiere estudiar si realmente existe una relación que a priori parece existir, entre la población en paro y el precio de las viviendas; si existe mucho paro el precio disminuye. Como se ha ido observando durante los otros documentos, la variable parados ha sido escogida por muchos de los autores comentados anteriormente, por ejemplo, en algunos documentos aparece la

reflexión de que el aumento de la tasa de paro junto a la caída de la renta marcó el punto de inflexión de la evolución del precio de la vivienda.

En quinto lugar, se va a incorporar la variable de Número Viviendas Iniciadas por cada mil habitantes, esta variable trata de indicar tanto los aumentos o las disminuciones por parte de la demanda como también aumentos o disminuciones por parte de la oferta. A corto plazo es un factor principal para determinar el precio de la demanda.

Además, cuando existe un aumento en el Número de Viviendas Iniciadas, los propietarios que tienen las viviendas disponibles para venderlas perciben que la demanda es superior a la oferta, debido a esta percepción hace que se aumente el precio de las viviendas ya disponibles para la venta.

En sexto lugar, se ha elegido la variable de la Densidad de Población, esta variable se ha escogido para el estudio debido a que se necesita captar como está distribuida la población en el territorio, en el presente trabajo en la Comunidad Valenciana, así como también, tener una aproximación del tipo de vivienda de cada municipio, ya que los municipios que tengan una alta densidad de población por kilómetro cuadrado, a su vez deberían tener una mayor proporción de bloques de pisos que de casas. Además, cabe destacar que se quiere elegir esta variable ya que se cree que puede tener una relación directa con el índice de precios de la vivienda.

En séptimo lugar, se ha escogido la variable del Producto Interior Bruto (PIB) per cápita, para este dato se ha tomado a precios corrientes, la variable PIB per cápita junto a otras variables entre ellas la renta o tipos de interés son las que mayor relación tienen en el comportamiento del precio de las viviendas. Además, el PIB per cápita es una manera de medir la riqueza del país, aunque en este caso será para medir la de la Comunidad Valenciana.

En octavo lugar, se ha elegido la variable el Precio por Metro Cuadrado, para este es una variable altamente correlacionada con las variables fundamentales del sector y ha sido utilizada en otros documentos de investigación. Además, el Precio por Metro Cuadrado depende también de la localización en donde se encuentre la vivienda, por lo que existe una relación directa con el precio de la vivienda, por ejemplo, la vivienda en Valencia capital cuesta más que una vivienda en los alrededores debido al factor de la localización. Es por eso por lo que esta variable es una buena variable para usarla como objeto de estudio.

Como última variable, se ha elegido estudiar el Número de Empresas de la CV que están en activo. Esta variable está relacionada directamente con el número de personas que se encuentran trabajando, cuantas más empresas existan en la Comunidad Valenciana más personas necesitaran contratar, consecuentemente habrá una disminución del paro, un aumento en el PIB de la comunidad y un aumento de la Renta de la población.

Los datos y las mediciones son cruciales para conseguir un modelo predictivo fiable, por lo que todos los datos del modelo son datos de libre acceso encontrados en las páginas como el INE, Gobierno de España en el apartado ministerio en actividades y servicios y en la página de la Generalidad Valenciana se trata de unas de las fuentes más importantes para conocer indicadores relacionados con el índice de precio de las viviendas.

Estas variables explicativas se quedan de esta manera:

$X_1 \rightarrow$ Renta. Anexo I (Tabla Renta)

$X_2 \rightarrow$ pobla3044. Anexo I (Tabla Pobra 30-44)

- X₃ → Euribor. Anexo I (Tabla Euribor)
- X₄ → Número Parados. Anexo I (Tabla Parados)
- X₅ → VivInici. Anexo I (Tabla Nº viviendas iniciadas)
- X₆ → DensiPobla. Anexo I (Tabla Densid Poblac.)
- X₇ → PIB. Anexo I (Tabla PIB)
- X₈ → preciom2. Anexo I (Tabla Precio m²)
- X₉ → empresasactivas. Anexo I (Tabla Nº empresas activas)

En la búsqueda de datos de estas variables explicativas, se ha conseguido en su gran mayoría datos que se muestran de forma trimestral o mensual, no obstante, con las variables Renta y Población 30-44 años, dichos datos se encuentran con datos anuales, por esta razón se ha decidido buscar 2 variables que se encuentren directamente relacionadas con las variables Renta y Población, pero que sus datos se muestren de forma trimestral. Las nuevas variables serán la variable Salario y la Variable número de nacimientos.

- X₁₁ → Salarios. Anexo I (Tabla Salario)
- X₂₂ → Nacimientos. Anexo I (Tabla Nacimientos)

7.-Análisis univariable

Se va a realizar un análisis univariable para cada una de las variables explicativas que se han elegido para realizar el estudio. Este análisis sirve para analizar cómo se comportan cada una de las variables de manera individual.

En primer lugar, se considera adecuado presentar una breve explicación teórica de los diferentes conceptos que se utilizan en los próximos apartados del presente trabajo. Para realizar un análisis univariable de todas las variables observadas en el apartado 6, se estudiarán las medidas de posición, medidas de dispersión y medidas de forma.

Entre las medidas de posición se encuentran:

- **Media:** Esta representa muy bien el “centro” de la distribución de datos dando una idea clara acerca de la posición de estos. La media se corresponde con la suma de todos los valores de datos dividido entre el número de datos que incluye la muestra.
- **Moda:** Es el valor de la muestra que se repite con mayor frecuencia.
- **Mediana:** Representa el valor de la variable de posición central en un conjunto de datos ordenados. Es decir, es aquel valor que, ordenando los datos de menor a mayor, deja a la derecha y a la izquierda el mismo número de observaciones.
- **Cuartiles y percentiles:** Este servirá para el análisis en profundidad del Gráfico de Cajas y Bigotes. En el primer cuartil se define como el valor que deja por debajo de él una cuarta parte (25%) de los datos, el segundo cuartil corresponde al valor que deja por debajo de él un 50% de los datos y, por último, el tercer cuartil deja por debajo de él tres cuartas partes (75%) de los datos.

Se define el percentil x como aquel valor que deja por debajo de él un $x\%$ de los datos. Así el percentil 25 corresponde al primer cuartil, el percentil 75 al tercer cuartil y el percentil 50 al segundo cuartil, el cual a su vez coincide con la mediana.

En las medidas de Dispersión se encuentran:

- **Rango:** Diferencia entre el valor máximo y mínimo de los datos contenidos en la muestra.
- **Varianza Muestral:** Junto a la desviación típica muestral, una de las mejores medidas de la variabilidad de los datos de una muestra.

$$\sigma^2 = \sum_{I=1}^n \frac{x_I^2}{N} - \bar{X}^2$$

- **Desviación Típica muestral:** Este se define como la raíz cuadrada de la varianza muestral, presenta la ventaja, respecto a esta última, de medirse en las mismas unidades que la variable analizada.

$$\sigma = \sqrt{\frac{\sum_i^N (X_i - \bar{X})^2}{N}}$$

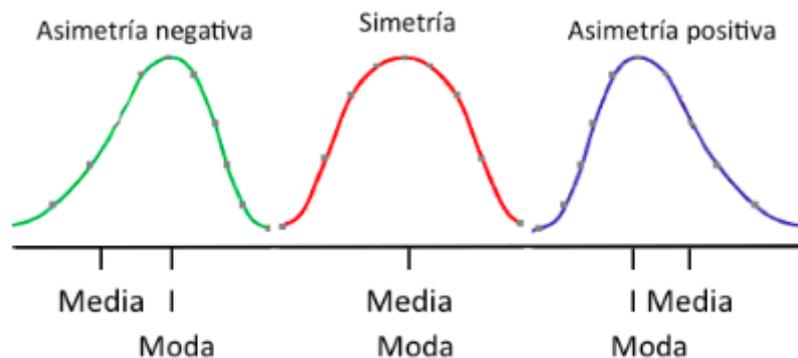
- **Rango Intercuartílico:** Este servirá para el análisis en profundidad del Gráfico de Cajas y Bigotes. Se define como la diferencia entre el tercer y el primer cuartil. A diferencia del rango, es una medida más robusta, ya que evita los valores extremos mínimos y máximos.

En las medidas de Forma:

- **Coefficiente de Asimetría (CA):** Si el CA es igual a cero los datos son simétricos respecto a su media, si CA es positivo los datos presentan asimetría positiva, y si es negativo presentan asimetría negativa.

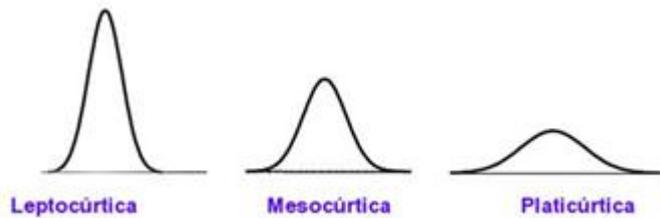
$$CA = \frac{\sum_{i=1}^n (x_i - \bar{x})^3 / (n-1)}{s_{n-1}^3}$$

Donde X es la media aritmética, Mo es la moda y S es la desviación estándar.



- **Coefficiente de apuntamiento o curtosis (CC):** El coeficiente de apuntamiento indica si el porcentaje de datos alrededor de su valor medio es superior, inferior o igual que el de una referencia determinada (distribución normal). Si la distribución es normal el CC tiene un valor igual a 3 y se dirá que es Mesocúrtica, si la distribución es más puntiaguda que una normal el CC es mayor que 3 se llamará Leptocúrtica y si la distribución es más plana que una normal el CC es menor que 3 y se llamará Platicúrtica.

$$CC = \frac{1}{N} \frac{\sum (x_i - \bar{x})^4}{s^4}$$



Posteriormente, se trabajará el Diagrama de Box-Whisker, también conocido como Gráfica de Cajas y Bigotes. Es una representación gráfica muy útil para mostrar visualmente los grupos de datos numéricos a través de sus cuartiles

Este diagrama puede presentar 3 distribuciones diferentes:

1. Asimetría positiva
2. Asimetría negativa
3. Distribución simétrica.

En el Diagrama de Box-Whisker la caja está delimitada por el 1º y 3º cuartil, la línea que separa la caja en dos partes corresponde a la mediana, la media de la muestra se representa mediante un signo más (+) en el interior de la caja. Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango (valores atípicos) está marcado e identificado individualmente

Por último, se tratará de observar el Histograma de frecuencias, este es un resumen gráfico de una tabla de frecuencias unidimensional., es decir, representan frecuencias agrupadas de una variable continua sobre intervalos. La naturaleza gráfica del histograma de frecuencias permite, por un lado, identificar las pautas de comportamiento que son difíciles de observar con un simple resumen numérico, y por otro lado, desarrollar una explicación razonada y relevante de dicha pauta.

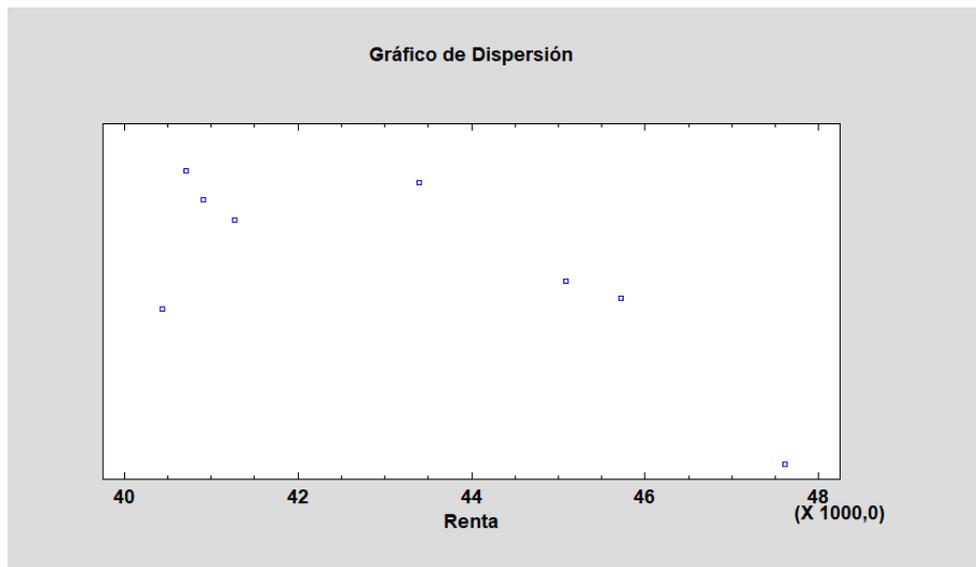
7.1.-X₁: Renta. Anexo I tabla renta

Recuento	8
Promedio	43144,6
Mediana	42334,5
Varianza	7,44921E6
Desviación Estándar	2729,32
Mínimo	40438,0
Máximo	47609,0
Rango	7171,0
Sesgo Estandarizado	0,666402
Curtosis Estandarizada	-0,755407

Tabla 1 Resumen estadístico Renta.

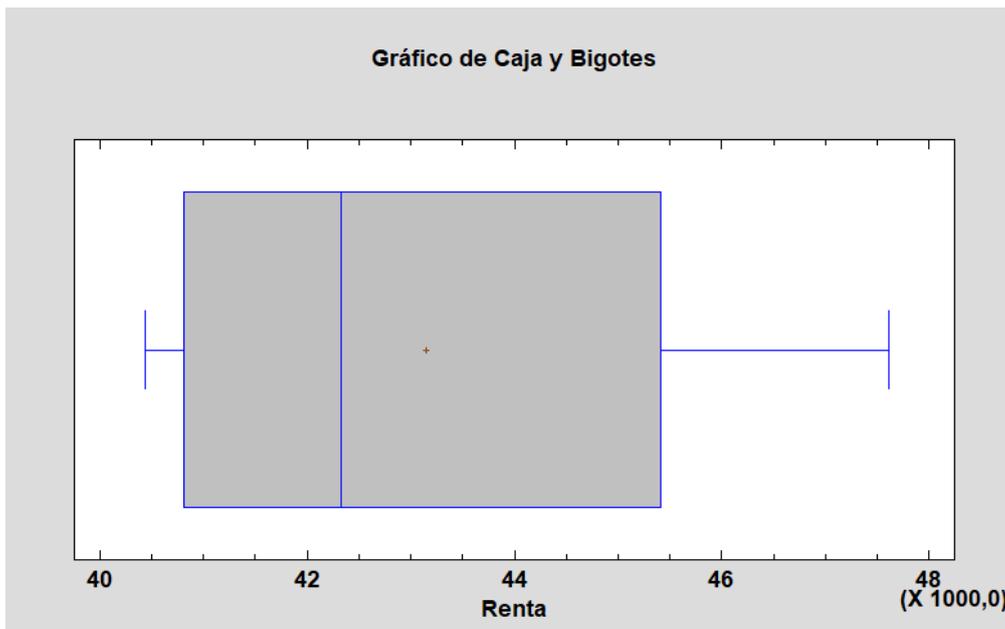
En la tabla 1, se presenta un resumen estadístico de la variable RENTA, en esta tabla se observa que existe una muestra 8 datos, se observa un promedio de los datos introducidos de 43144,6.y observamos que existe una gran varianza por tanto hay diferentes valores de renta que se distribuyen a derecha e izquierda del promedio o media.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3 por tanto es Platicúrtica es decir hay una baja concentración de los valores.



Gráfica 1 Dispersión Renta.

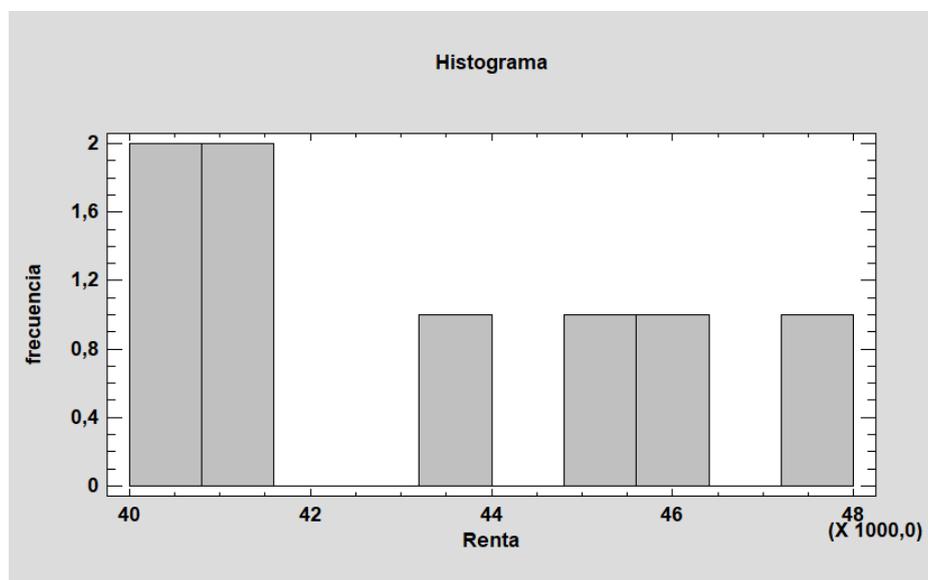
En la gráfica 1, relacionada con la dispersión de la variable Renta se observa que existen muy pocos datos y que están muy dispersos. Este hecho es lógico, ya que en el resumen estadístico se observa que la varianza es muy elevada, hecho que da lugar a una dispersión elevada.



Gráfica 2 Caja y Bigotes Renta

En la gráfica 2, en la caja y bigotes de la variable Renta, se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 3 Histograma Renta

En la gráfica 3, el histograma sobre la variable Renta, se observa que es un histograma asimétrico hacia la derecha, debido que las primeras barras son máximas y luego se mantiene en las siguientes. Aplicando la teoría al ser una variable con tan poco recuentos no es muy fiable la información que detalle el Histograma.

7.2.-X₁₁: Salarios. Anexo I tabla salarios

Recuento	52
Promedio	1679,65
Mediana	1674,76
Varianza	6031,39
Desviación Estándar	77,662
Mínimo	1529,71
Máximo	1823,65
Rango	293,94
Sesgo Estandarizado	-0,0705964
Curtosis Estandarizada	-1,49495

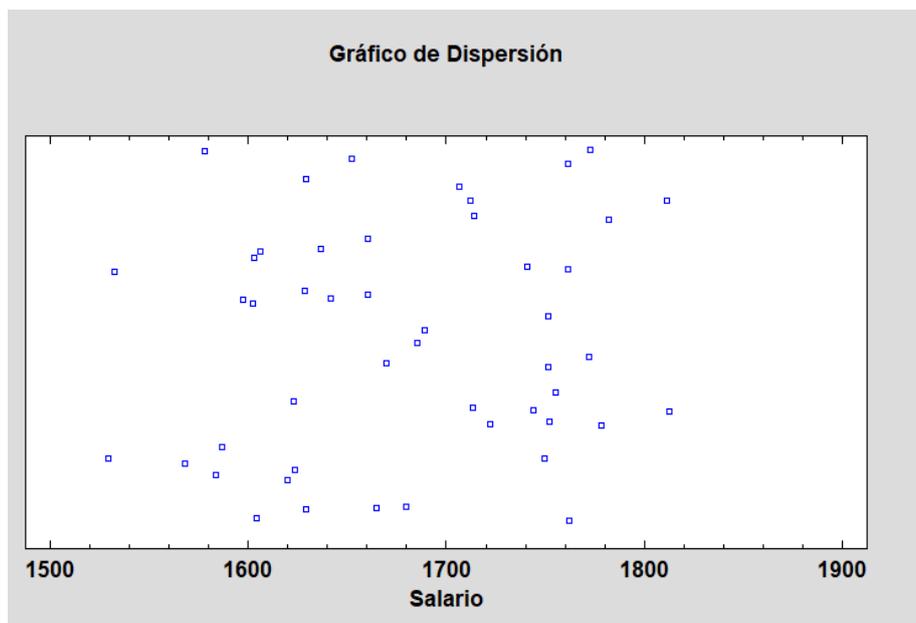
Tabla 2 Resumen Estadístico para Salario.

En la tabla 2, se presenta un resumen estadístico de la variable Salarios, en esta tabla se observa que existe una muestra 52 datos y que la media de estos datos es de 1.679,65.

En cuanto a la varianza se observa que, al ser un valor muy elevado, indicará que en el análisis de dispersión la muestra va a salir muy dispersa, debido a que el valor es muy alto.

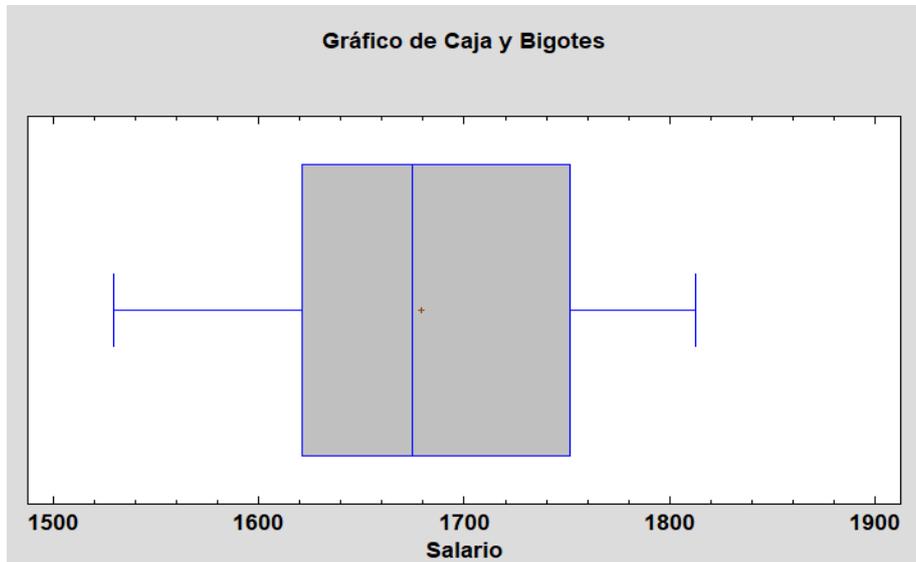
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es superior a la mediana aparecerá a la derecha en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es Platicúrtica es decir hay una baja concentración de los valores.



Gráfica 4 Dispersión Salario.

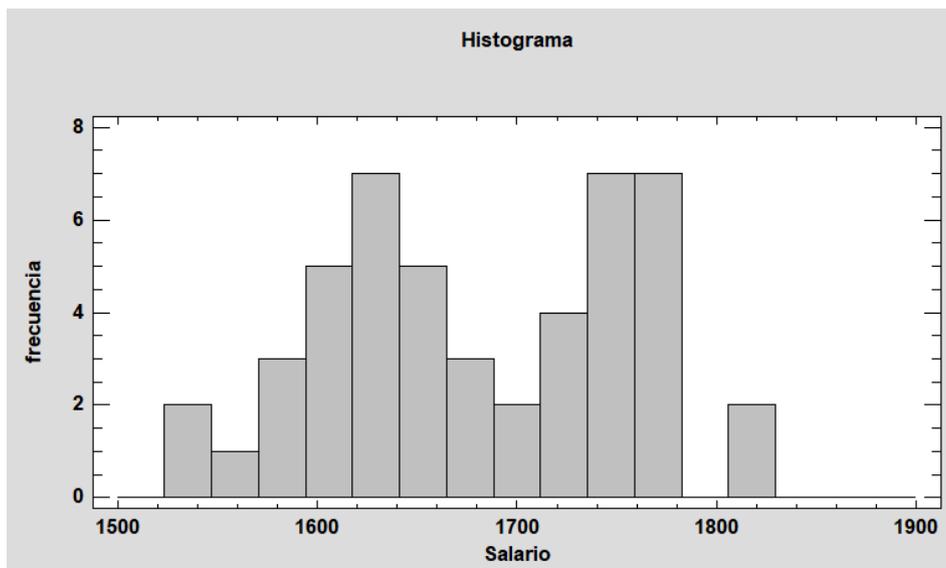
En la gráfica 4, se observa la dispersión de la variable Salario, en ella se muestra que lo que indicaba el resumen estadístico se está cumpliendo, además los datos están dispersos esto se veía representado en la varianza y el coeficiente de curtosis entre otros.



Gráfica 5 Caja y Bigotes Salario

En la gráfica 5, en la caja y bigotes de la variable Salario, se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 6 Histograma Salario.

En la gráfica 6, se puede decir que se trata de un histograma asimétrico hacia la izquierda.

7.3.-X₂: Población 30-44 años. Anexo I tabla pobl.30-44

Recuento	23
Promedio	1,14915E6
Mediana	1,1676E6
Varianza	1,65238E10
Desviación Estándar	128545,
Mínimo	895966,
Máximo	1,31056E6
Rango	414594,
Sesgo Estandarizado	-0,95074
Curtosis Estandarizada	-0,708896

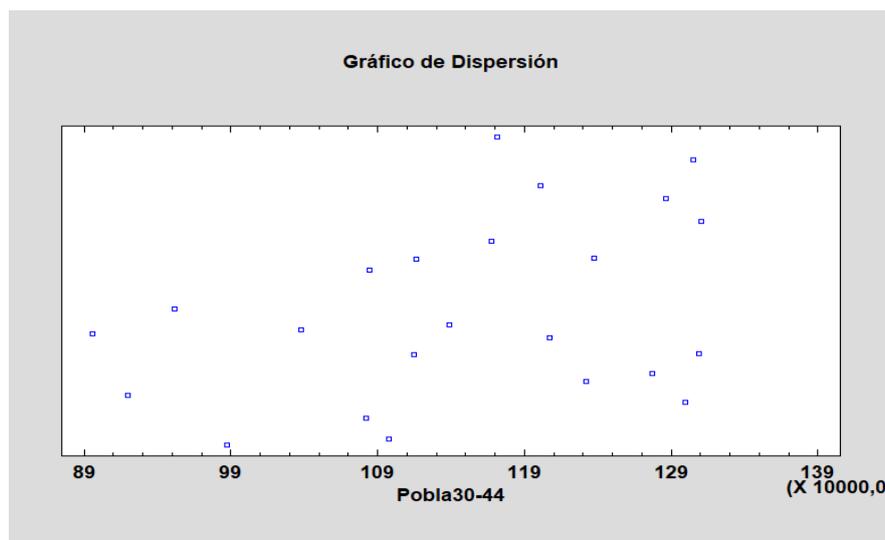
Tabla 3 Resumen Estadístico Pobl30-44

En la tabla 3, se presenta un resumen estadístico de la variable Población 30-44, en esta tabla se observa que existe una muestra 23 datos, presentando una media de 1.149.150.-Número de personas entre 30-44 años.

En cuanto a la varianza se observa que, al ser un valor intermedio, indicará que en el análisis de dispersión la muestra va a bastante dispersa, debido a que el valor es superior a 1.

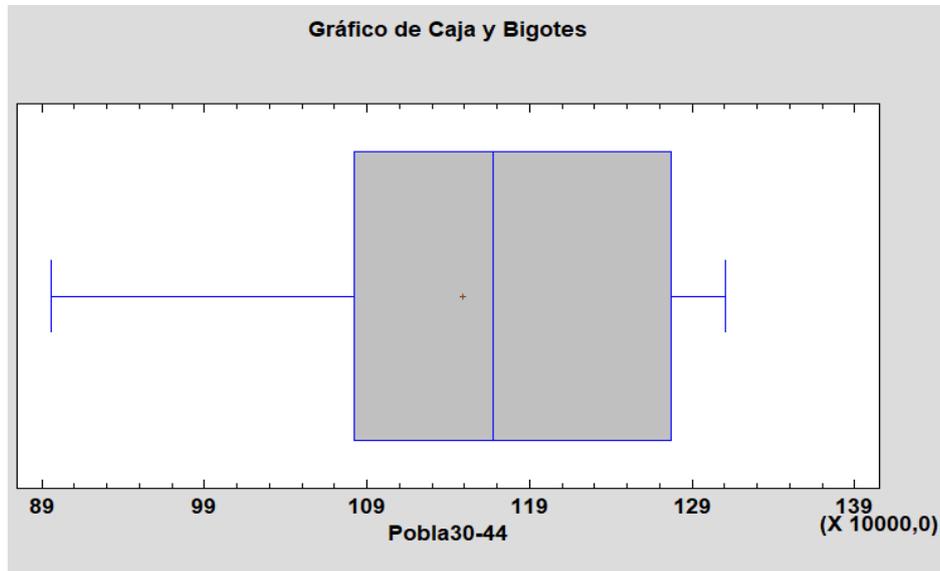
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es inferior a la mediana aparecerá a la izquierda en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es Platicúrtica es decir hay una baja concentración de los valores.



Gráfica 7 Dispersión Pobl30-44

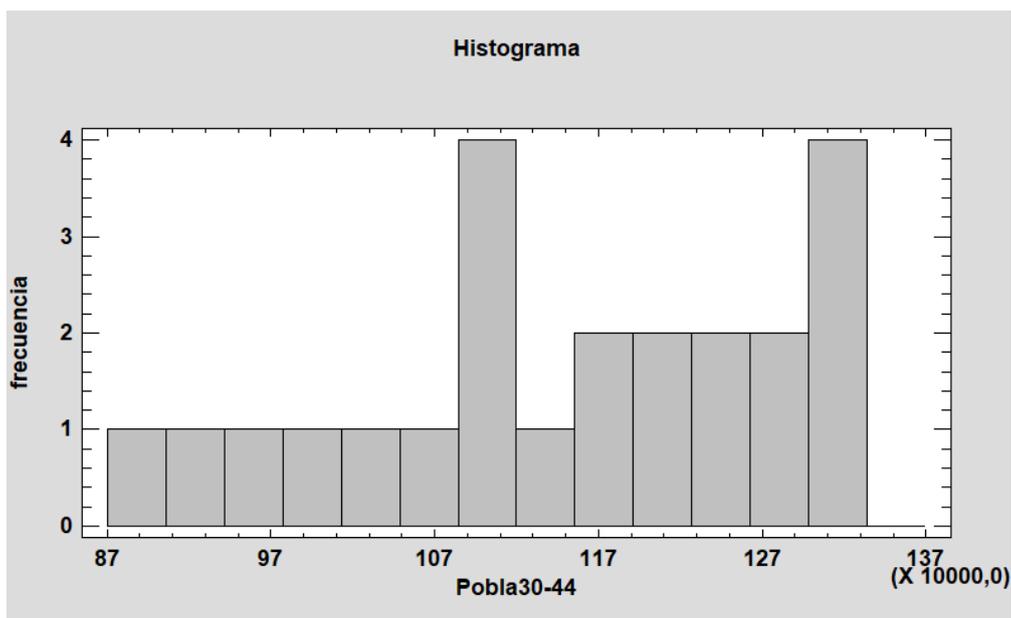
En la gráfica 7, dispersión de la variable pobla30-44, se aprecia cómo además de tener una mayor muestra que en las anteriores, los datos están dispersos, pero menos que los anteriores, esto se ve representado en la tabla anterior en la varianza o también en la desviación típica.



Gráfica 8 Cajas y Bigotes Población 30-44.

En la gráfica 8, en la caja y bigotes de la variable Población 30-44, se aprecia una asimetría negativa, además se observa como la media de la muestra está comprendida entre el primer cuartil y la mediana.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 9 Histograma Población 30-44

En la gráfica 9, histograma de la variable Población entre 30 y 44 años, se observa cómo es un histograma asimétrico hacia la izquierda con 2 valores muy altos tocando el máximo.

7.4.-X₂₂: Nacimientos. Anexo I tabla nacimientos

Recuento	48
Promedio	11412,7
Mediana	11310,5
Varianza	2,42973E6
Desviación Estándar	1558,76
Mínimo	8595,0
Máximo	14645,0
Rango	6050,0
Sesgo Estandarizado	0,939988
Curtosis Estandarizada	-0,942878

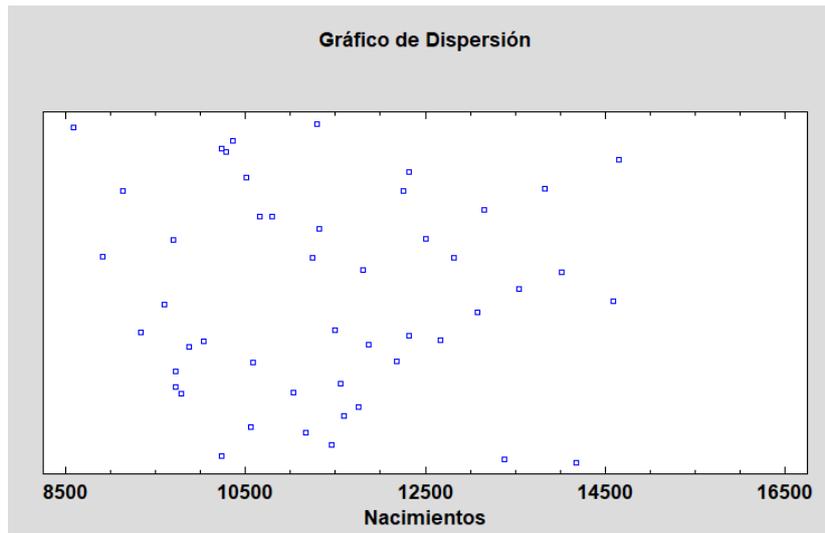
Tabla 4 Resumen estadístico Nacimientos.

En la tabla 4, se presenta un resumen estadístico de la variable Nacimientos, en esta tabla se observa que existe una muestra 48 datos, donde su media es de 11.412.-Nacimientos.

En cuanto a la varianza se observa que, al ser un valor alto, indicará que en el análisis de dispersión la muestra va a ser bastante dispersa, debido a que el valor es superior a 1.

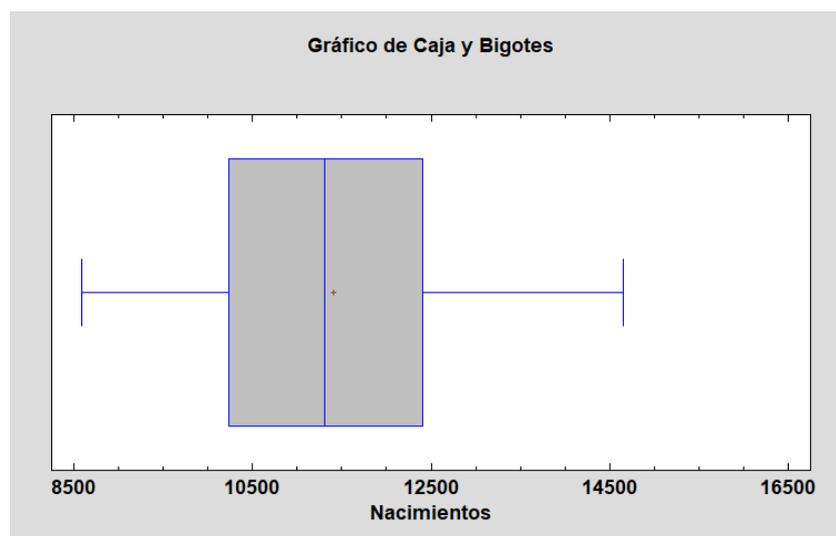
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es superior a la mediana aparecerá representado a la derecha en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es Platicúrtica es decir hay una baja concentración de los valores.



Gráfica 10 Dispersión Nacimientos.

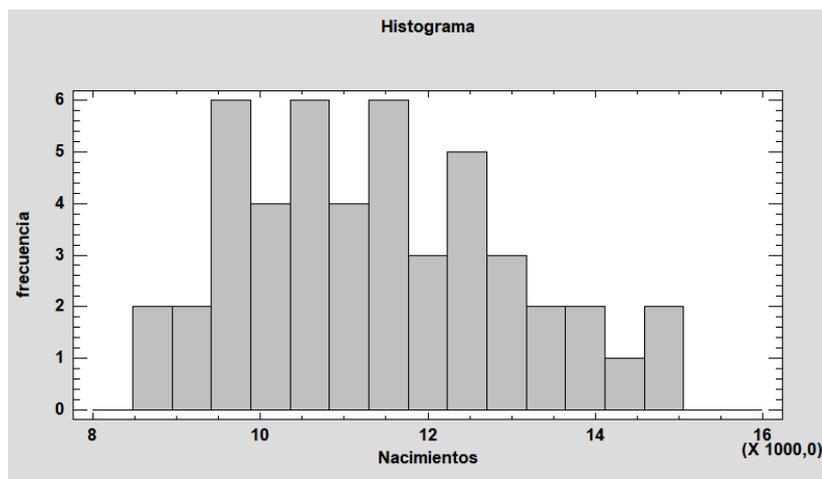
En la gráfica 10, dispersión de la variable Nacimientos, se observa cómo además de lo que ya indicaban en el resumen estadístico, los datos están dispersos esto es debido a que en la varianza y el coeficiente de curtosis ya lo indicaba.



Gráfica 11 Caja y Bigotes Nacimientos.

En la gráfica 11, en la caja y bigotes de la variable Nacimientos, se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 12 Histograma Nacimientos.

En la gráfica 12, Histograma de nacimientos, se observa que tiene un histograma asimétrico hacia la derecha.

7.5.-X₃: Euribor. Anexo I tabla euribor

Recuento	269
Promedio	0,0184565
Mediana	0,0177
Varianza	0,000307422
Desviación Estándar	0,0175335
Mínimo	-0,0051
Máximo	0,0539
Rango	0,059
Sesgo Estandarizado	2,37716
Curtosis Estandarizada	-3,76898

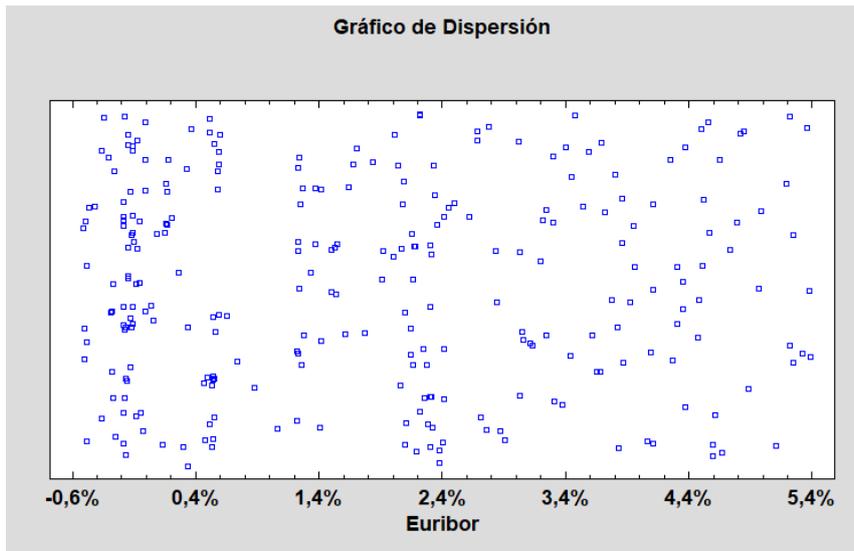
Tabla 5 Resumen estadístico Euribor.

En la tabla 5, se presenta un resumen estadístico de la variable Euribor, en esta tabla se observa que existe una muestra 269 datos, con una media de 0.0184565%.

En cuanto a la varianza se observa que, al ser un valor pequeño, indicará que en el análisis de dispersión la muestra va a estar agrupada, debido a que el valor es cercano a 0.

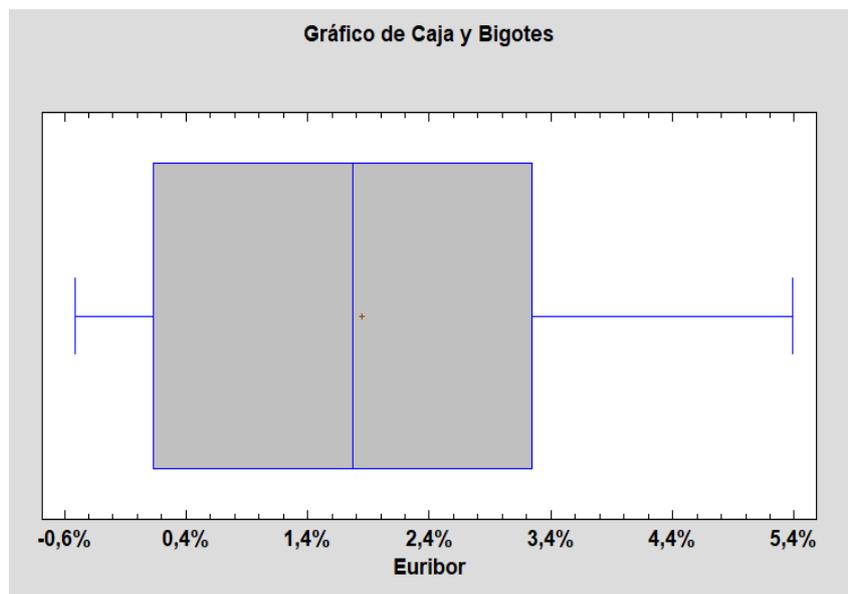
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es superior a la mediana aparecerá representado a la derecha en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es Platicúrtica es decir hay una baja concentración de los valores.



Gráfica 13 Dispersión Euribor

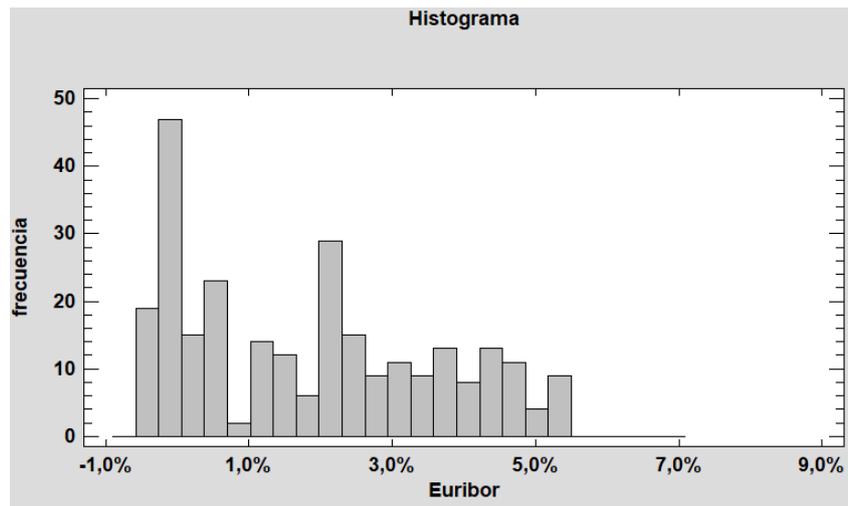
En la gráfica 13, dispersión de la variable Euribor, se puede observar cómo además de tener una mayor muestra, los datos están menos dispersos que los anteriores eso también se ha visto representado en la tabla anterior en la varianza y en la desviación estándar. Y se observa como existe una agrupación de los datos.



Gráfica 14 Caja y Bigotes Euribor.

En la gráfica 14, en la caja y bigotes de la variable Euribor, se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 15 Histograma Euribor.

En la gráfica 15, histograma para la variable Euribor, se puede apreciar que tiene un histograma asimétrico hacia la derecha.

7.6.-X₄: Parados. Anexo I Tabla nº parados

Recuento	77
Promedio	3677,54
Mediana	3719,8
Varianza	1,97497E6
Desviación Estándar	1405,34
Mínimo	1766,9
Máximo	6278,2
Rango	4511,3
Sesgo Estandarizado	0,631813
Curtosis Estandarizada	-2,28992

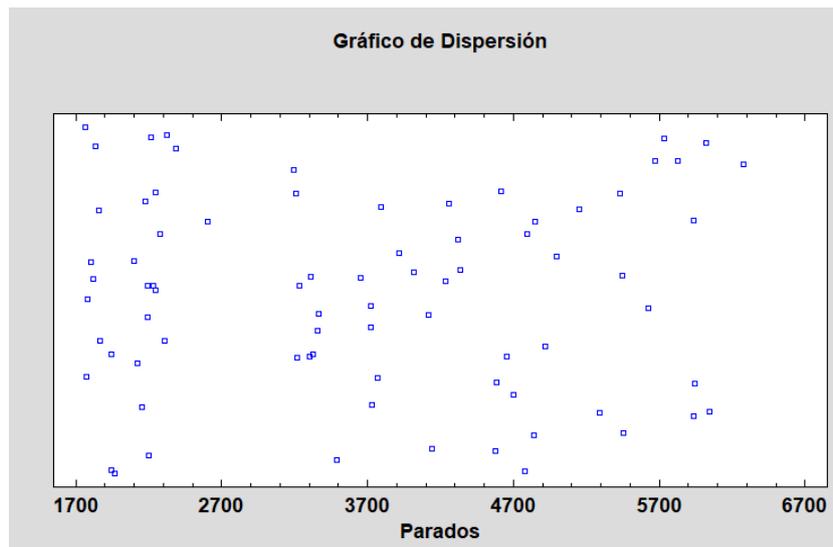
Tabla 6 Resumen estadístico Parados

En la tabla 6, se presenta un resumen estadístico de la variable Parados, en esta tabla se observa que existe una muestra 77 datos con una media de 3677.-número de parados

En cuanto a la varianza se observa que, al ser un valor pequeño, indicará que en el análisis de dispersión la muestra va a ser bastante dispersa, debido a que el valor es superior a 1

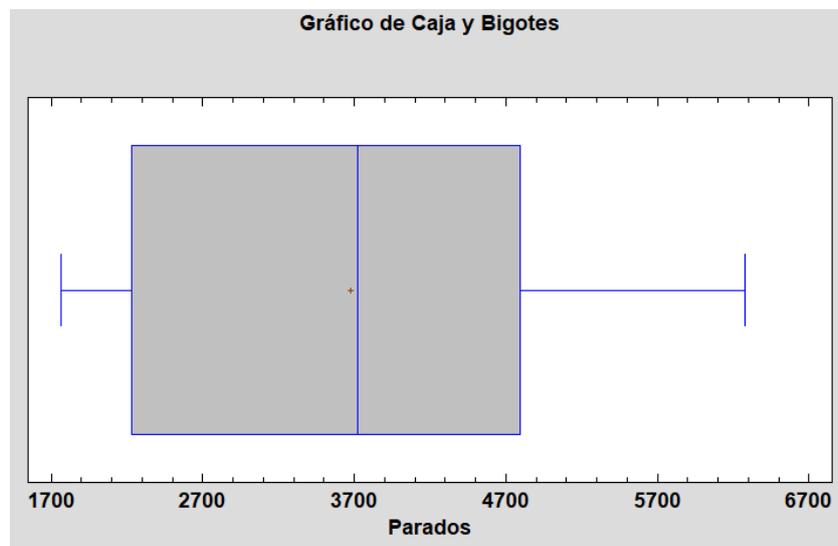
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es inferior a la mediana aparecerá representado a la izquierda en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es platicúrtica, es decir hay una baja concentración de los valores.



Gráfica 16 Dispersión Parados

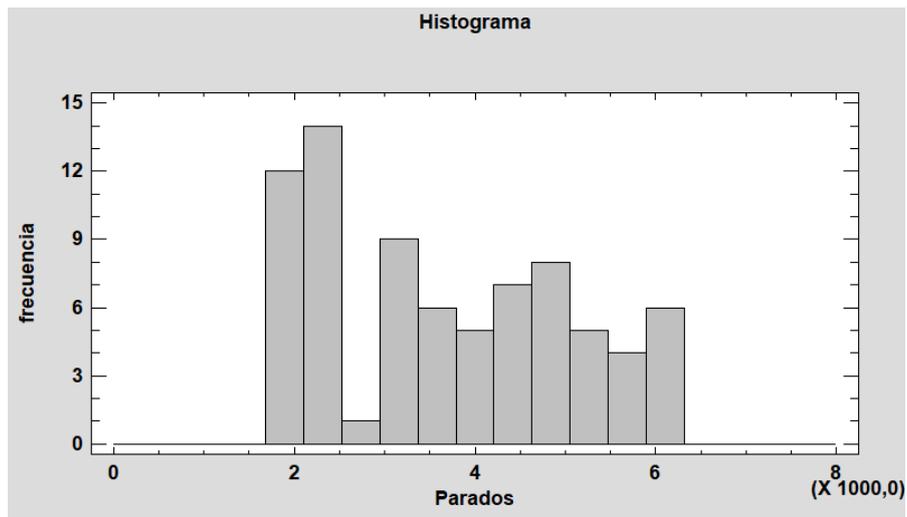
En la gráfica 16, dispersión de la variable Parados, se observa que existen una gran cantidad de datos pero que están bastantes dispersos, esto se puede llegar a imaginar antes de crear este gráfico debido a que en la tabla resumen anterior la varianza es elevada y la desviación estándar también es lejana a 0, lo que nos indica que tiene una dispersión mayor.



Gráfica 17 Caja y Bigotes Parados

En la gráfica 17, en la caja y bigotes de la variable Parados, se aprecia una asimetría negativa, además se observa como la media de la muestra está comprendida entre el primer cuartil y la mediana.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 18 Histograma Parados

En la gráfica 18, histograma para la variable Parados, se aprecia un histograma asimétrico hacia la derecha.

7.7.-X₅: Viviendas Iniciadas. Anexo I tabla viviendas iniciadas

Recuento	156
Promedio	672,423
Mediana	520,5
Varianza	461536,
Desviación Estándar	679,364
Mínimo	62,0
Máximo	4573,0
Rango	4511,0
Sesgo Estandarizado	18,7989
Curtosis Estandarizada	41,4748

Tabla 7 Resumen estadístico VivIni

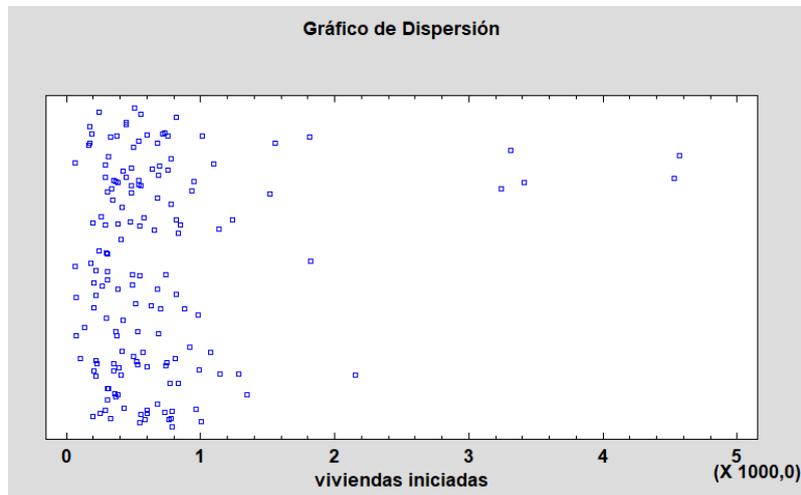
En la tabla 7, se presenta un resumen estadístico de la variable Viviendas Iniciadas, en esta tabla se observa que existe una muestra 156 datos, con una media de 672 viviendas iniciadas.

En cuanto a la varianza se observa que, al ser un valor muy grande, indicará que en el análisis de dispersión la muestra va a estar muy dispersa, debido a que el valor como se ha comentado ya es superior a 400.000.

Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es superior a la mediana aparecerá representado a la derecha en el gráfico de cajas y bigotes.

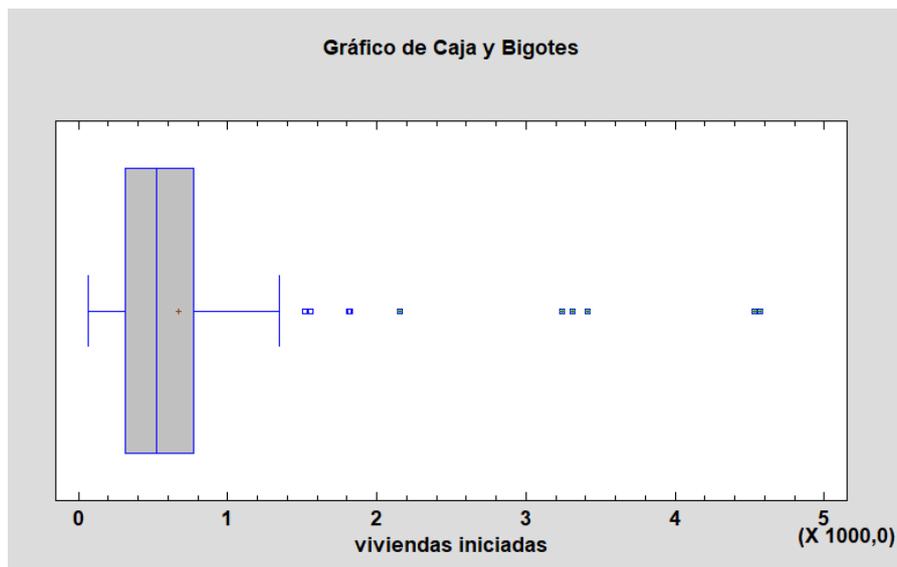
Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es positivo y superior a 3 lo que indica que habrá colas menos largas y que los datos se agruparán

más, provocando así una distribución de los datos más puntiaguda que una normal, debido a que es superior a 3, por tanto, es Leptocúrtica, hay una concentración de los valores.



Gráfica 19 Dispersión VivIni.

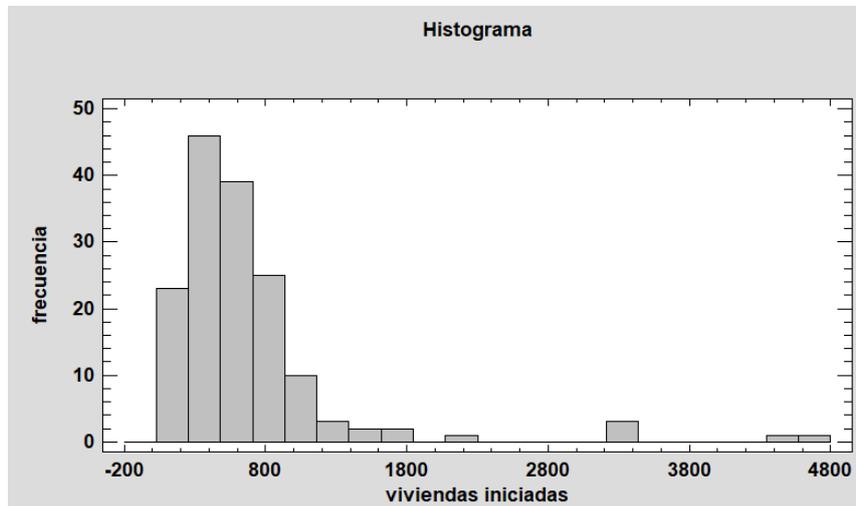
En la gráfica 19, dispersión de la variable Viviendas Iniciadas, se observa que existen una gran cantidad de datos pero que están bastantes juntos, esto se puede llegar a imaginar antes de crear este gráfico debido a que en la tabla resumen anterior el coeficiente Curtosis es positivo, pero cabe destacar que existe una dispersión de algunos datos que podrían llegar a ser anómalos, aunque si se observa la varianza y la desviación estándar también podrían ser buenos, debido a la varianza y desviación estándar.



Gráfica 20 Caja y Bigotes VivIni

En la gráfica 20, en la caja y bigotes de la variable Vivlni, se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos existen algunos datos que están superando los límites de prolongación, por lo que en esta variable existen datos anómalo.



Gráfica 21 Histograma Vivlni.

En la gráfica 21, histograma para la variable de viviendas iniciadas, se puede apreciar que el histograma se trata de un histograma asimétrico a la derecha

7.8.-X₆: Densidad de Población. Anexo I tabla densidad población.

Recuento	56
Promedio	5,02808E6
Mediana	5,01722E6
Varianza	5,66805E9
Desviación Estándar	75286,5
Mínimo	4,88503E6
Máximo	5,12927E6
Rango	244237,
Sesgo Estandarizado	-0,547085
Curtosis Estandarizada	-1,71763

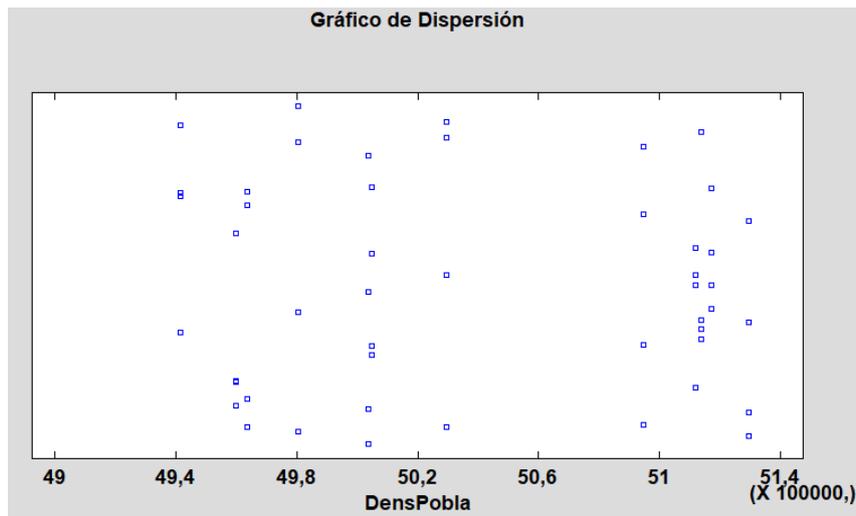
Tabla 8 Resumen estadístico DensPobla.

En la tabla 8, se presenta un resumen estadístico de la variable DensPobla, en esta tabla se observa que existe una muestra 56 datos, con una media de 5.028.080.-número de habitantes.

En cuanto a la varianza se observa que, al ser un valor grande, indicará que en el análisis de dispersión la muestra va a estar dispersa, debido a que el valor es superior a 5.

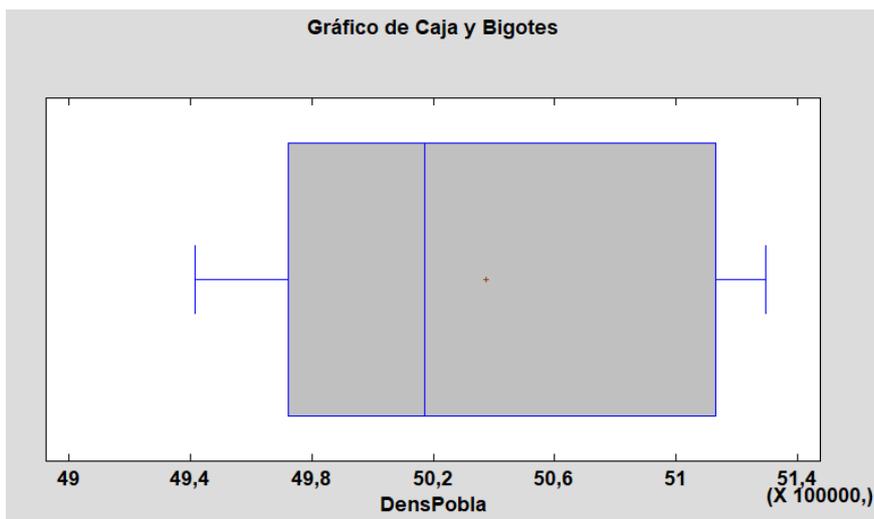
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es superior a la mediana aparecerá representado a la derecha en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es platicúrtica y no hay un gran concentración.



Gráfica 22 Dispersión DensPobla

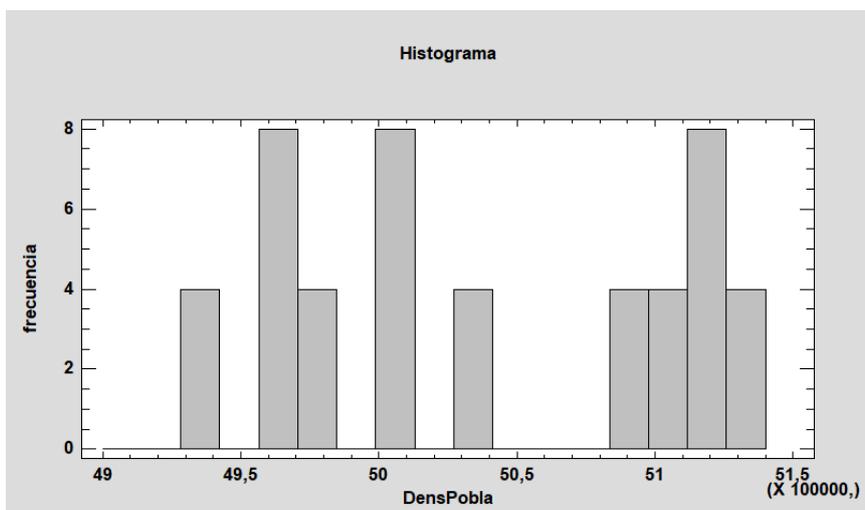
En la gráfica 22, dispersión de la variable Densidad Población, se observa que existen datos pero que están bastantes dispersos, esto se puede llegar a imaginar antes de crear este gráfico debido a que en la tabla resumen anterior el coeficiente Curtosis es negativo pero muy baja, pero cabe destacar que existe una dispersión de algunos datos que podrían llegar a ser anómalos, aunque observando la varianza y la desviación estándar podrían entrar dentro debido a que los dos son lejanos a 0.



Gráfica 23 Caja y Bigotes DensPobla.

En la gráfica 23, en la caja y bigotes de la variable DensPobla, se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 24 Histograma DensPobla

En la gráfica 24, histograma para variable Densidad de Población, se aprecia que existe un histograma asimétrico hacia la izquierda.

7.9.-X₇: PIB. Anexo I tabla PIB

Recuento	104
Promedio	227986,
Mediana	253753,
Varianza	3,66515E9
Desviación Estándar	60540,5
Mínimo	109165,
Máximo	324901,
Rango	215736,
Sesgo Estandarizado	-2,34285
Curtosis Estandarizada	-2,13693

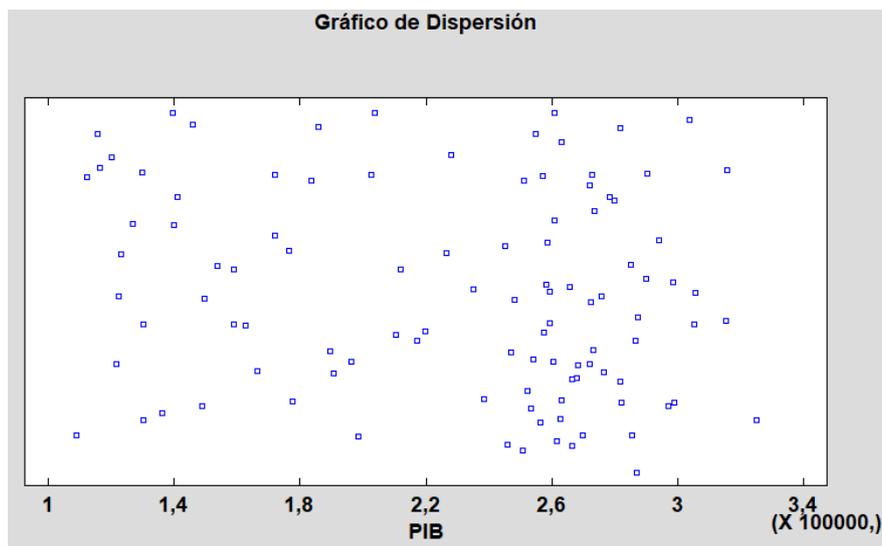
Tabla 9 Resumen estadístico PIB.

En la tabla 9, se presenta un resumen estadístico de la variable PIB, en esta tabla se observa que existe una muestra 104 datos con una media de 227986 per cápita.

En cuanto a la varianza se observa que, al ser un valor grande, indicará que en el análisis de dispersión la muestra va a estar dispersa, debido a que el valor es superior a 3.

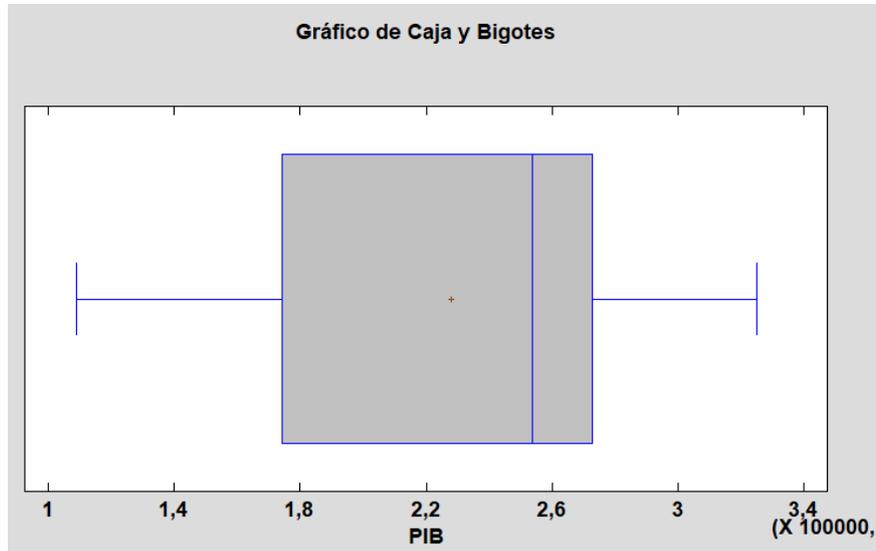
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es inferior a la mediana aparecerá representado a la izquierda en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3, por tanto, es Platicúrtica y sus datos no se encuentran concentrados.



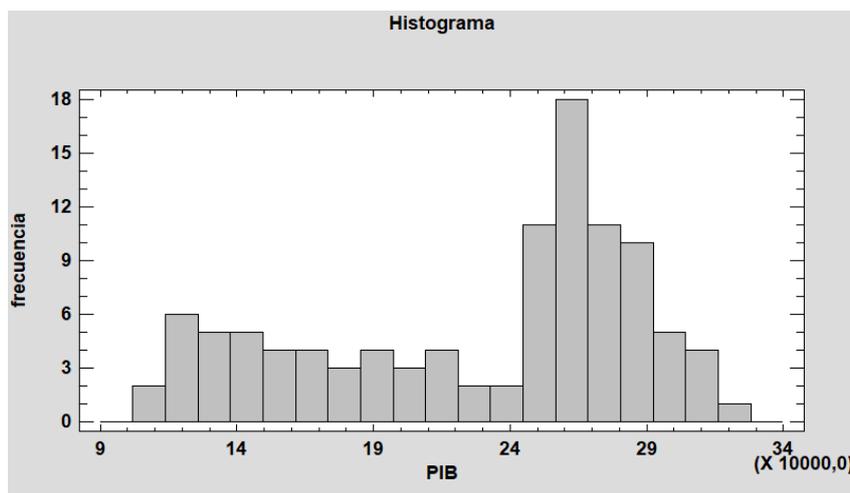
Gráfica 25 Dispersión PIB.

En la gráfica 25, dispersión de la variable PIB, se observa lo que se había apreciado antes en el resumen estadístico, existe una dispersión bastante abultada, aunque ya se podría imaginar debido a los datos obtenidos por la varianza, por la desviación típica y por el coeficiente de curtosis estandarizado que son datos que a priori muestran que va a existir dispersión.



Gráfica 26 Caja y Bigotes PIB

En la gráfica 26, en la caja y bigotes de la variable PIB, se aprecia una asimetría negativa, además se observa como la media de la muestra está comprendida entre el primer cuartil y la mediana. Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 27 Histograma PIB.

En la gráfica 27, histograma para la variable PIB, se observa que el histograma aparece de una manera asimétrica hacia la izquierda

7.10.-X₈: Precio de M². Anexo I tabla precio m²

Recuento	68
Promedio	231,449
Mediana	197,6
Varianza	7072,24
Desviación Estándar	84,0966
Mínimo	128,7
Máximo	423,8
Rango	295,1
Sesgo Estandarizado	2,11502
Curtosis Estandarizada	-1,72792

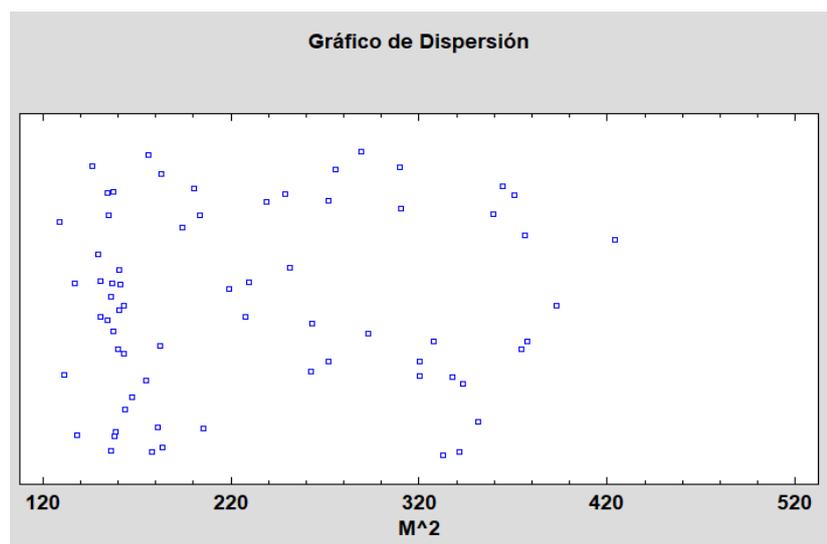
Tabla 10 Resumen estadístico M².

En la tabla 10, se presenta un resumen estadístico de la variable M², en esta tabla se observa que existe una muestra 68 datos y una media de 231,44.-€ por metro cuadrado.

En cuanto a la varianza se observa que, al ser un valor muy elevado, indicará que en el análisis de dispersión la muestra va a estar muy dispersa, debido a que el valor es superior a 7072.

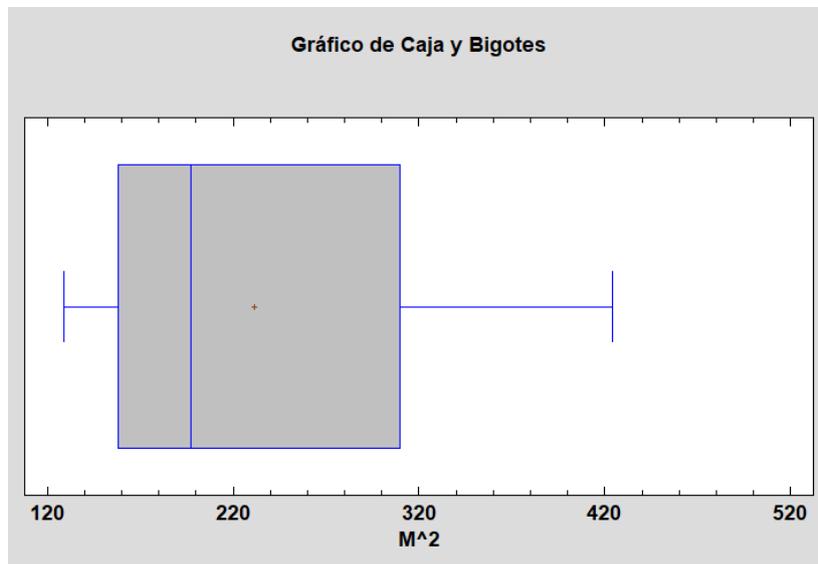
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es superior a la mediana aparecerá representado a la derecha en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3 por tanto es platicúrtica y no hay un gran concentración.



Gráfica 28 Dispersión M²

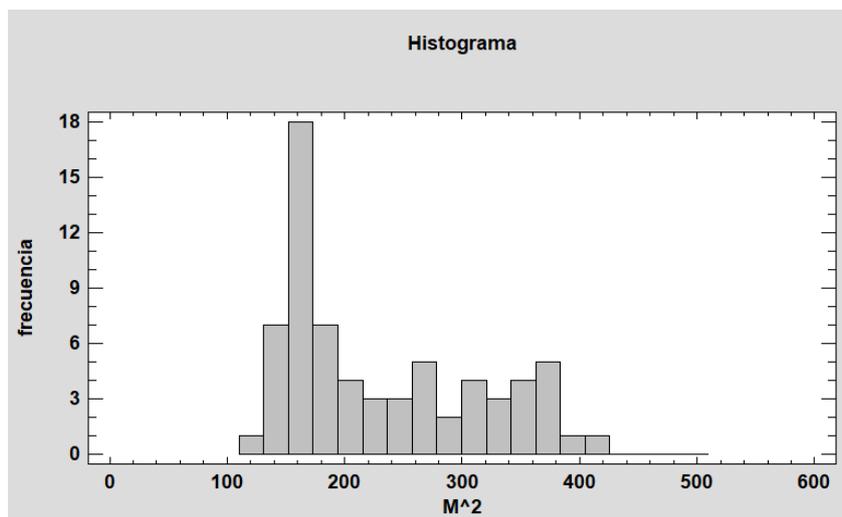
En la gráfica 28, dispersión de la variable M^2 , se observa que existen una gran cantidad de datos pero que están bastantes juntos, pero cabe destacar que existe una dispersión de algunos datos debido a que la varianza y la desviación estándar son lejanas a 0.



Gráfica 29 Caja y Bigotes M^2 .

En la gráfica 29, en la caja y bigotes de la variable M^2 , se aprecia una asimetría positiva, además se observa como la media de la muestra está comprendida entre la mediana y el tercer cuartil.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 30 Histograma M^2 .

En la gráfica 30, histograma para la variable M^2 , se aprecia un histograma asimétrico hacia la derecha

7.11.-X₉: Empresas Activas. Anexo I tabla empresas activas.

Recuento	37
Promedio	142559,
Mediana	143657,
Varianza	3,89879E7
Desviación Estándar	6244,03
Mínimo	132882,
Máximo	153692,
Rango	20810,0
Sesgo Estandarizado	-0,115266
Curtosis Estandarizada	-1,50927

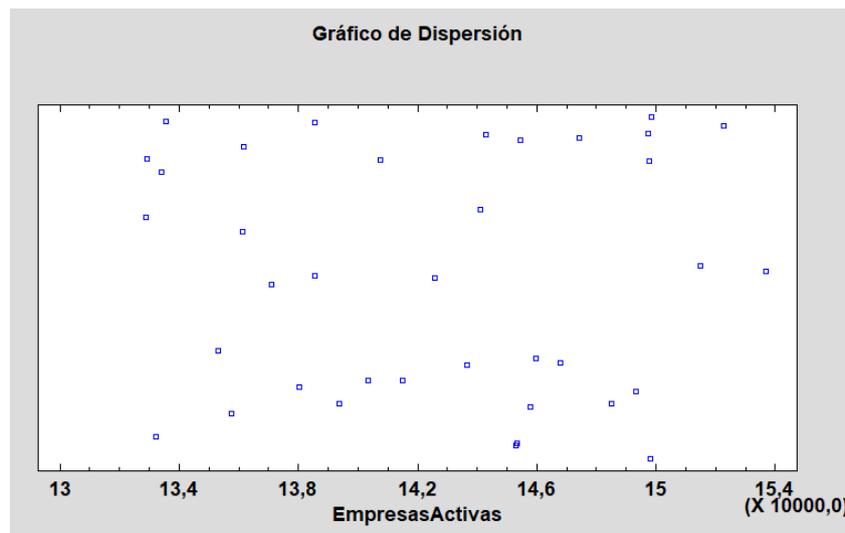
Tabla 11 Resumen estadístico Empresas Activas.

En la tabla 11, se presenta un resumen estadístico de la variable Euribor, en esta tabla se observa que existe una muestra 37 datos y una media de 142.559.-Empresas activas en la Comunidad Valenciana.

En cuanto a la varianza se observa que, al ser un valor grande, indicará que en el análisis de dispersión la muestra va a estar dispersa, debido a que el valor es superior a 3.

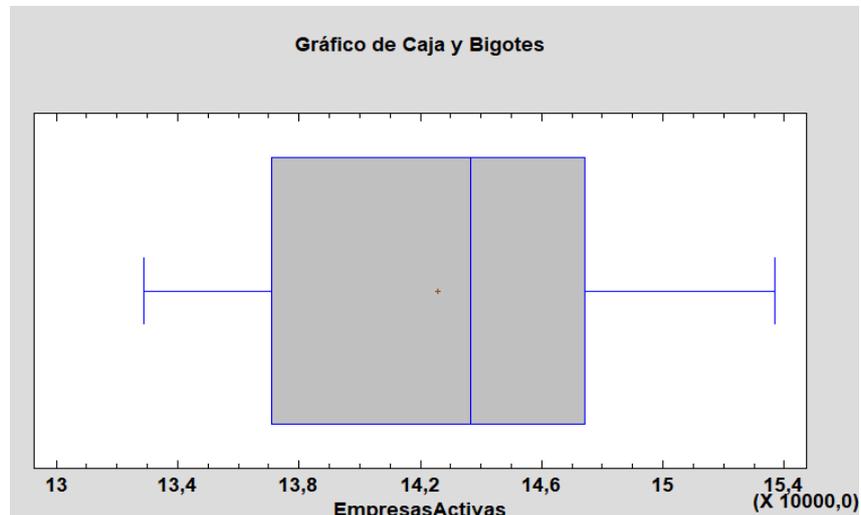
Además, se observa que en la mediana se verá si sus valores se quedarán a la izquierda o la derecha, como el promedio es inferior a la mediana aparecerá representado a la izquierda en el gráfico de cajas y bigotes.

Por último, otro dato relevante del que poder hablar es el Coeficiente de Curtosis, en este caso es negativo lo que indica que habrá colas más largas y que los datos se agruparán menos, provocando así una distribución de los datos más plana que una normal, debido a que es menor a 3 por tanto es platocúrtica y no hay un gran concentración.



Gráfica 31 Dispersión Empresas Activas

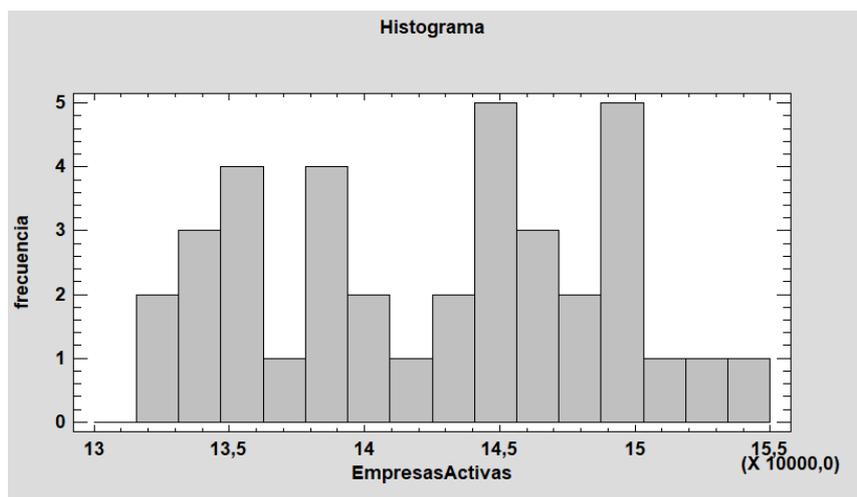
En la gráfica 31, dispersión de la variable EmpresasActivas, se aprecia lo que se ha observado antes en el resumen estadístico, existe una dispersión bastante abultada, aunque ya se podría imaginar debido a los datos obtenidos por la varianza, por la desviación típica y por el curtosis estandarizado que son datos que a priori indican que va a existir dispersión.



Gráfica 32 Caja y Bigotes Empresas Activas.

En la gráfica 32, en la caja y bigotes de la variable Empresas Activas, se aprecia una asimetría negativa, además se observa como la media de la muestra está comprendida entre el primer cuartil y la mediana.

Los bigotes, tanto el mínimo como el máximo, muestran que todos los datos están comprendidos entre los límites de prolongación y que no existe ningún dato anómalo.



Gráfica 33 Histograma Empresas Activas.

En la gráfica 33, histograma para la variable Empresas Activas, se trata de un histograma lo más parecido al bimodal, ya que en mitad del histograma decaen los datos.

8.-Análisis Bivariante

A continuación, se presentan los resultados de los análisis bivariantes realizados. Concretamente se han realizado diversos modelos de regresión lineal simples donde se relacionan cada una de las variables X (Salarios, Nacimientos, DensPobla, PIB, EmpresasActivas, Euribor, M², Viviendas iniciadas, Parados) con la variable endógena Y, el Índice de precio de las Viviendas.

Para este estudio se ha tenido que convertir algunas de las variables cuyos datos existían de manera mensual y se han pasado a trimestrales para poder hacer el desarrollo del estudio, algunas de las variables que se han convertido son Núm. Viviendas iniciadas y el Euribor.

Para cada análisis bivalente se va a tratar de estudiar los siguientes datos R², observar el P-Valor si es o no significativo (Fisher y Tstudent) tanto en el modelo como en parámetros, si los residuos se distribuyen de manera normal y por último si existe autocorrelación. Se va a realizar para cada una de las variables independientes.

²La prueba exacta de Fisher para la comprobación del modelo se debe tener en cuenta lo siguiente:

Si aceptamos la hipótesis nula, es decir no existe evidencia para rechazar H₀, el modelo que se construye no tiene sentido, es decir, todos los parámetros β_i toman valor cero, lo que quiere decir que ninguna de las variables independientes sirve para explicar la variable dependiente.

- Si rechazamos la hipótesis nula, alguna o algunas de las variables independientes que hemos incluido en el modelo sirve para explicar la variable dependiente. Por tanto, el objetivo ahora es estudiar qué variables de todas sirven para explicar Y.

Además de lo visto anteriormente se puede decidir mediante la función del p-valor.

- Si P-Valor es inferior a α rechazaremos la hipótesis nula, por tanto, el modelo es significativo
- Si P-Valor es mayor a α no existen evidencias para rechazar la hipótesis nula, por tanto, el modelo será no significativo.
- Siendo el valor de $\alpha = 0,5$

Quedando estas hipótesis:

- Hipótesis Nula H₀ : $\beta_i = 0$
- Hipótesis Alternativa H₁ : $\beta_i \neq 0$

Contraste de Significación Individual.

Para la comprobación de la significación de las variables, se utiliza la prueba de T-Student para la validación de estas.

² (Apuntes Métodos Estadísticos Aplicados a la Economía, UPV 2019)

- Prueba T-Student
 - $H_0 : \beta_i = 0$
 - $H_1 : \beta_i \neq 0$

(Hipótesis nula) $H_0 : \beta_i = 0$ significa que el valor de acompaña a la variable explicativa, tiene un valor de cero y por tanto cuando multiplica esta Beta a la variable explicativa, su resultado es cero, por eso se afirma que la variable no es significativa.

(Hipótesis alternativa) $H_1 : \beta_i \neq 0$ significa que el valor que acompaña a la variable explicativa es diferente de cero y por tanto cuando se produce la multiplicación del valor Beta por la variable explicativa su resultado es distinto de cero, por eso se afirma que la variable es significativa.

Si los cálculos se realizan con un programa estadístico, se puede decidir en función del p-valor:

Siendo el valor de $\alpha = 0,5$

- Si $p\text{-valor} < \alpha$, se rechaza la hipótesis nula, y por tanto se acepta la hipótesis alternativa, siendo significativa esa variables.
- -Si $p\text{-valor} > \alpha$, no existe evidencia para rechazar la hipótesis alternativa, por tanto la variable no es significativa.

³Hay que explicar la Bondad del ajuste o también conocido como R^2

Esta es, una vez que se han realizado todos los contrastes pertinentes y se ha obtenido el modelo correcto, tenemos que observar qué porcentaje de bondad tiene el ajuste, o sea, si la relación lineal explica suficientemente la variabilidad de la variable dependiente, o por el contrario, otros factores que no se controlan explican más variabilidad que el modelo que se propone. Esto se lleva a cabo por medio del Coeficiente de Determinación o de Bondad de Ajuste.

Para calcularlo se realiza el cociente entre las sumas de cuadrados explicada por el modelo y la suma de cuadrados total:

$$R^2 = \frac{SCE}{SCT}$$

Además, cabe destacar que el R^2 mide la proporción de la variabilidad total que es explicada por el modelo, es decir indica qué porcentaje de las variables explicativas X justifica la variable independiente Y.

El objetivo, en principio, es detectar posibles relaciones lineales entre las variables (Análisis de correlación) y, si éstas presentan un fuerte grado de correlación, buscar la forma funcional que mejor explique la variable dependiente a partir de la independiente (Análisis de regresión).

Esto sirve para poder estudiar, por ejemplo, si existe relación entre la renta y el consumo de las familias y, en caso de que exista, tratar de encontrar la mejor relación funcional que explique el consumo a partir de la renta.

Una vez se estudia si existe o no la relación entre las variables se hará una prueba de la normalidad de los residuos con su histograma normal de los residuos para el estudio.

Respecto al cumplimiento de la hipótesis de normalidad es indispensable para poder realizar las pruebas de significación y obtener los intervalos de confianza que permitan realizar

³ Apuntes Econometría, UPV 2021

predicciones. Además de las pruebas numéricas, es posible comprobar la normalidad del error representando gráficamente los residuos en un histograma o como se verá en otros casos en un papel probabilístico.

Como último estudio de este se hace la autocorrelación:

La autocorrelación aparece cuando cualquier pareja de valores de la perturbación están relacionados entre sí, $Cov(U_i U_j) \neq 0$ con $i \neq j$, se dice que existe autocorrelación, y deja de cumplirse una de las hipótesis del modelo de mínimos cuadrados ordinarios.

Entonces el modelo se ve afectado en diferentes aspectos:

- Las estimaciones b siguen siendo insesgadas, pero ya no son uniformemente de mínima varianza.
- Los estadísticos t y F para el contraste de los parámetros no son válidos. Los estadísticos ya no siguen la distribución correspondiente, y sus valores son más grandes de lo que cabría esperar, lo que llevaría a aceptar como variables explicativas aquellas que no lo son.
- Las predicciones que ofrece el modelo no son adecuadas.

El coeficiente de autocorrelación de orden k se define como:

$$P_k = (Y_k / Y_0)$$

Con lo que queda cuantificada la importancia de la relación lineal que existe entre un error y su anterior.

Las causas que provocan la autocorrelación son:

1. Las variables estudiadas son series temporales.
2. Las variables estudiadas son series espaciales.
3. El modelo no ha sido correctamente formulado.
4. La variable por explicar se ha medido de forma que aparece el error.

Determinar la existencia de autocorrelación es algo complicado, porque bajo este nombre se encuentran dos cuestiones importantes como son la forma de la autocorrelación y su orden. La mayoría de las pruebas de hipótesis asumen una forma o un orden para determinar la autocorrelación, lo que hace que las pruebas sean buenas para determinar la existencia de un tipo muy determinado de autocorrelación, pero totalmente inútiles para otros.

Como se verá en el trabajo se realizará una mejor forma de comprobar la existencia de autocorrelación es, sin duda, obtener y representar las funciones de autocorrelación simple y parcial de los residuos.

La función de autocorrelación simple (FAS) no es más que el conjunto de los coeficientes de autocorrelación simple ρ_k (con $k = 1, 2, 3, 4, \dots$), y miden la relación lineal que hay entre el residuo en el instante de tiempo t , e_t , y el residuo en el instante $t-k$, e_{t-k} .

La función de autocorrelación parcial (FAP) es el conjunto de coeficientes de autocorrelación parcial α_k (con $k = 1, 2, 3, 4, \dots$), y miden la relación directa que existe entre el residuo en el instante de tiempo t , e_t , y el residuo en el instante $t-k$, e_{t-k} , eliminando el posible efecto de los residuos intermedios a ambos.

8.1.-Modelo de regresión simple: IPV en función de la variable Salarios

En primer lugar, se va a estudiar cómo afecta la variable X_1 Salarios sobre la variable endógena IPV, planteando el siguiente modelo de regresión lineal simple: **IPV = B0+ B1*SALARIO + U**

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	185,706	51,5844	3,60003	0,0008
Salario	-0,042518	0,0306851	-1,38562	0,1725

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	487,462	1	487,462	1,92	0,1725
Residuo	11679,1	46	253,893		
Total (Corr.)	12166,5	47			

Tabla 12 Análisis Varianza Salarios

R-cuadrada = 4,00658 por ciento

R-cuadrado (ajustado para g.l.) = 1,91977 por ciento

Validación Modelo

- Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCE}{k-1}}{\frac{SCR}{n-k}}$$

En la tabla 12, como el P-Valor es 0,1725 es superior a Alfa, por tanto No existe evidencia para rechazar la hipótesis alternativa, por consiguiente el modelo no es significativo.

Validación Parámetros.

- Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0008, es inferior a Alfa, se rechaza la hipótesis nula, por tanto la constante es significativa.

P- Valor Salario = 0,1725 es superior a Alfa, por tanto No existe evidencia para rechazar la hipótesis alternativa, por consiguiente el parámetro salario no es significativo.

Si se observa en el R^2 ajustado esta muestra que el 1.9197% es el peso que tiene la variable Salarios sobre la variable endógena IPV. Este dato viene a indicar que esta variable no afectará mucho sobre la variable IPV debido al pequeño porcentaje que tiene sobre el total.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de Salarios.

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	23,25	0,0789742
Estadístico W de Shapiro-Wilk	0,916984	0,00233104
Valor-Z para asimetría	1,04496	0,296041
Valor-Z para curtosis	-2,58204	0,00982183

Tabla 13 Pruebas Normalidad Residuos Salarios.

Test hipótesis Residuos

H_0 : Los residuos se distribuyen normalmente

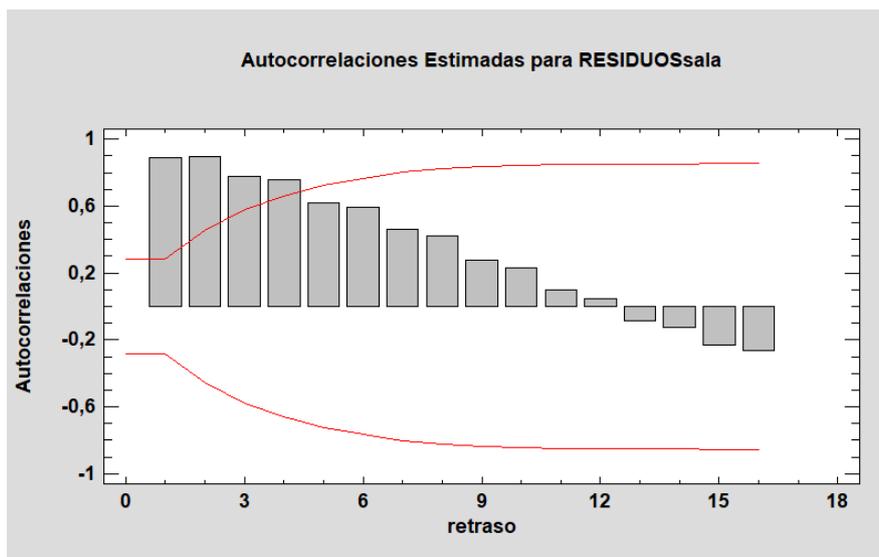
H_1 : Los residuos no se distribuyen normalmente $\alpha > 0,05$

En la tabla 13, como P-Valor, 0,00233104, es menor que alfa, 0,05, se rechaza la hipótesis nula, con lo que se concluye que los residuos no se distribuyen de forma normal.

			<i>Límite en 95,0%</i>	<i>Límite en 95,0%</i>
<i>Retraso</i>	<i>Autocorrelación</i>	<i>Error Estd.</i>	<i>Inferior</i>	<i>Superior</i>
1	0,891459	0,144338	-0,282897	0,282897
2	0,897043	0,232262	-0,455227	0,455227
3	0,774759	0,295761	-0,579681	0,579681
4	0,756264	0,335388	-0,657349	0,657349
5	0,622222	0,369209	-0,723638	0,723638
6	0,59641	0,390445	-0,765259	0,765259
7	0,459805	0,408984	-0,801596	0,801596
8	0,420015	0,419616	-0,822433	0,822433
9	0,277939	0,428285	-0,839424	0,839424
10	0,23161	0,432026	-0,846757	0,846757
11	0,0976968	0,434605	-0,851812	0,851812
12	0,0480416	0,435063	-0,852709	0,852709
13	-0,0842097	0,435173	-0,852925	0,852925
14	-0,124342	0,435513	-0,85359	0,85359
15	-0,233338	0,436251	-0,855039	0,855039
16	-0,260773	0,438844	-0,86012	0,86012

Tabla 14 Autocorrelación Residuos Salario

En la Tabla 14, se observa como existe una autocorrelación de nivel 4.



Gráfica 34 Autocorrelación Residuos Salarios

En la gráfica 34, autocorrelación residuos Salarios, se aprecia lo que se ha podido observar lo mismo que aparecía en la tabla, existe una autocorrelación de nivel 4 en esta variable.

8.2.- Modelo de regresión simple: IPV en función de la variable Nacimientos

En segundo lugar, se va a estudiar cómo afecta la variable X_2 Nacimientos sobre la variable endógena IPV. Se plantea este modelo de regresión lineal simple $IPV = B_0 + B_1 * NACIMIENTOS + U$

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	30,8115	12,3646	2,49191	0,0164
Nacimientos	0,00731539	0,00107364	6,81361	0,0000

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	6111,26	1	6111,26	46,43	0,0000
Residuo	6055,27	46	131,636		
Total (Corr.)	12166,5	47			

Tabla 15 Análisis Varianza Nacimientos

R-cuadrada = 50,2301 por ciento

R-cuadrado (ajustado para g.l.) = 49,1481 por ciento

Validación Modelo

- Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCE}{k-1}}{\frac{SCE}{n-k}}$$

En la tabla 15, como el P-Valor es 0,0000 es inferior a Alfa, por tanto se rechaza la hipótesis nula y se concluye afirmando que el modelo es significativo.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto la constante es significativa.

P- Valor Nacimientos = 0,0164 es inferior a Alfa, se rechaza la hipótesis nula, por el parámetro nacimiento es significativo.

Si se observa el R² ajustado esta muestra que el 49,1481 % es el peso que tiene la variable Nacimientos sobre la variable endógena IPV. Este dato viene a indicar que el número de nacimientos afectará a la variable IPV.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de Nacimientos.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	19,5	0,191963
Estadístico W de Shapiro-Wilk	0,956511	0,0729676
Valor-Z para asimetría	0,181709	0,855807
Valor-Z para curtosis	-3,015	0,00256992

Tabla 16 Pruebas Normalidad Residuos Nacimientos.

Test de Normalidad Residuos

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

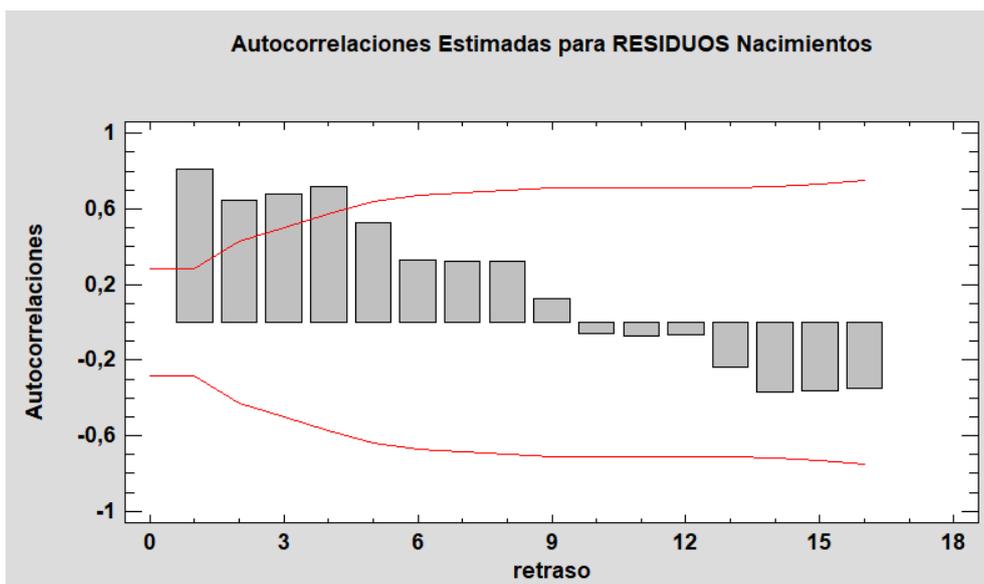
En la tabla 16, como P-Valor, 0,00256992, es menor que alfa, 0,05, rechazamos la hipótesis nula, con lo que concluimos que los residuos no se distribuyen de una manera o forma normal.

			Límite en 95,0%	Límite en 95,0%
Retraso	Autocorrelación	Error Estd.	Inferior	Superior
1	0,809256	0,144338	-0,282897	0,282897
2	0,645181	0,219364	-0,429947	0,429947
3	0,680206	0,255861	-0,501479	0,501479
4	0,721212	0,291107	-0,57056	0,57056
5	0,524775	0,326215	-0,63937	0,63937
6	0,328176	0,343352	-0,672958	0,672958
7	0,322075	0,349826	-0,685647	0,685647
8	0,323803	0,35595	-0,69765	0,69765
9	0,122161	0,362034	-0,709575	0,709575
10	-0,061321	0,362892	-0,711257	0,711257

11	-0,0731701	0,363108	-0,71168	0,71168
12	-0,0649743	0,363415	-0,712281	0,712281
13	-0,237005	0,363657	-0,712756	0,712756
14	-0,371751	0,366861	-0,719035	0,719035
15	-0,362972	0,374627	-0,734256	0,734256
16	-0,348672	0,381883	-0,748478	0,748478

Tabla 17 Autocorrelación Residuos Nacimientos.

En la tabla 17, se puede observar que existe una autocorrelación de los residuos de nivel 4.



Gráfica 35 Autocorrelaciones Residuos Nacimientos.

En la gráfica 35, se ha podido observar tanto en la tabla como a nivel gráfico que existe una autocorrelación de nivel 4 en esta variable.

8.3.- Modelo de regresión simple: IPV en función de la variable Euribor

En tercer lugar, se va a estudiar cómo afecta la variable X_3 Euribor sobre la variable endógena IPV. Se plantea el siguiente modelo de regresión lineal simple $IPV=B_0+B_1*EURIBOR + U$

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	107,561	1,68825	63,7116	0,0000
Euribor	797,046	85,9463	9,27377	0,0000

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	9275,77	1	9275,77	86,00	0,0000
Residuo	5824,13	54	107,854		

Total (Corr.)	15099,9	55			
---------------	---------	----	--	--	--

Tabla 18 Análisis Varianza Euribor.

R-cuadrada = 61,4293 por ciento

R-cuadrado (ajustado para g.l.) = 60,7151 por ciento

Validación Modelo

•Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCE}{k-1}}{\frac{SCE}{n-k}}$$

En la tabla 18, como el P-Valor es 0,0000 es inferior a Alfa, por tanto se rechaza la hipótesis nula y se concluye afirmando que el modelo es significativo.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, la constante es significativa.

P- Valor Euribor = 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro Euribor es significativo.

Si se observa el R² ajustado este muestra que el 60.715% es el peso que tiene la variable Euribor sobre la variable endógena IPV. Este dato viene a indicar que esta variable afectará bastante a la variable IPV debido al porcentaje que tiene.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de Euribor.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	15,25	0,50641
Estadístico W de Shapiro-Wilk	0,975349	0,30502
Valor-Z para asimetría	0,287315	0,773867
Valor-Z para curtosis	-1,84853	0,0645248

Tabla 19 Prueba Normalidad Residuos Euribor

Test hipótesis normalidad residuos.

H₀: Los residuos se distribuyen normalmente

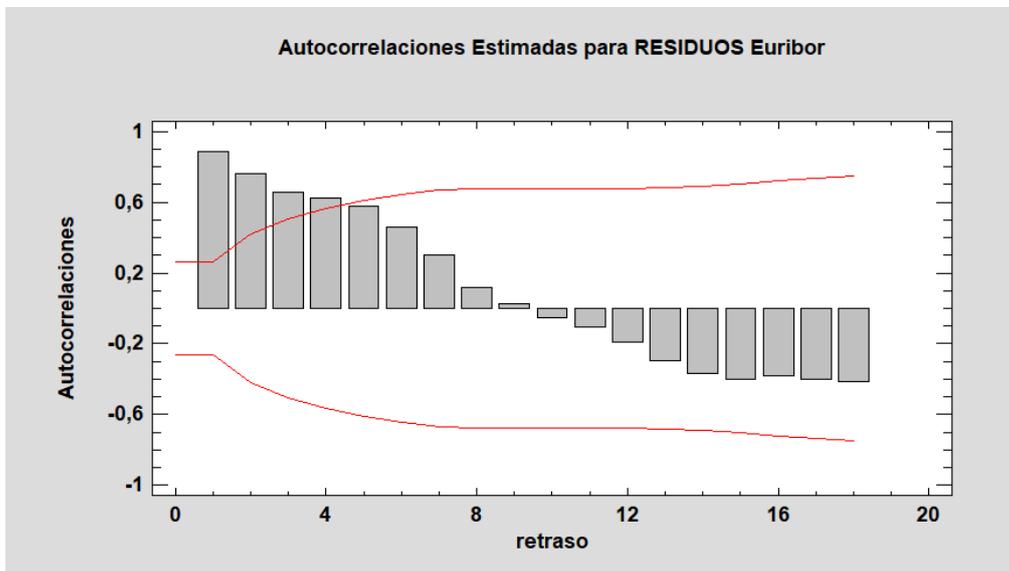
H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

En la tabla 19, como P-Valor, 0,0645248, es mayor o igual a alfa, 0,05, no existe evidencia para rechazar la hipótesis alternativa, con lo que se concluye que los residuos se distribuyen normalmente, de una manera normal.

			Límite en 95,0%	Límite en 95,0%
Retraso	Autocorrelación	Error Estd.	Inferior	Superior
1	0,889847	0,133631	-0,261912	0,261912
2	0,765091	0,214795	-0,420991	0,420991
3	0,65954	0,258926	-0,507486	0,507486
4	0,622494	0,287364	-0,563224	0,563224
5	0,57656	0,310511	-0,608592	0,608592
6	0,457416	0,329074	-0,644974	0,644974
7	0,300161	0,340238	-0,666856	0,666856
8	0,120364	0,344934	-0,67606	0,67606
9	0,0259703	0,345684	-0,677529	0,677529
10	-0,0547571	0,345718	-0,677597	0,677597
11	-0,107442	0,345873	-0,6779	0,6779
12	-0,18911	0,346469	-0,679067	0,679067
13	-0,295641	0,348307	-0,682671	0,682671
14	-0,369909	0,35276	-0,691397	0,691397
15	-0,402029	0,35962	-0,704843	0,704843
16	-0,383726	0,367558	-0,720401	0,720401
17	-0,399306	0,374643	-0,734288	0,734288
18	-0,415719	0,382167	-0,749036	0,749036

Tabla 20 Autocorrelación Residuos Euribor

En la tabla 20, se observa que existe una autocorrelación sobre los residuos Euribor de nivel 4.



Gráfica 36 Autocorrelación Residuos Euribor.

En la gráfica 36, como se ha podido observar tanto en la tabla como a nivel gráfico, existe una autocorrelación de nivel 4 en esta variable.

8.4.- Modelo de regresión simple: IPV en función de la variable Parados

En cuarto lugar, se va a estudiar cómo afecta la variable X_4 Parados sobre la variable endógena IPV. Se plantea el siguiente modelo de regresión lineal simple $IPV = B_0 + B_1 * PARADOS + U$

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	155,258	6,09858	25,4581	0,0000
Parados	-0,0091353	0,00138224	-6,60906	0,0000

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	6752,26	1	6752,26	43,68	0,0000
Residuo	8347,64	54	154,586		
Total (Corr.)	15099,9	55			

Tabla 21 Análisis Varianza Parados

R-cuadrada = 44,7172 por ciento

R-cuadrado (ajustado para g.l.) = 43,6935 por ciento

Validación Modelo

•Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCE}{k-1}}{\frac{SCE}{n-k}}$$

En la tabla 21, como el P-Valor es 0,0000 es inferior a Alfa, por tanto, se rechaza la hipótesis nula y se concluye afirmando que el modelo es significativo.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, la constante es significativa.

P- Valor Parados = 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro parados es significativo.

Entre el R^2 y el R^2 ajustado hay que fijarse en el ajustado y este muestra que el 43,69% es el peso que tiene la variable Parados sobre la variable endógena IPV. Por lo que se podría decir que existe una buena relación entre las variables.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de Parados.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	33,5714	0,00620264
Estadístico W de Shapiro-Wilk	0,888558	0,0000888242
Valor-Z para asimetría	1,08747	0,276827
Valor-Z para curtosis	-3,96388	0,000073775

Tabla 22 Pruebas Normalidad Residuos Parados.

Test Hipótesis normalidad residuos.

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

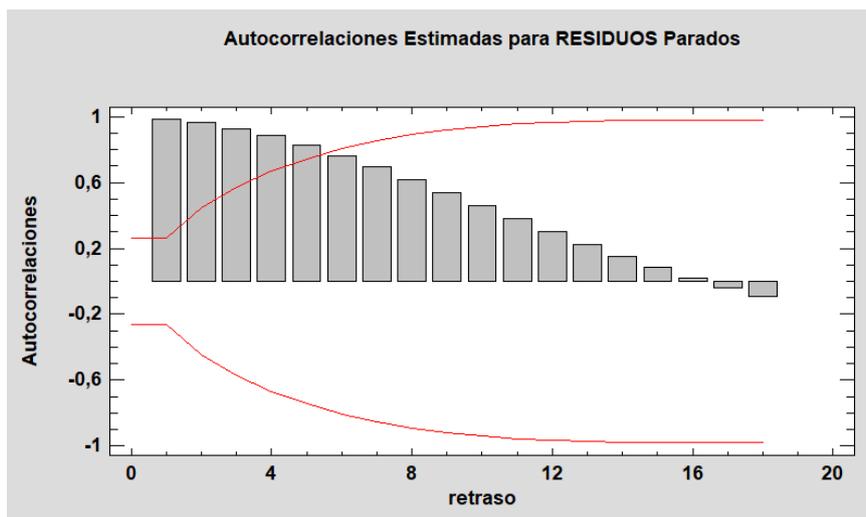
En la tabla 22, como P-Valor, 0,000073775, es menor que alfa, 0,05, se rechaza la hipótesis nula, y se concluye que los residuos no se distribuyen de una forma normal.

Por último, para el análisis de la variable Parados se va a estudiar las autocorrelaciones de esta variable.

Retraso	Autocorrelación	Error Estd.	Límite en 95,0% Inferior	Límite en 95,0% Superior
1	0,985246	0,133631	-0,261912	0,261912
2	0,962477	0,229184	-0,449193	0,449193
3	0,928403	0,292591	-0,57347	0,57347
4	0,884521	0,341164	-0,668671	0,668671
5	0,826991	0,379914	-0,74462	0,74462
6	0,762998	0,410805	-0,805164	0,805164
7	0,693733	0,435376	-0,853323	0,853323
8	0,616734	0,454687	-0,891172	0,891172
9	0,538146	0,469387	-0,919984	0,919984
10	0,460884	0,480278	-0,94133	0,94133
11	0,381948	0,488112	-0,956684	0,956684
12	0,302709	0,493421	-0,967088	0,967088
13	0,226201	0,496726	-0,973566	0,973566
14	0,154146	0,498562	-0,977165	0,977165
15	0,0847201	0,499412	-0,978831	0,978831
16	0,0217434	0,499669	-0,979334	0,979334
17	-0,0376096	0,499686	-0,979367	0,979367
18	-0,0929458	0,499736	-0,979467	0,979467

Tabla 23 Autocorrelación Residuo Parados.

En la tabla 23, se observa que existe una autocorrelación de nivel 5 para los residuos Parados.



Gráfica 37 Autocorrelación Residuos Parados

En la gráfica 37, como se ha podido observar, en ambos lados existe una autocorrelación de nivel 5.

8.5.- Modelo de regresión simple: IPV en función de la variable Viviendas Iniciadas

En tercer lugar, se va a estudiar cómo afecta la variable X_5 Viviendas Iniciadas sobre la variable endógena IPV. Se plantea el siguiente modelo de regresión lineal simple **IPV = B0 + B1 * VIVENDAS INICIADAS + U**.

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	107,187	2,79054	38,4107	0,0000
Vivi Iniciadas	0,0036221	0,00100032	3,62094	0,0007

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	2532,9	1	2532,9	13,11	0,0007
Residuo	9659,33	50	193,187		
Total (Corr.)	12192,2	51			

Tabla 24 Análisis Varianza VivIni.

R-cuadrada = 20,7747 por ciento

R-cuadrado (ajustado para g.l.) = 19,1902 por ciento

Validación Modelo

•Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCF}{k}}{\frac{SCF}{n-k}}$$

En la tabla 24, como el P-Valor es 0,0007 es inferior a Alfa, por tanto, se rechaza la hipótesis nula y se concluye afirmando que el modelo es significativo.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, la constante es significativa.

P- Valor Viviendas Iniciadas = 0,0007 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, las viviendas iniciadas es un parámetro significativo.

Si se observa el R² ajustado este muestra que es del 19.19% es el peso que tiene la variable Vivilnici sobre la variable endógena IPV. Este dato viene a indicar que esta variable afectará poco a la variable IPV debido al porcentaje que tiene.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de Viviendas Iniciadas.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	38,6154	0,00123555
Estadístico W de Shapiro-Wilk	0,855226	0,000015483
Valor-Z para asimetría	1,55392	0,120202
Valor-Z para curtosis	-2,07271	0,0381988

Tabla 25 Pruebas Normalidad Residuos Vivlni.

Test normalidad de residuos.

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

En la tabla 25, como P-Valor, 0,000015483 es menor que alfa, 0,05, se rechaza la hipótesis nula, con lo que se concluye que los residuos no se distribuyen de forma normal.

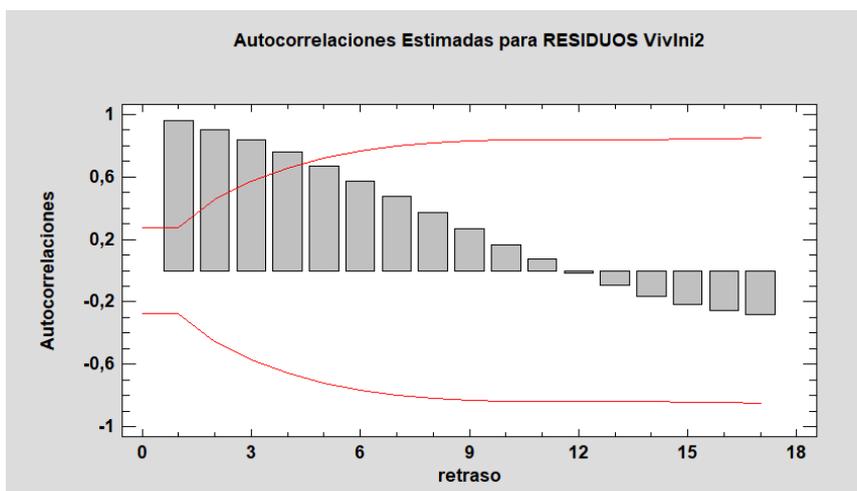
Por último, para el análisis de la variable Viviendas Iniciadas se va a estudiar las autocorrelaciones de esta variable.

			Límite en 95,0%	Límite en 95,0%
Retraso	Autocorrelación	Error Estd.	Inferior	Superior
1	0,958129	0,138675	-0,271799	0,271799
2	0,89923	0,233536	-0,457722	0,457722
3	0,840036	0,292642	-0,573569	0,573569
4	0,76153	0,335828	-0,658212	0,658212
5	0,671146	0,367539	-0,720365	0,720365
6	0,572204	0,390397	-0,765165	0,765165
7	0,472475	0,406205	-0,796149	0,796149
8	0,372737	0,41664	-0,8166	0,8166

9	0,267509	0,423004	-0,829074	0,829074
10	0,164781	0,426245	-0,835426	0,835426
11	0,0713225	0,427468	-0,837823	0,837823
12	-0,0150732	0,427697	-0,838272	0,838272
13	-0,0949102	0,427707	-0,838292	0,838292
14	-0,16239	0,428112	-0,839085	0,839085
15	-0,21708	0,429295	-0,841404	0,841404
16	-0,255445	0,4314	-0,845531	0,845531

Tabla 26 Autocorrelación Residuos Vivlni.

En la tabla 26, se puede observar que para los Residuos Viviendas Inciadas, existe autocorrelación de nivel 4.



Gráfica 38 Autocorrelación Residuo Vivlni

En la gráfica 38, como se ha podido observar tanto en la tabla como a nivel gráfico, existe una autocorrelación de nivel 4 en esta variable.

Las estimaciones de los parámetros utilizando mínimos cuadrados ordinarios dejan de ser eficientes, ya que no es el estimador lineal e insesgado de mínima varianza, el estimador que tendría esas propiedades sería el que se obtiene aplicando mínimos cuadrados generalizados

8.6.- Modelo de regresión simple: IPV en función de la variable Densidad Población

Se plantea el siguiente modelo de regresión lineal $IPV = B_0 + B_1 * DENSIDAD POBLACIÓN + U$

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	-300,317	162,733	-1,84547	0,0714
DensPobla	0,0000823052	0,0000323009	2,54808	0,0142

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	1504,85	1	1504,85	6,49	0,0142
Residuo	10661,7	46	231,776		
Total (Corr.)	12166,5	47			

Tabla 27 Análisis Varianza Densidad Población.

R-cuadrada = 12,3688 por ciento

R-cuadrado (ajustado para g.l.) = 10,4637 por ciento

Validación Modelo

- Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCE}{k-1}}{\frac{SCR}{n-k}}$$

En la tabla 27, como el P-Valor es 0,0142 es inferior a Alfa, por tanto se rechaza la hipótesis nula y se concluye afirmando que el modelo es significativo.

Validación Parámetros.

- Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0714 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto el parámetro no es significativa.

P- Valor Densidad población = 0,0142 es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro densidad de población es significativo.

Si se observa el R^2 ajustado este muestra que el 10.4637% es el peso que tiene la variable DensPobla sobre la variable endógena IPV. Este dato viene a indicar que esta variable afectará sobre la variable IPV, pero menos que otras variables que se han ido viendo anteriormente, esto debido al porcentaje pequeño que tiene.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de Densidad Población.

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	12,0	0,679029
Estadístico W de Shapiro-Wilk	0,962573	0,128258
Valor-Z para asimetría	1,00399	0,315383
Valor-Z para curtosis	-0,127119	0,898841

Tabla 28 Prueba Normalidad Residuos Densidad Población.

Test normalidad residuos

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

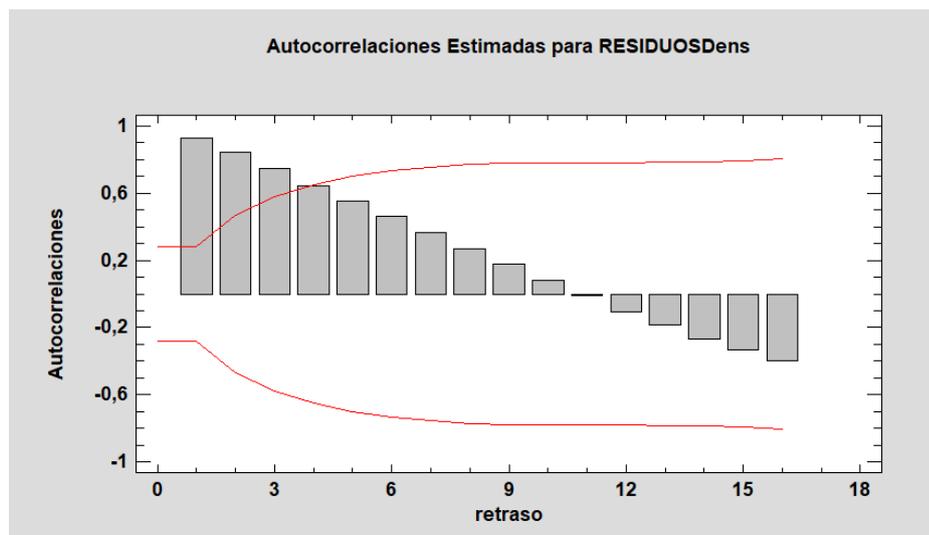
En la tabla 28, como P-Valor, 0,128258 es mayor que alfa, 0,05, no existe evidencia para rechazar la hipótesis alternativa, con lo que se concluye que los residuos se distribuyen de forma normal.

Por último, para el análisis de la variable Densidad Población se va a estudiar las autocorrelaciones de esta variable.

<i>Retraso</i>	<i>Autocorrelación</i>	<i>Error Estd.</i>	<i>Límite en 95,0% Inferior</i>	<i>Límite en 95,0% Superior</i>
1	0,928194	0,144338	-0,282897	0,282897
2	0,843429	0,238183	-0,466831	0,466831
3	0,745185	0,29389	-0,576016	0,576016
4	0,645064	0,330922	-0,648596	0,648596
5	0,555176	0,356156	-0,698054	0,698054
6	0,461229	0,37375	-0,732539	0,732539
7	0,366624	0,385426	-0,755423	0,755423
8	0,269072	0,392624	-0,769531	0,769531
9	0,17957	0,396447	-0,777024	0,777024
10	0,0809369	0,398138	-0,780338	0,780338
11	-0,0116563	0,398481	-0,78101	0,78101
12	-0,103855	0,398488	-0,781023	0,781023
13	-0,185414	0,399051	-0,782128	0,782128
14	-0,266136	0,400842	-0,785638	0,785638
15	-0,33455	0,404507	-0,79282	0,79282
16	-0,396231	0,410231	-0,804039	0,804039

Tabla 29 Autocorrelación Residuos DensPobla.

En la tabla 29, autocorrelación residuos denspobla, se aprecia que existe una autocorrelación de nivel 3.



Gráfica 39 Autocorrelación Residuos DensPobla

En la gráfica 39, como se ha observado tanto en la tabla como a nivel gráfico, existe una autocorrelación de nivel 3 en esta variable.

8.7.- Modelo de regresión simple: IPV en función de la variable PIB

En cuarto lugar, se va a estudiar cómo afecta la variable X_7 PIB sobre la variable endógena IPV. Planteando el modelo de regresión lineal simple $IPV = B_0 + B_1 * PIB + U$.

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	109,382	33,1616	3,29846	0,0017
PIB	0,0000258028	0,000120325	0,214442	0,8310

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	12,8478	1	12,8478	0,05	0,8310
Residuo	15087,1	54	279,39		
Total (Corr.)	15099,9	55			

Tabla 30 Análisis Varianza PIB.

R-cuadrada = 0,0850856 por ciento

R-cuadrado (ajustado para g.l.) = 0 por ciento

Validación Modelo

•Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCF}{k}}{\frac{SCF}{n-k}}$$

En la tabla 30, como el P-Valor es 0,8310 es superior a Alfa, por tanto, no existe evidencia para rechazar la hipótesis alternativa, y por consiguiente el modelo no es válido.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0017, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, la constante es significativa.

P- Valor Parados = 0,8310 es superior a Alfa, por tanto, no existe evidencia para rechazar la hipótesis alternativa y el parámetro parados no es significativo.

Si se observa el R² ajustado este muestra que el 0% es el peso que tiene la variable PIB sobre la variable endógena IPV. Este dato viene a indicar que esta variable no afectará sobre la variable IPV debido al porcentaje que tiene.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de PIB.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	34,9286	0,0040647
Estadístico W de Shapiro-Wilk	0,889442	0,0000950829
Valor-Z para asimetría	1,21183	0,225575
Valor-Z para curtosis	-3,30353	0,000954872

Tabla 31 Normalidad Residuos PIB.

Test normalidad residuos.

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

En la tabla 31, como P-Valor, 0,0000950829, es menor que alfa, 0,05, se rechaza la hipótesis nula, con lo que se concluye que los residuos no se distribuyen de forma normal.

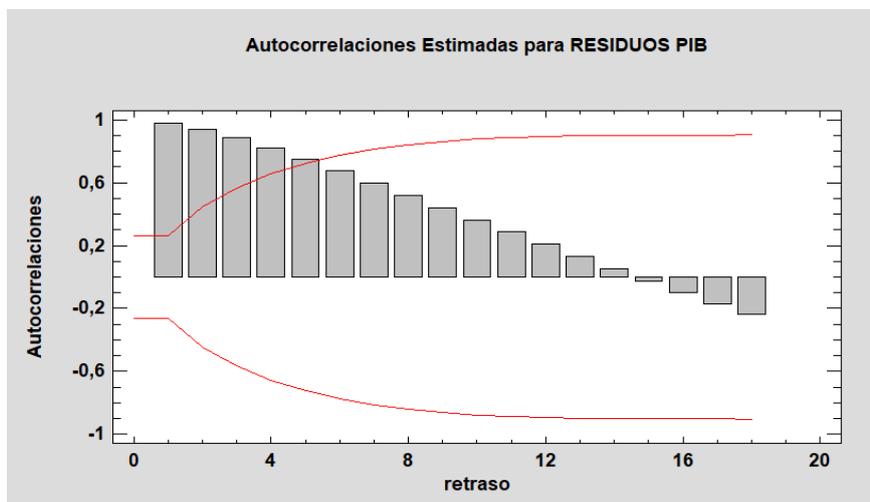
Por último, para el análisis de la variable PIB se va a estudiar las autocorrelaciones de esta variable.

			Límite en 95,0%	Límite en 95,0%
Retraso	Autocorrelación	Error Estd.	Inferior	Superior
1	0,976691	0,133631	-0,261912	0,261912
2	0,937355	0,227873	-0,446623	0,446623
3	0,885283	0,288627	-0,5657	0,5657
4	0,82288	0,33361	-0,653866	0,653866
5	0,750399	0,368075	-0,721415	0,721415
6	0,674717	0,394449	-0,773107	0,773107
7	0,596848	0,414546	-0,812497	0,812497
8	0,518593	0,429617	-0,842036	0,842036
9	0,441834	0,440654	-0,863668	0,863668

10	0,364576	0,448495	-0,879036	0,879036
11	0,286752	0,453756	-0,889348	0,889348
12	0,20822	0,456981	-0,895668	0,895668
13	0,130674	0,458672	-0,898982	0,898982
14	0,0495004	0,459336	-0,900284	0,900284
15	-0,0277171	0,459432	-0,900471	0,900471
16	-0,101718	0,459461	-0,90053	0,90053
17	-0,170348	0,459863	-0,901317	0,901317
18	-0,234043	0,460989	-0,903523	0,903523

Tabla 32 Autocorrelación Residuos PIB.

En la tabla 32, se observa que existe una autocorrelación de nivel 5, sobre la variable PIB.



Gráfica 40 Autocorrelación Residuos PIB

En la gráfica 40, como se ha podido observar tanto en la tabla como a nivel gráfico, existe una autocorrelación de nivel 5 en esta variable.

8.8.- Modelo de regresión simple: IPV en función de la variable M²

Se va a estudiar cómo afecta la variable X₈ M² sobre la variable endógena IPV. Se plantea el siguiente modelo de regresión lineal simple **IPV=B0+B1*M²+U**.

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	75,2691	3,42932	21,9487	0,0000
M ²	0,196423	0,0154527	12,7113	0,0000

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	11317,5	1	11317,5	161,58	0,0000
Residuo	3782,39	54	70,0442		
Total (Corr.)	15099,9	55			

Tabla 33 Analisis Varianza M²

R-cuadrada = 74,9509 por ciento

R-cuadrado (ajustado para g.l.) = 74,487 por ciento

Validación Modelo

- Modelo

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \text{Al menos uno es } \neq 0 \quad F_{calc} = \frac{CME}{CMR} = \frac{\frac{SCE}{k-1}}{\frac{SCE}{n-k}}$$

En la tabla 33, como el P-Valor es 0,0000 es inferior a Alfa, por se rechaza la hipótesis nula y se concluye indicando que el modelo es significativo.

Validación Parámetros.

- Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor constante = 0,0000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, la constante es significativa.

P- Valor M² = 0,000 es superior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro M² si es significativo.

Si se observa el R² ajustado este muestra que el 74,48% es el peso que tiene la variable Parados sobre la variable endógena IPV.

A continuación, se procede a pasar a observar las pruebas de normalidad de los residuos de M².

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	12,5357	0,706353
Estadístico W de Shapiro-Wilk	0,962434	0,0786151
Valor-Z para asimetría	0,83023	0,406407
Valor-Z para curtosis	-1,52666	0,126846

Tabla 34 Pruebas Normalidad Residuos M²

Test normalidad residuos

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

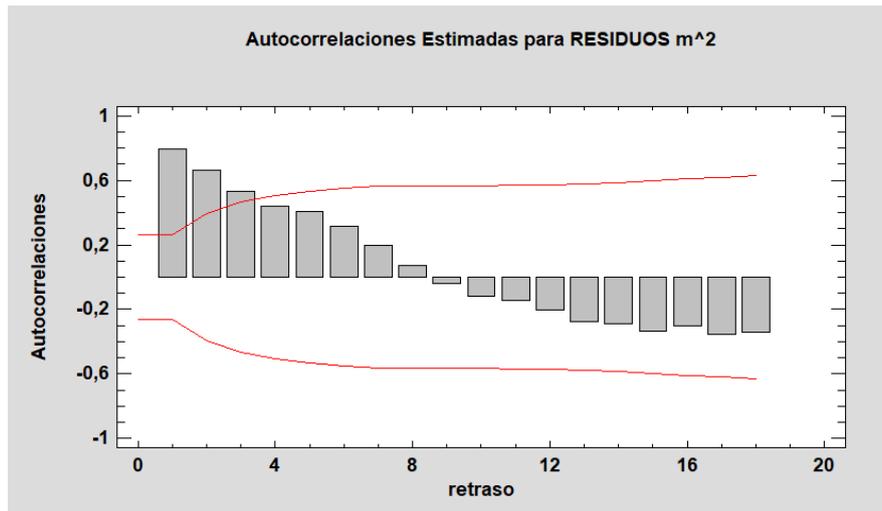
En la tabla 34, como P-Valor, 0,0786151, es mayor o igual a alfa, 0,05, no existe evidencia para rechazar la hipótesis alternativa, con lo que se concluye que los residuos se distribuyen de forma normal.

Por último, para el análisis de la variable M² se va a estudiar las autocorrelaciones de esta variable.

<i>Retraso</i>	<i>Autocorrelación</i>	<i>Error Estd.</i>	<i>Límite en 95,0% Inferior</i>	<i>Límite en 95,0% Superior</i>
1	0,792996	0,133631	-0,261912	0,261912
2	0,662737	0,200788	-0,393538	0,393538
3	0,52887	0,236648	-0,463822	0,463822
4	0,440976	0,256888	-0,503493	0,503493
5	0,405243	0,270068	-0,529324	0,529324
6	0,313876	0,280716	-0,550195	0,550195
7	0,195116	0,286915	-0,562344	0,562344
8	0,0731362	0,289275	-0,566969	0,566969
9	-0,0407272	0,289605	-0,567616	0,567616
10	-0,116551	0,289707	-0,567816	0,567816
11	-0,144376	0,290543	-0,569455	0,569455
12	-0,204557	0,291821	-0,571961	0,571961
13	-0,275251	0,294371	-0,576957	0,576957
14	-0,287275	0,298931	-0,585896	0,585896
15	-0,336956	0,303821	-0,59548	0,59548
16	-0,302195	0,310423	-0,608419	0,608419
17	-0,35385	0,315632	-0,618629	0,618629
18	-0,340882	0,322639	-0,632361	0,632361

Tabla 35 Autocorrelaciones Residuos M².

En la tabla 35, se observa que existe una autocorrelación de nivel 3 para los residuos de la variable M².



Gráfica 41 Autocorrelaciones Estimadas Residuos M^2 .

En la gráfica 41, como se ha podido observar tanto en la tabla como a nivel gráfico, existe una autocorrelación de nivel 3 en esta variable.

8.9.-Conclusiones sobre el análisis bivariante.

Como conclusión del apartado se puede decir que el análisis bivariante sirve para conocer en mejor medida que variables afectan tienen una mayor relación con la variable IPV.

Tal y como se ha podido observar, las variables que más relación tienen son las siguientes:

1. La variable M^2 con un 74%
2. La variable Parados con un 43%
3. La variable Nacimientos con un 49%
4. La variable Euribor con un 60%

A priori, estas variables son las que más relación tienen, aunque más adelante se hará un análisis de multicolinealidad para verificar si existe o no alguna relación lineal entre las variables independientes.

9.-Series de Tiempo

⁴Una serie de tiempo es una sucesión de datos de una variable ordenada en el tiempo, como en el caso de este trabajo, la variable Índice de Precio de las Viviendas de manera trimestral.

Pero lo interesante de una serie temporal es poder predecir valores en un futuro mediante el estudio del comportamiento pasado.

Es por eso por lo que las series de tiempo son importantes para analizar en un futuro si se cumplen todas las pautas que se verán a continuación, predicciones de cómo va a fluctuar la variable estudiada.

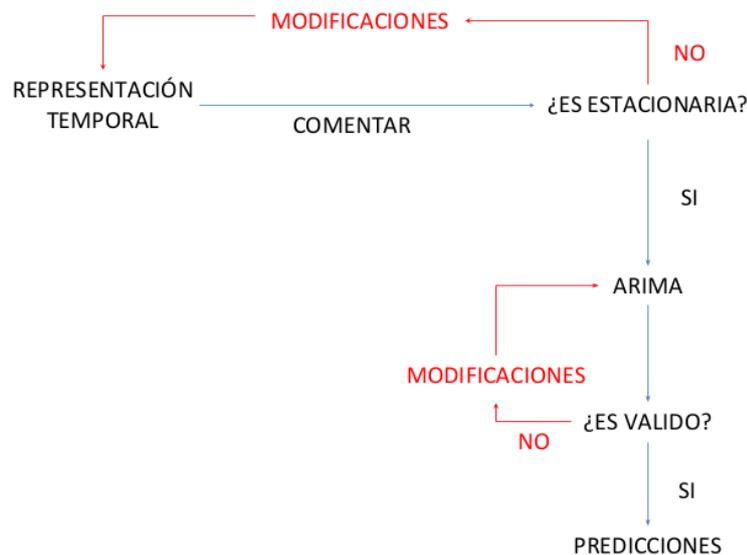


Imagen 1 Resumen Serie Temporal.

Como se aprecia en la Imagen 1 es un breve esquema de lo que se tiene que analizar en cada una de las variables, en el caso de este trabajo se va a realizar con las variables anteriormente estudiadas y además ordenadas cronológicamente, para poder construir el modelo y así poder describir el comportamiento histórico.

A continuación, se va a realizar todos análisis de series de tiempo (Modelo ARIMA) para cada una de nuestras variables X.

En primer lugar, hay que destacar que se quiere conseguir que una serie de tiempo llegue a ser estacionaria para ello tiene que cumplir los siguientes pasos:

1. Debe tener tendencia nula o constante
2. No tener variaciones estacionales
3. Tener la varianza constante

⁴ Apuntes Econometría UPV 2021

4. Tener autocorrelación

En caso de que no cumpla alguno de los pasos, se tendrán que hacer modificaciones en la variable hasta que se cumplan los pasos anteriormente citados.

Una vez se cumplan los 4 pasos, ya se puede realizar el modelo ARIMA (p,d,q) x (P,D,Q), siendo cada una de las letras las siguientes connotaciones:

P= Orden que se le aplica a AR

d= Orden no estacional que será entre (0,1,2,3), esto variará dependiendo de lo que cueste suavizar la tendencia.

q= Orden que se le aplica a MA

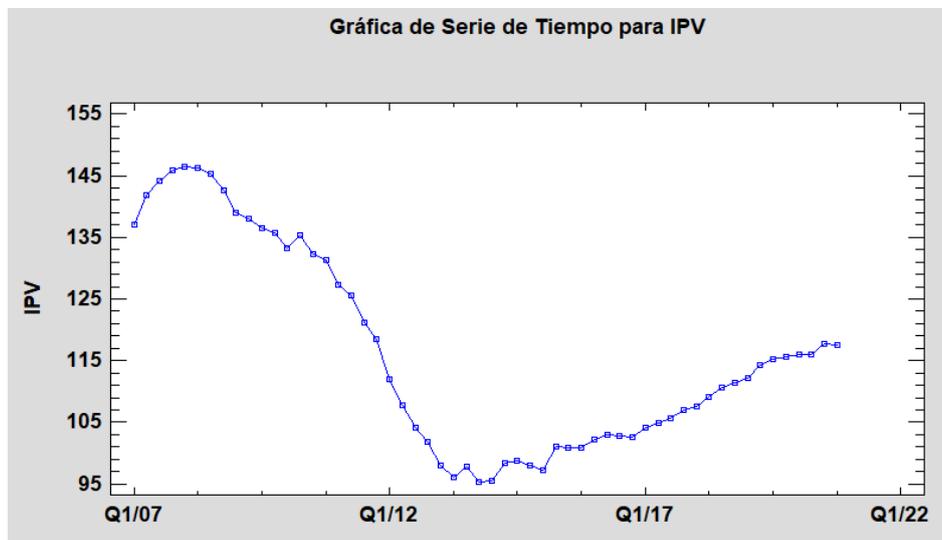
P= Orden que se le aplica al SAR cuando la variable tiene estacionalidad.

D= Orden estacional que será entre (0,1,2) dependiendo de lo que cueste suavizar la estacionalidad.

Q= Orden que se le aplica a SMA

Una vez explicado el Modelo ARIMA, se buscará el correspondiente, pero cabe destacar que es imprescindible disponer de una serie totalmente estacionaria. La propuesta del modelo ARIMA vendrá de los gráficos FAS y FAP de la serie estacionaria.

9.1.-Modelo serie temporal IPV



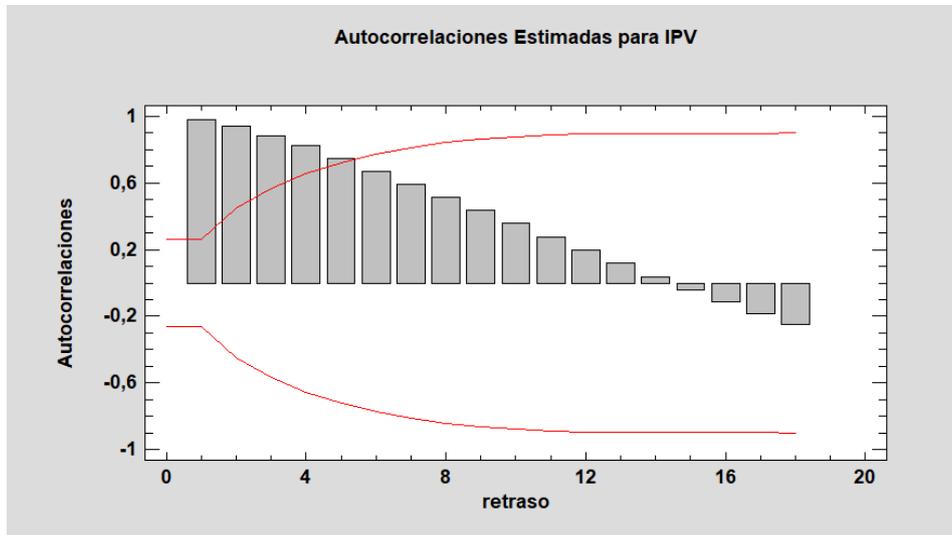
Gráfica 42 Serie Tiempo IPV 1

Como se puede observar en esta grafica está compuesta por 2 tramos, un primer tramo de tendencia descendente (negativa) y una tendencia ascendente (positiva) a partir del primer trimestre de 2015.

En segundo lugar, se puede observar la varianza es constante ya que no se ve una forma cónica, que pueda apreciarse a simple vista.

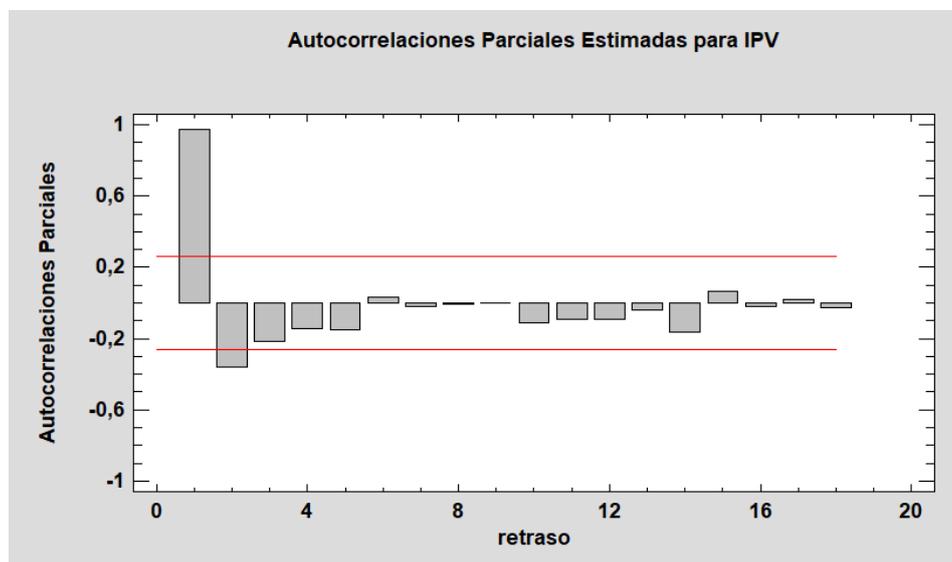
En tercer lugar, se mira la estacionalidad, aunque a simple vista se puede apreciar que no existe estacionalidad

En cuarto lugar, las autocorrelaciones del FAS Y FAP.



Gráfica 43 FAS IPV

En la gráfica 43, en el FAS se puede apreciar que existe autocorrelación.



Gráfica 44 FAP IPV

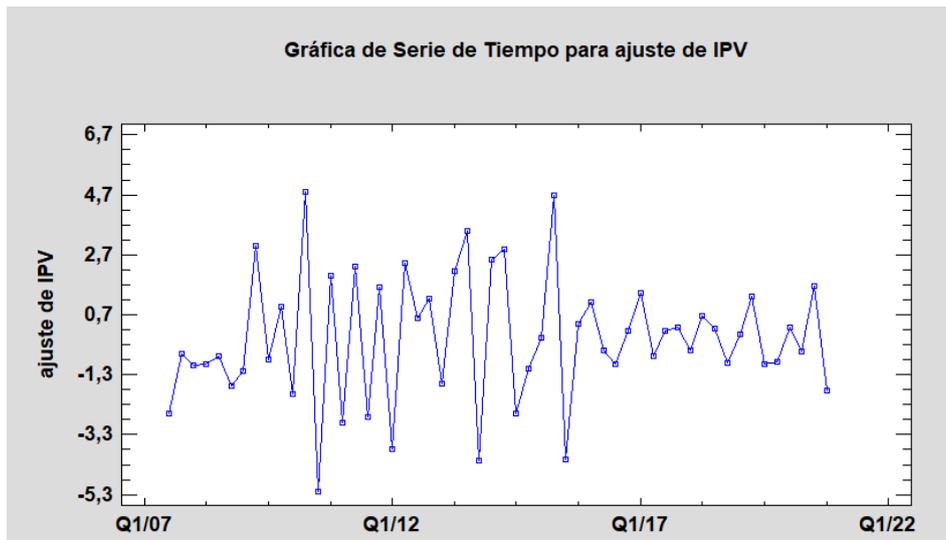
En la gráfica 44, en el FAP se aprecia que existe autocorrelación.

Mirando el comportamiento de la serie se concluye que la serie no es estacionaria, por tanto, se debe plantear una serie de modificaciones para conseguir que la serie sea estacionaria.

Orden para conseguir que la serie sea estacionaria es:

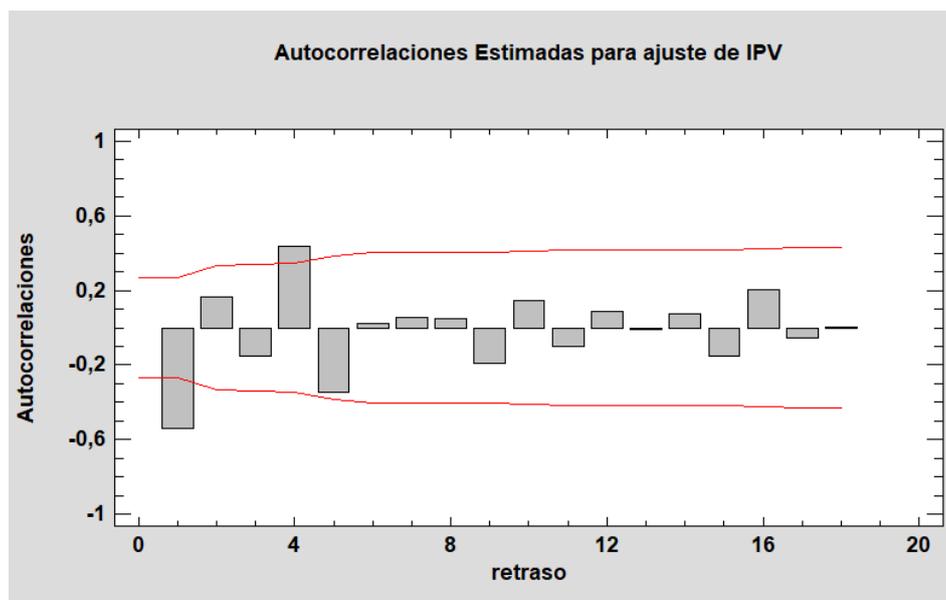
1. Conseguir la varianza sea constante
2. Conseguir una tendencia nula o constante
3. No tener variaciones estacionales
4. Tener autocorrelación

Para que la serie sea estacionaria solo debe solucionarse el problema de la tendencia aplicando 2 diferenciaciones de orden no estacional, presentando el siguiente resultado gráfico.

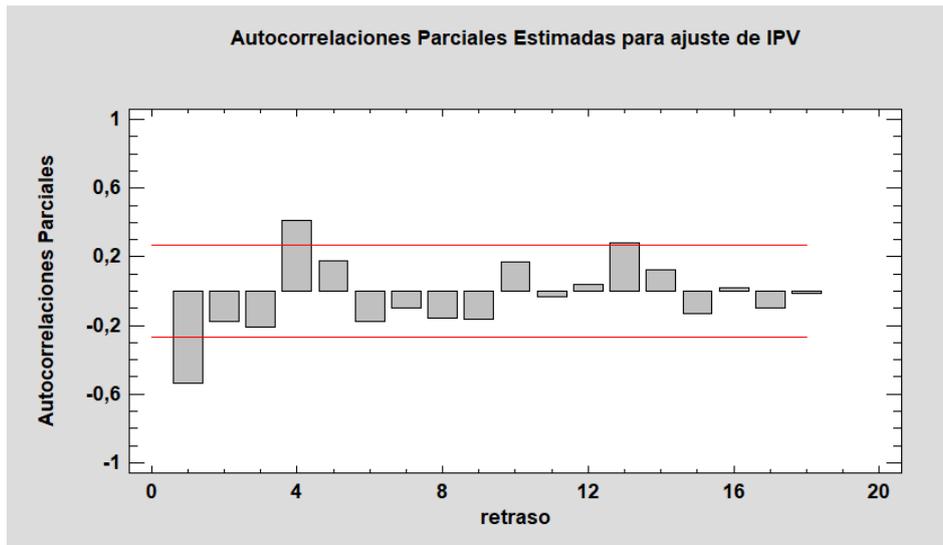


Gráfica 45 Serie Tiempo Ajustada 2 IPV.

En la gráfica 45, se puede observar que se suaviza la tendencia, siendo esta una tendencia más suave y siguiendo una tendencia más o menos neutral.



Gráfica 46 FAS Ajustado IPV



Gráfica 47 FAP Ajustado IPV

Tanto en n la gráfica 46 y 47, el FAS y el FAP, se observa que se van consiguiendo unos valores significativos en los dos casos, se afirma esto debido a que las autocorrelaciones no son tan grandes y por lo tanto consigue mejor significación.

A continuación, se va a pasar a observar la validación del modelo **ARIMA (0,2,2) sin constante**.

Parámetro	Estimado	Error Estd.	t	Valor-P
MA(1)	0,93102	0,113587	8,19652	0,000000
MA(2)	-0,621881	0,114796	-5,41729	0,000002

Tabla 36 Resumen Modelo ARIMA IPV.

Pronóstico Histórico: sí

Varianza estimada de ruido blanco = 2,78949 con 52 grados de libertad

Desviación estándar estimada de ruido blanco = 1,67018

Número de iteraciones: 7

Validación Parámetros.

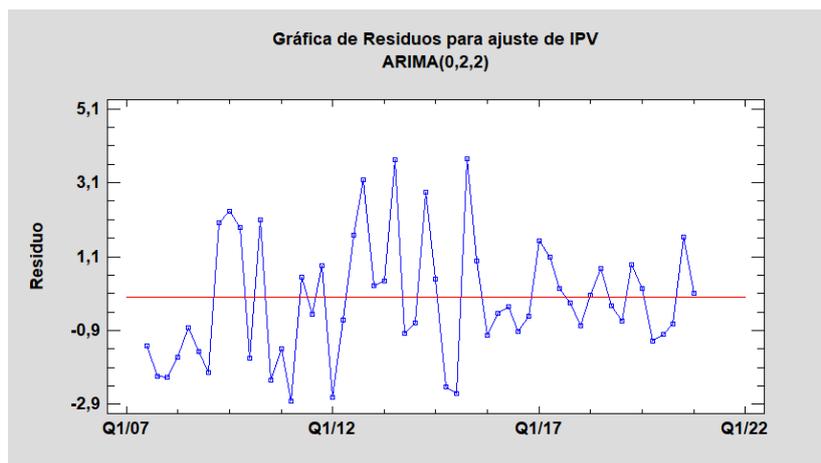
•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor MA1 = 0,000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor MA2= 0,000002 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.



Gráfica 48 Observación de residuos IPV.

En segundo lugar, en la gráfica 48, se observa si los residuos tienen o no una media nula y se observa que sí; que existe homocedasticidad, debido a que no tiene aspecto de forma cónica.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	12,1481	0,73372
Estadístico W de Shapiro-Wilk	0,969671	0,186439
Valor-Z para asimetría	1,06787	0,285578
Valor-Z para curtosis	-0,255247	0,798528

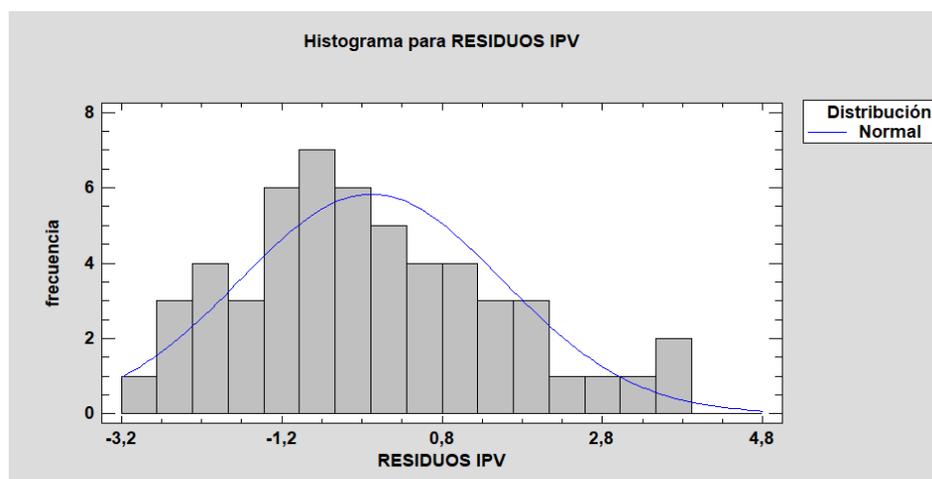
Tabla 37 Pruebas Normalidad Residuos IPV

Test normalidad residuos

H_0 : Los residuos se distribuyen normalmente

H_1 : Los residuos no se distribuyen normalmente $\alpha > 0,05$

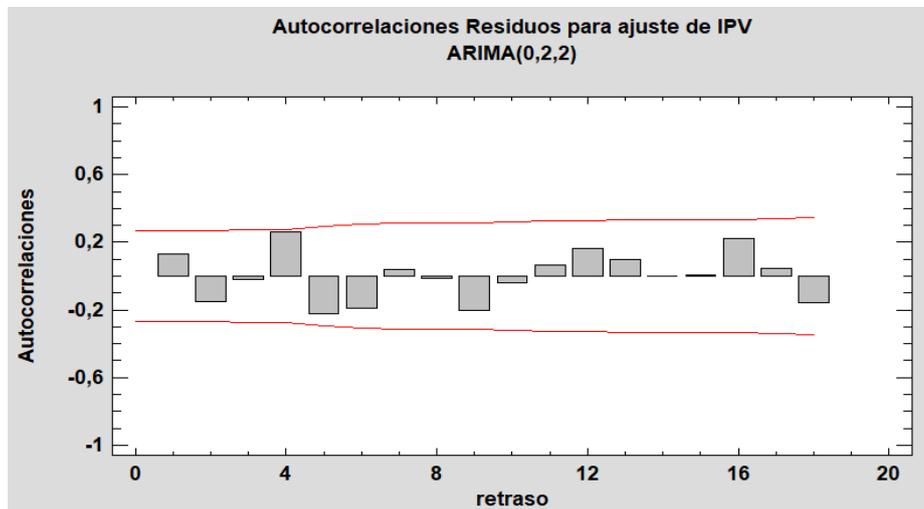
En tercer lugar, en la tabla 37, se escoge la prueba más restrictiva, que sería el estadístico de W de Shapiro-wilk con P-valor 0,156439, este dato es mayor que Alfa, por tanto, no existe evidencia para rechazar la hipótesis alternativa y se concluye que los residuos son normales.



Gráfica 49 Histograma Residuos IPV

La gráfica 49 sirve para comprobar si la normalidad de los residuos se distribuye de una manera normal, y observando la línea azul parece como se asimila a la campana de Gauss, por lo que se puede afirmar que sí.

Por último, se va a observar si existe o no autocorrelación en la FAS.



Gráfica 50 FAS Residuos IPV.

Analizando la gráfica 50 de FAS de los residuos se observa que no existe problemas de autocorrelación.

Esto indica que el modelo ARIMA (0,2,2) es válido para la variable IPV. Por lo que se va a realizar una predicción.

		Límite en 95%	
Periodo	Pronóstico	Inferior	Superior
Q1/21	118,084	114,704	121,464
Q2/21	118,751	113,805	123,696
Q3/21	119,352	111,628	127,076
Q4/21	119,887	108,58	131,193

Tabla 38 Pronósticos IPV.

Para las predicciones se tendrán que escoger como mucho hasta 3 periodos debido a que el programa de Statgraphics arrastra un error y deja de ser fiable.

Q1 Predicción puntual: 118.084

Intervalo confianza (114.704 <IPV< 121.464) 95% Confianza.

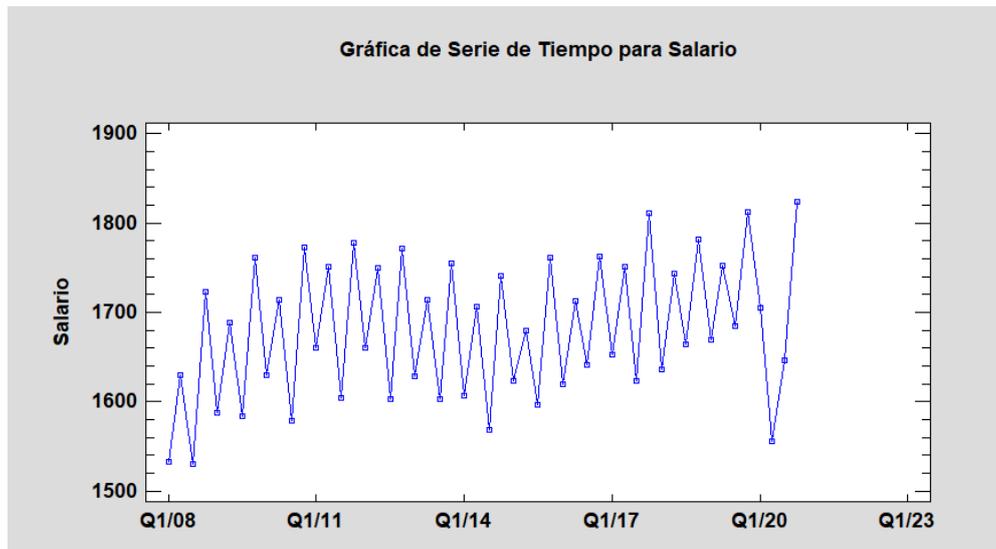
Q2 Predicción puntual: 118,751

Intervalo confianza (113.805 <IPV< 123.696) 95% Confianza.

Q3 Predicción puntual: 119,352

Intervalo confianza (111.628 <IPV< 127.076) 95% Confianza.

9.2.-Modelo serie temporal Salario



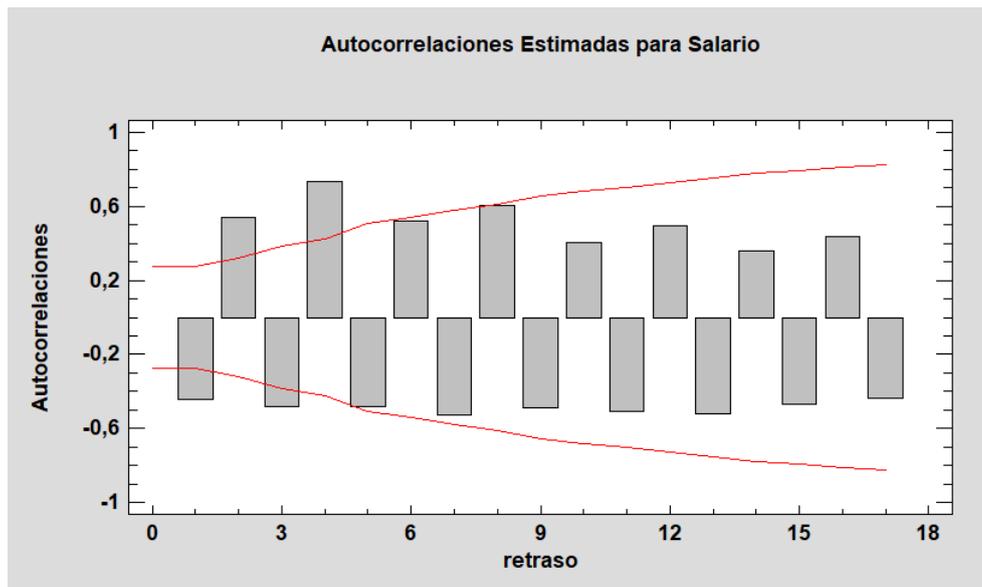
Gráfica 51 Serie Tiempo Salario

En la gráfica 51, se puede observar la tendencia de esta variable, es una tendencia positiva, por lo que ya no cumple el primer criterio de ser estacionaria, lo que habrá que hacer es transformar esta serie de tiempo en una tendencia nula, aplicando diferenciaciones de orden no estacional.

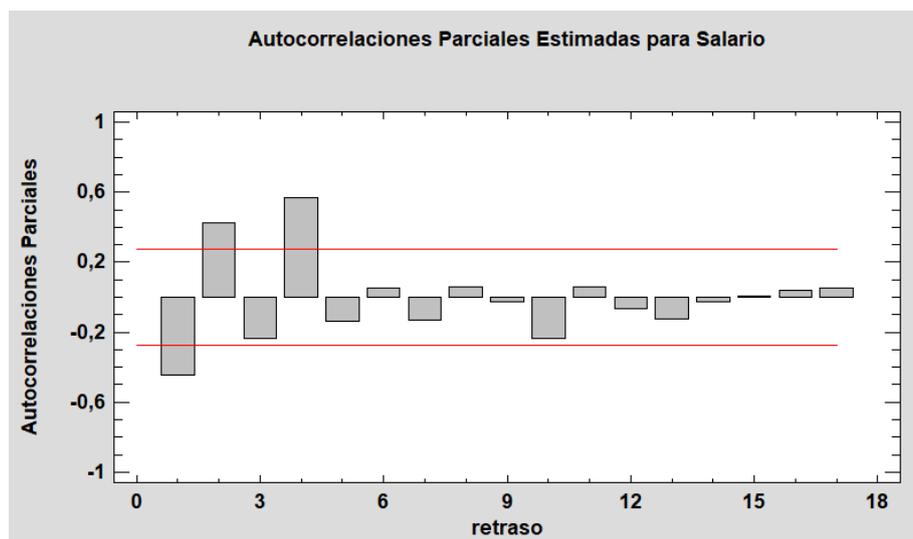
En segundo lugar, se puede observar la varianza no es constante, ya que se puede observar una forma cónica creciente a lo largo de la serie.

En tercer lugar, se mira la estacionalidad, observando que esta serie presenta variaciones estacionales para poder hacer la serie estacionaria se aplicarán diferenciaciones de orden estacional.

En cuarto lugar, tiene que existir las autocorrelaciones del FAS Y FAP.



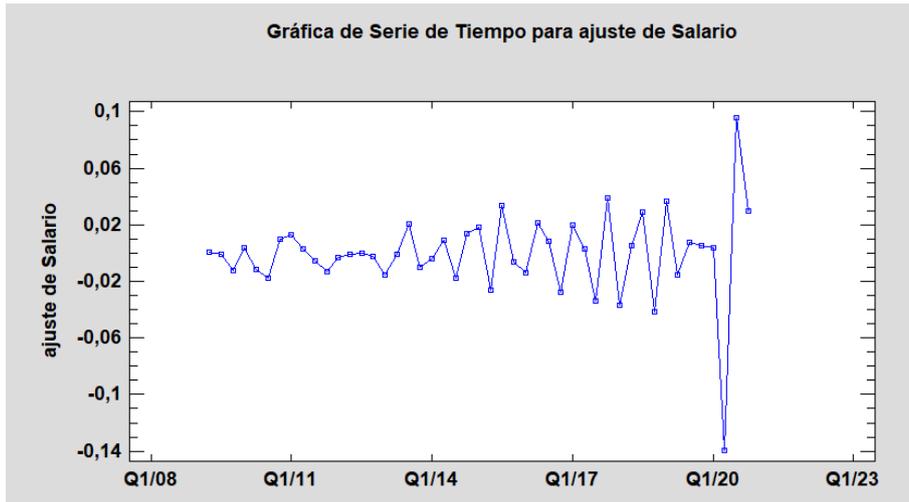
Gráfica 52 FAS Salario



Gráfica 53 FAP Salario

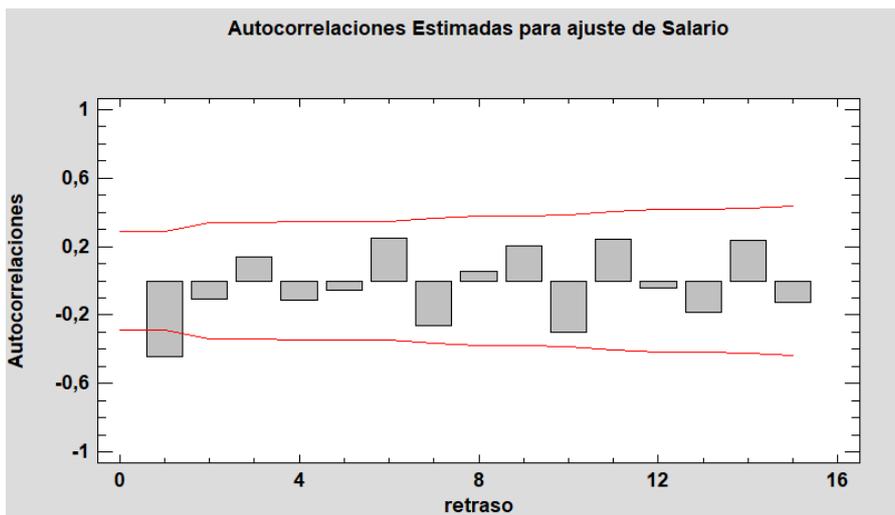
En las gráficas 52 y 53, se puede observar tanto en la FAS como en el FAP se observa cómo tiene autocorrelación.

En principio habrá que mejorar la tendencia, la estacionalidad y la heteroscedasticidad, para solucionar los problemas de heteroscedasticidad habrá que hacer la función logarítmica y para la tendencia se han realizado las transformaciones de 1 diferenciación de orden no estacional y 1 transformación de diferenciación de orden estacional.

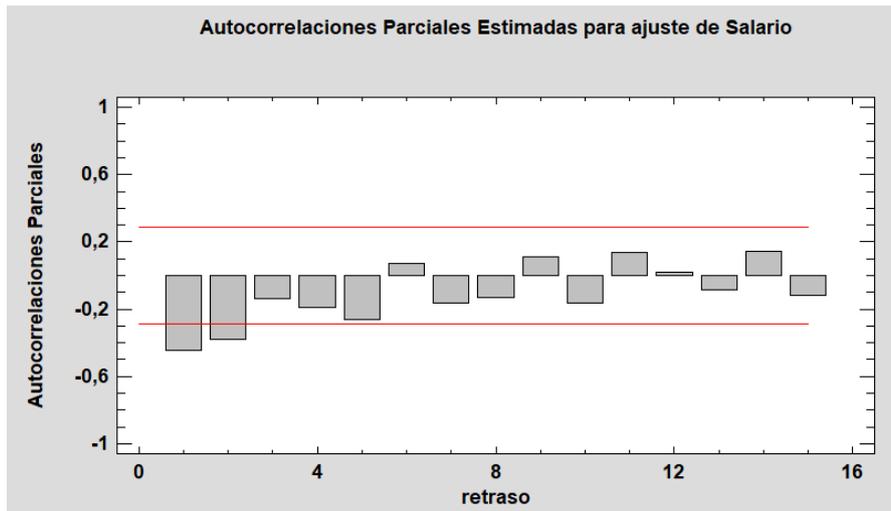


Gráfica 54 Serie Tiempo Ajustada Salario

En la gráfica 54, se observa el resultado de transformar la serie en logarítmica y hacer 1 diferenciación de orden no estacional y otra de orden estacional para conseguir hacer la serie estacionaria.



Gráfica 55 FAS Ajustado Salarios.



Gráfica 56 FAP Ajustado Salarios.

En las gráficas 55 y 56, se observan las autocorrelaciones de FAS Y FAP se observa que existe autocorrelación.

Una vez se ha conseguido que la serie sea estacionaria se plantea la realización de un modelo ARIMA.

El modelo automático que decide sacar Statgraphics, es un modelo de pronóstico seleccionado: ARIMA(1,2,2)x(1,1,2)⁴

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	0,0108374	0,171384	0,0632344	0,949940
MA(1)	1,72463	0,061254	28,1554	0,000000
MA(2)	-0,96672	0,0561749	-17,2091	0,000000
SAR(1)	0,726689	0,142563	5,0973	0,000012
SMA(1)	1,69063	0,083369	20,2789	0,000000
SMA(2)	-0,867661	0,0654872	-13,2493	0,000000

Tabla 39 Resumen Modelo ARIMA Salarios.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor AR1 = 0,949940, es superior a Alfa no existe evidencia para rechazar la hipótesis alternativa por tanto es un parámetro no significativo.

P- Valor MA1= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor MA2= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

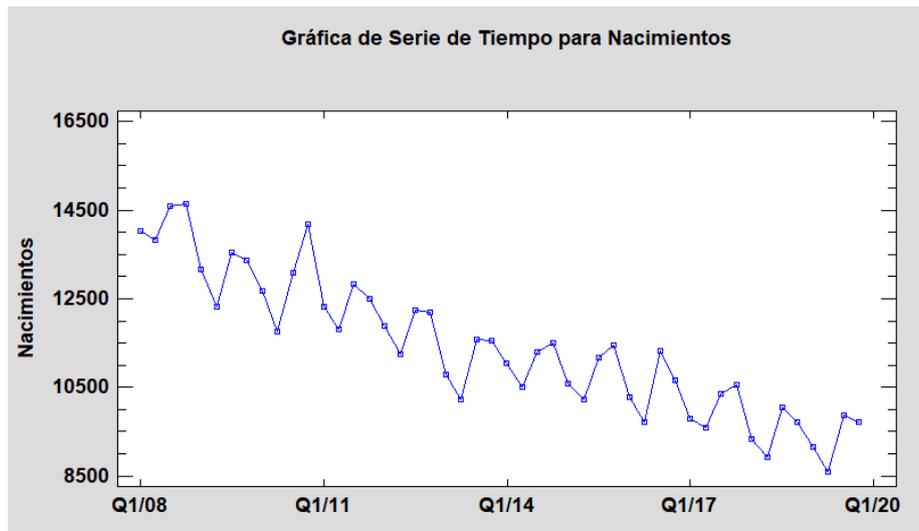
P- Valor SAR1= 0,00012 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA1= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA2= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

Dado que uno de los parámetros no es significativo, el modelo ARIMA(1,2,2)x(1,1,2)₄ no es válido.

9.3.-Modelo serie temporal Nacimientos



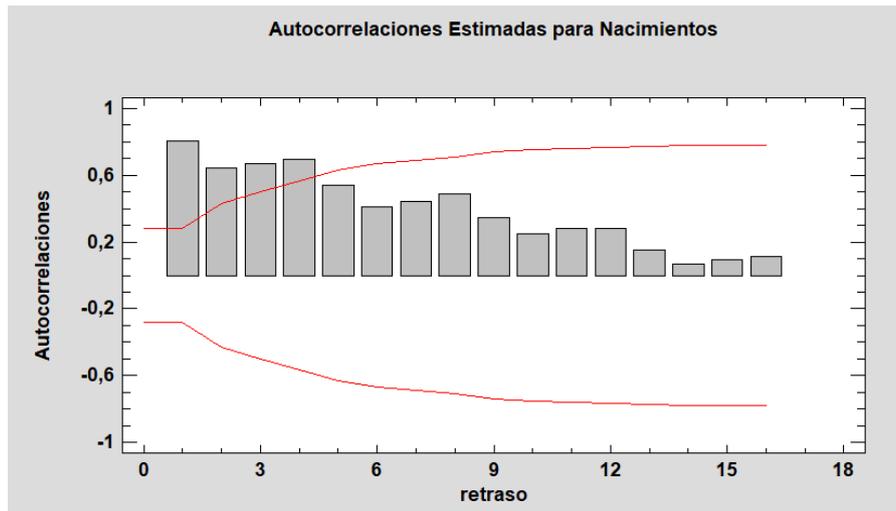
Gráfica 57 Serie Tiempo Nacimientos

En la gráfica 57, se puede observar una tendencia negativa por lo que ya no cumple uno de los criterios para ser una serie estacionaria y debido a esto habrá que aplicar diferenciaciones de orden no estacional para que la serie de tiempo de nacimientos tenga la tendencia nula.

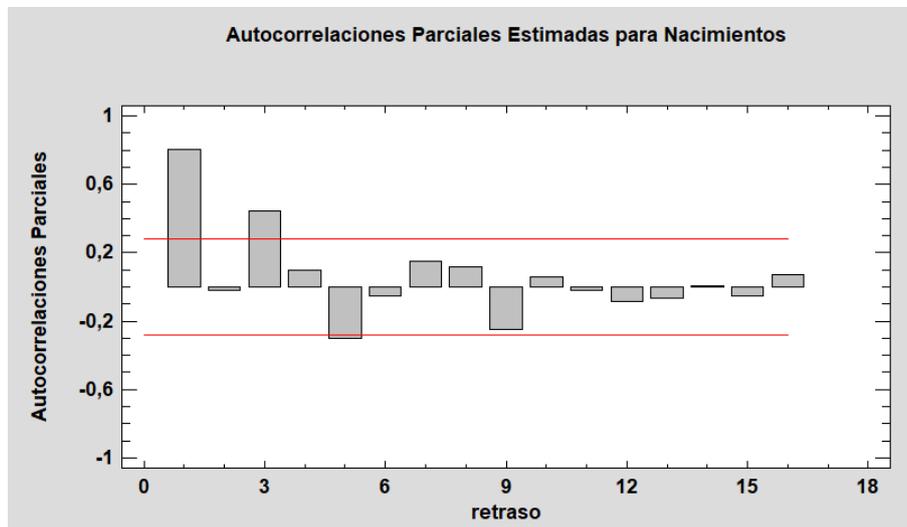
En segundo lugar, se puede observar la varianza es constante, ya que no se ve una forma cónica que pueda apreciarse a simple vista.

En tercer lugar, se observa la existencia de variaciones estacionales en los trimestres.

En cuarto lugar, las autocorrelaciones del FAS Y FAP.



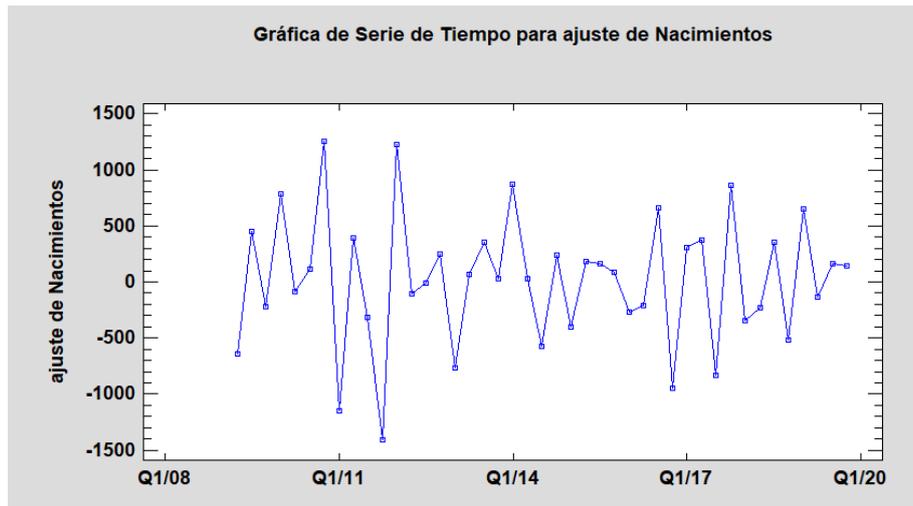
Gráfica 58 FAS Nacimiento.



Gráfica 59 FAP Nacimiento

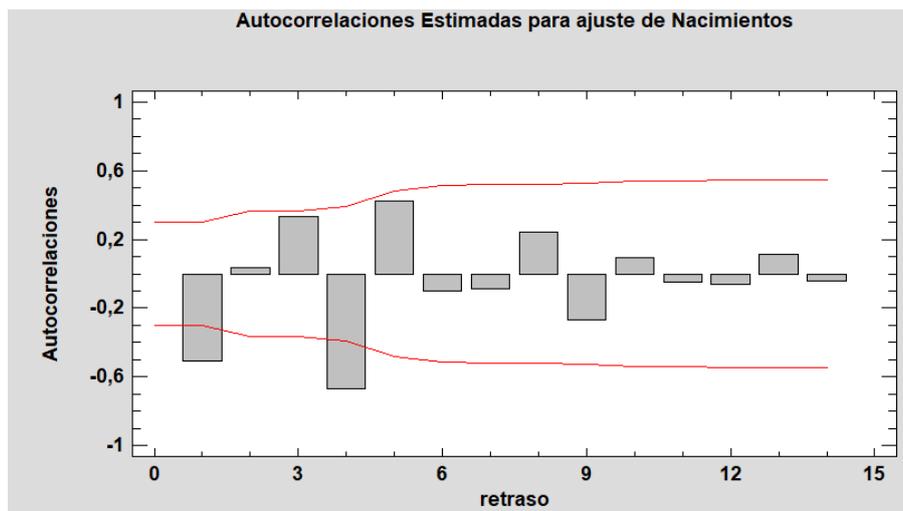
En la gráfica 58 y 59 se puede observar tanto en la FAS como en el FAP indican que existe autocorrelación.

En principio habrá que mejorar la tendencia y la estacionalidad, para solucionar la tendencia se han realizado las transformaciones de 1 diferenciación de orden no estacional y para mejorar la estacionalidad 1 transformación de diferenciación de orden estacional. Después de las modificaciones realizadas para conseguir que la serie sea estacionaria se observan las siguientes imágenes.

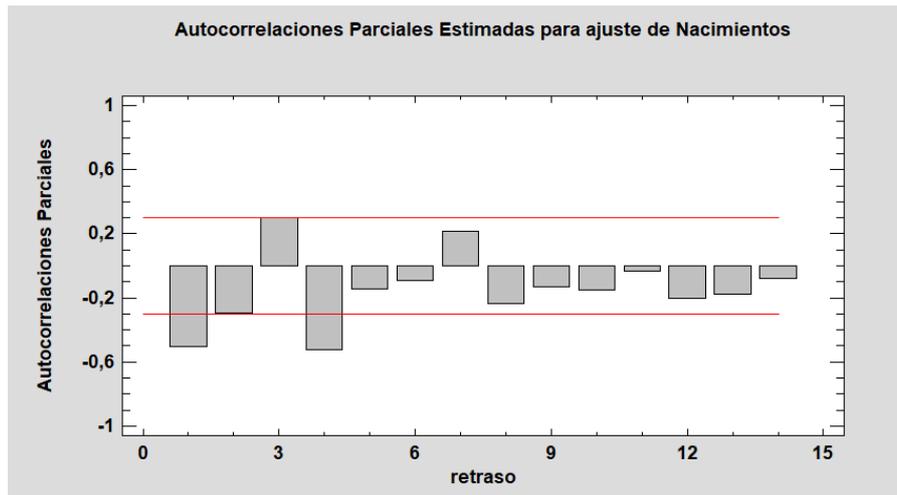


Gráfica 60 Serie Tiempo Ajustada Nacimiento

En la gráfica 60, se puede observar cómo se ha aplicado lo anteriormente comentado y se ha conseguido obtener una Serie de Tiempo bien ajustada para la variable junto con las autocorrelaciones que se verán a continuación.



Gráfica 61 FAS Ajustado Nacimiento



Gráfica 62 FAP Ajustado Nacimiento

En las gráficas 61 y 62 se observan las autocorrelación de FAS Y FAP.

Una vez se tiene la serie estacionaria se plantea un modelo ARIMA(2,0,2)*(0,1,2)⁴

El modelo de pronóstico seleccionado: ARIMA(2,0,2)x(0,1,2)⁴

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	0,331113	0,237406	1,39471	0,171203
AR(2)	0,636388	0,230505	2,76084	0,008828
MA(1)	-0,466757	0,256	-1,82327	0,076136
MA(2)	0,384899	0,142315	2,70456	0,010178
SMA(1)	1,44727	0,119594	12,1015	0,000000
SMA(2)	-0,76515	0,0914297	-8,36872	0,000000

Tabla 40 Resumen Modelo ARIMA Nacimientos.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor AR1 = 0.171203, es superior a Alfa no existe evidencia para rechazar la hipótesis alternativa por tanto es un parámetro no significativo.

P- Valor AR2= 0,008828 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

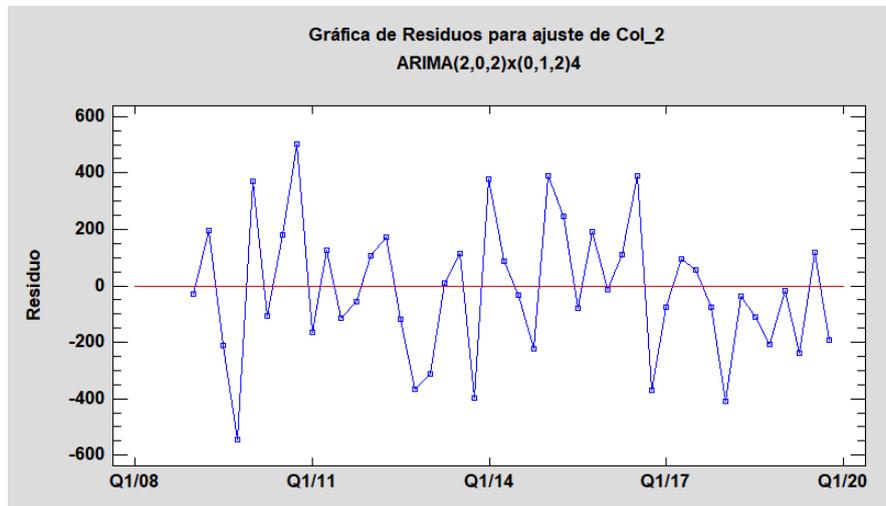
P- Valor MA1= 0,076136 es superior a Alfa, es superior a Alfa no existe evidencia para rechazar la hipótesis alternativa por tanto es un parámetro no significativo.

P- Valor MA2= 0,010178 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA1= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA2= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

De acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior los parámetros AR1 y MA1 no se podrían eliminar del modelo.



Gráfica 63 Observación de residuos Nacimientos.

En la gráfica 63, se observa que los residuos tienen media nula, representada por la línea horizontal de color rojo, y se aprecia que la serie es homogénea, ya no tiene aspecto de forma cónica.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	13,1818	0,51226
Estadístico W de Shapiro-Wilk	0,986157	0,868667
Valor-Z para asimetría	0,0886922	0,929321
Valor-Z para curtosis	-0,204194	0,838197

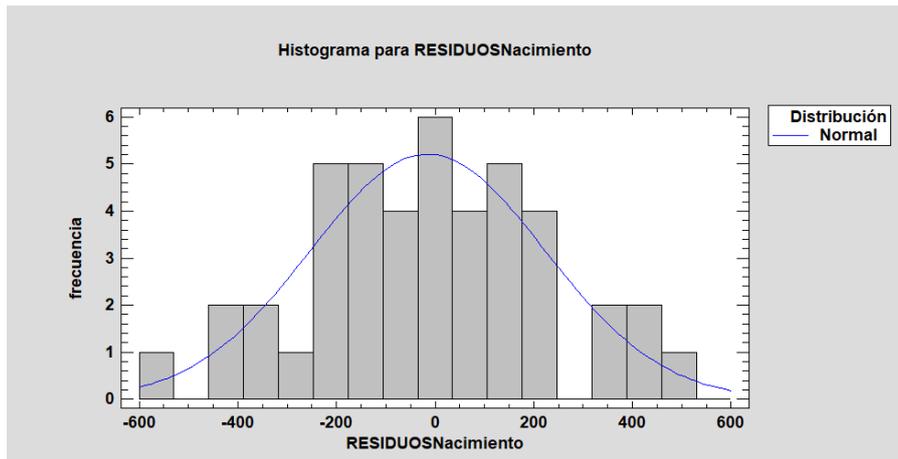
Tabla 41 Prueba Normalidad Residuos Nacimiento.

Test normalidad residuos

H0: Los residuos se distribuyen normalmente

H1: Los residuos no se distribuyen normalmente $\alpha > 0,05$

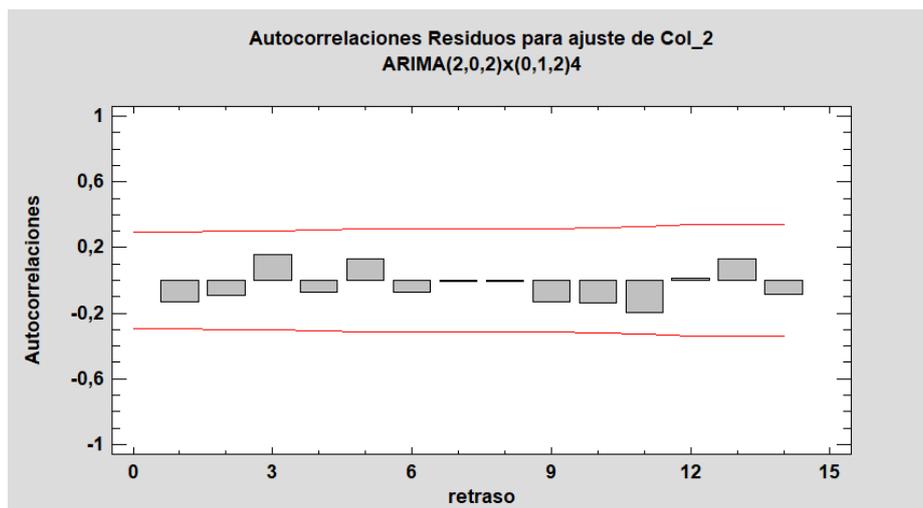
En la tabla 41, se escoge la prueba más restrictiva, que sería el estadístico de Chi-Cuadrado con P-valor 0,51226, este dato es mayor que Alfa, por tanto, no existe evidencia para rechazar la hipótesis alternativa y se concluye que los residuos son normales.



Gráfica 64 Histograma Residuos Nacimiento.

La gráfica 64 sirve para comprobar si la normalidad de los residuos se distribuye de una manera normal, y observando la línea azul parece como se asimila a la campana de Gauss, por lo que se puede afirmar que sí.

Por último, se va a observar si existe o no autocorrelación en la FAS.



Gráfica 65 FAS Residuos Nacimientos.

Analizando la Gráfica 65 del FAS de los residuos se observa que no existe problemas de autocorrelación.

Esto indica que el modelo ARIMA (2,0,2) x (0,1,2)⁴ es válido para la variable Nacimientos. Por lo que se va a realizar una predicción.

		Límite en 95%	Límite en 95%
Periodo	Pronóstico	Inferior	Superior
Q1/20	8652,86	8059,64	9246,07
Q2/20	8587,61	7828,71	9346,5
Q3/20	9566,15	8747,92	10384,4
Q4/20	9475,35	8563,48	10387,2
Q1/21	8517,44	7603,43	9431,46

Tabla 42 Pronosticos Nacimientos.

Para las predicciones se escogen 3 periodos ya que a medida que se hacen más predicciones futuras, para el cálculo de estas se tiene en cuenta la estimación anterior, generando un error que se va arrastrando y acumulando conforme se hacen más predicciones.

Q1 Predicción puntual: 8.652,86

Intervalo confianza (8.059,64<IPV< 9.246,07) 95% Confianza.

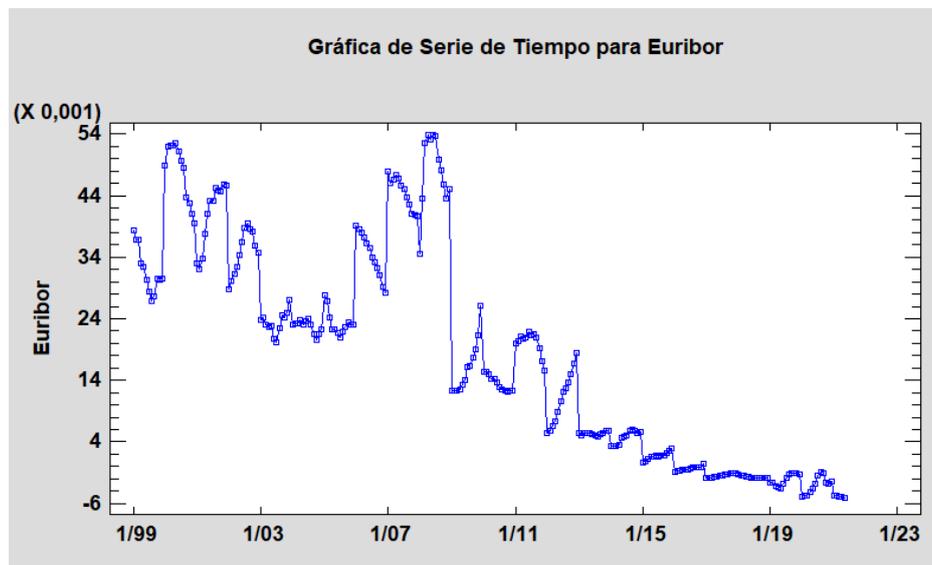
Q2 Predicción puntual: 8.587,61

Intervalo confianza (7.828,71<IPV< 9.346,5) 95% Confianza.

Q3 Predicción puntual: 9.566,15

Intervalo confianza (8.747,92 <IPV< 10.384,4) 95% Confianza

9.4.-Modelo serie temporal Euribor



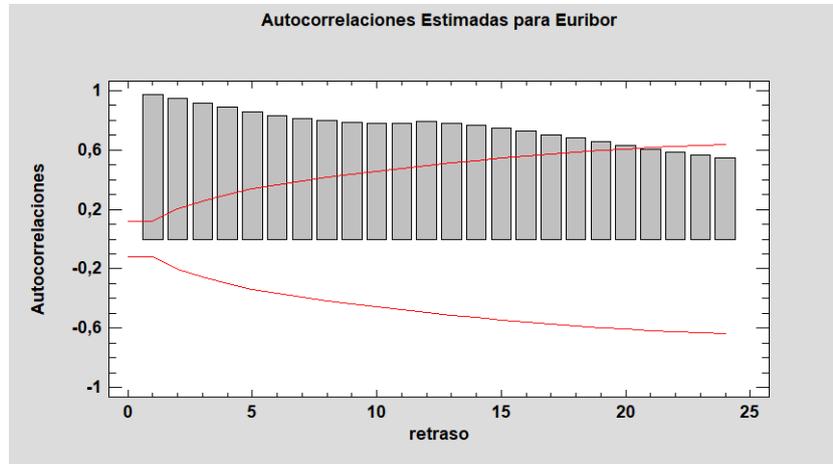
Gráfica 66 Series Tiempo Euribor

En la gráfica 66, se observa que existen muchas variaciones de tendencia durante todos los periodos, pero actualmente se encuentra en una tendencia negativa, por lo que se tendrán que aplicar diferenciaciones de orden no estacional.

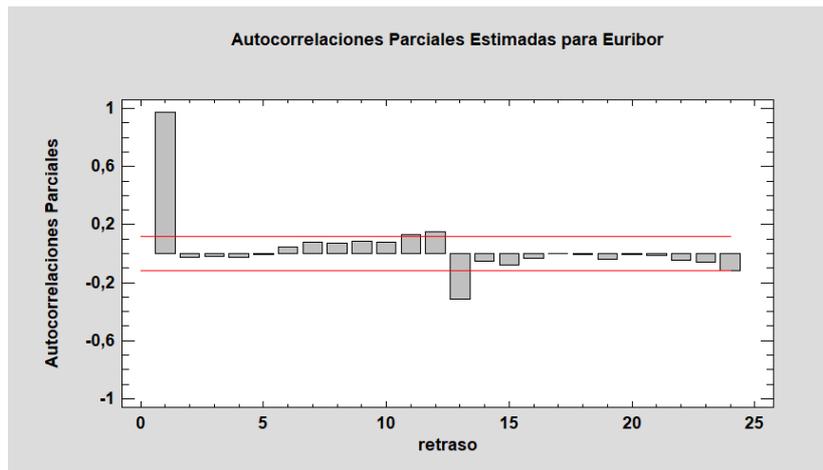
En segundo lugar, se puede observar que no existe problema de heterocedasticidad debido a que no se observa una forma cónica, que pueda apreciarse a simple vista.

En tercer lugar, no se aprecia problemas de estacionalidad, no obstante, cuando se observa el análisis FAS, se puede observar una forma sinusoidal que indica que si existes problemas de estacionalidad.

En cuarto lugar, las autocorrelaciones del FAS Y FAP.



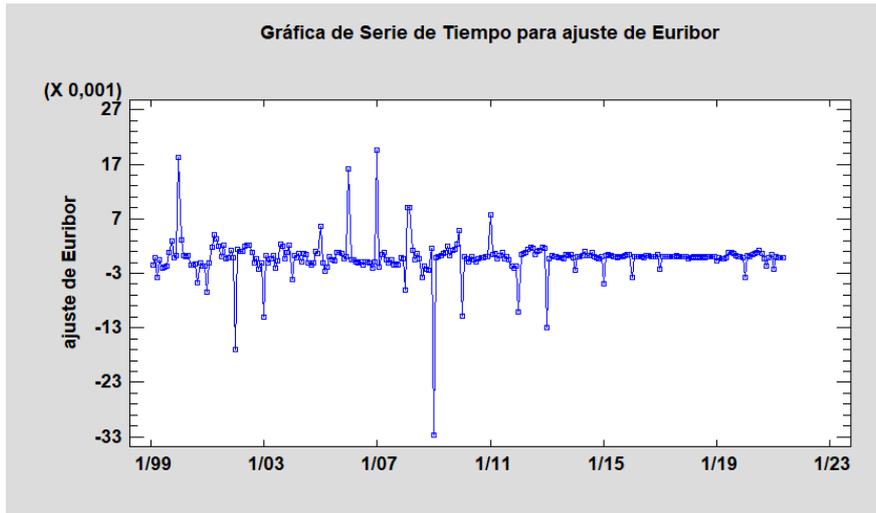
Gráfica 67 FAS Euribor



Gráfica 68 FAP Euribor

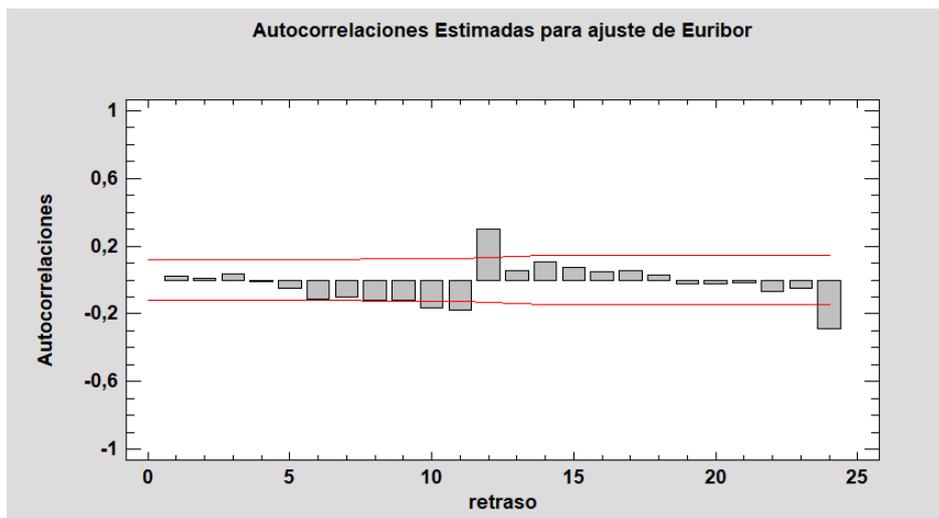
En las gráficas 67 y 68, se observa el FAS Y FAP, se observa que existe autocorrelación.

Para conseguir que la serie sea estacionaria se debe solucionar el problema de tendencia, para solucionar los problemas de tendencia se aplican dos diferenciaciones de orden no estacional, y una diferenciación de orden estacional.

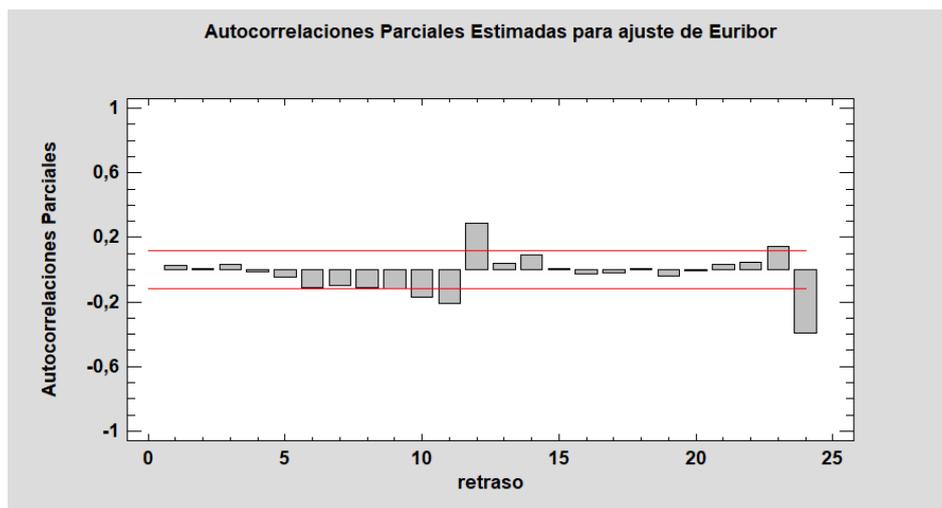


Gráfica 69 Serie Tiempo Ajustada Euribor.

En la gráfica 69, se observa cómo se corrige la tendencia quedando prácticamente nula llegando a ser cercana a 0.



Gráfica 70 FAS Ajustado Euribor



Gráfica 71 FAP Ajustado Euribor.

En las gráficas 70 y 71, se observa como el FAS tiene una caída rápida de los valores y en el FAP cuando la serie no suele ser estacionaria los primeros órdenes suelen tener una alta autocorrelación.

A continuación, se va a pasar a realizar un modelo ARIMA (0,2,1)(2,1,2)4 y su validación.

Modelo de pronóstico seleccionado: ARIMA(0,2,1)x(2,1,2)4

Parámetro	Estimado	Error Estd.	t	Valor-P
MA(1)	0,967462	0,0160917	60,122	0,000000
SAR(1)	-0,845704	0,062343	-13,5653	0,000000
SAR(2)	-0,260002	0,0464354	-5,59922	0,000000
SMA(1)	-0,0242973	0,0301686	-0,805385	0,421339
SMA(2)	0,908234	0,0243596	37,2844	0,000000

Tabla 43 Resumen ARIMA Euribor.

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor MA1 = 0.0000 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

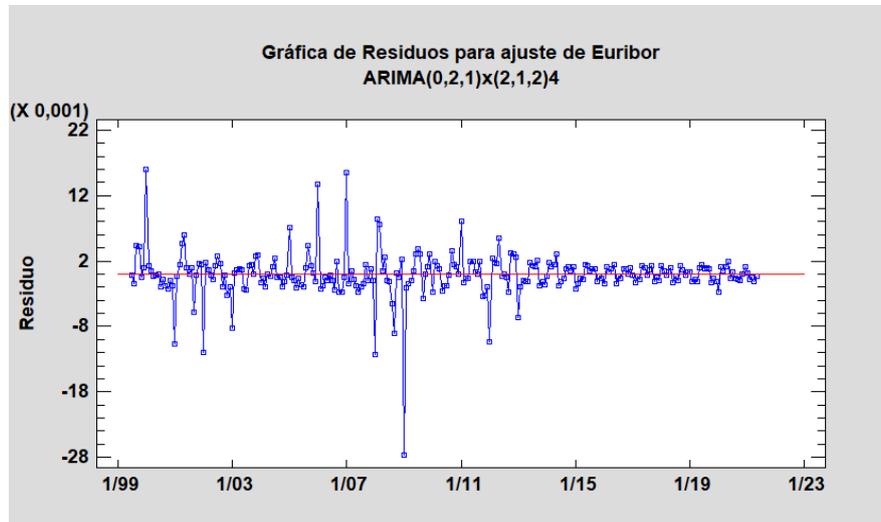
P- Valor SAR1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SAR2= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA1= 0,421339 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto, el parámetro no es significativo

P- Valor SMA2= 0,000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

De acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior los parámetros SMA1 no se podrían eliminar del modelo.



Gráfica 72 Series Tiempo Parados

En la gráfica 72, se observa que los residuos tienen media nula, representada por la línea horizontal de color rojo, y se aprecia que la serie es homogénea, ya no tiene aspecto de forma cónica.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	166,183	0,0
Estadístico W de Shapiro-Wilk	0,735722	0,0
Valor-Z para asimetría	5,22993	1,69907E-7
Valor-Z para curtosis	8,92478	0,0

Tabla 44 Pruebas Normalidad Residuos Euribor

Test normalidad residuos

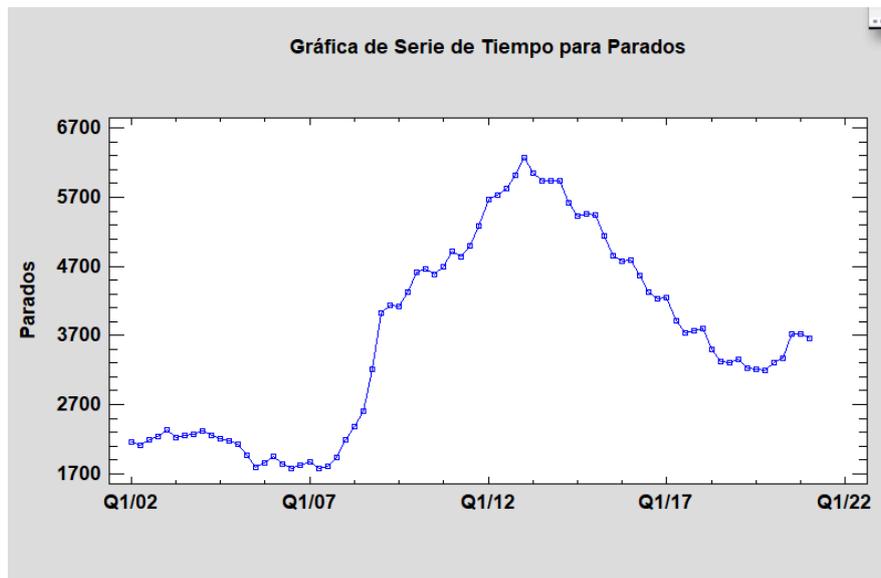
H_0 : Los residuos se distribuyen normalmente

H_1 : Los residuos no se distribuyen normalmente $\alpha > 0,05$

En la tabla 44, se escoge la prueba más restrictiva, que serían 3 de 4 estadísticos con P-valor 0,0, este dato es menor que Alfa, por tanto, si existe evidencia para rechazar la hipótesis alternativa y se concluye que los residuos no se distribuyen de una manera normal.

Por lo tanto, se concluye que no existe un modelo ARIMA correcto para la variable Euribor.

9.5.-Modelo serie temporal Parados



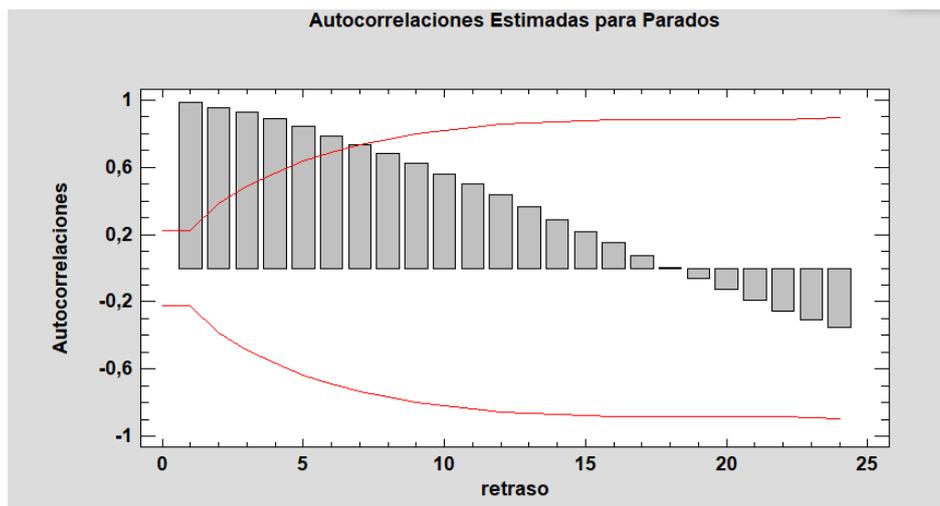
Gráfica 73 Series Tiempo Parados

En primer lugar, se puede observar en esta grafica que la tendencia está compuesta por 2 tramos, un primer tramo de tendencia ascendente (positiva) y el segundo tramo que se observa una tendencia decreciente (negativa).

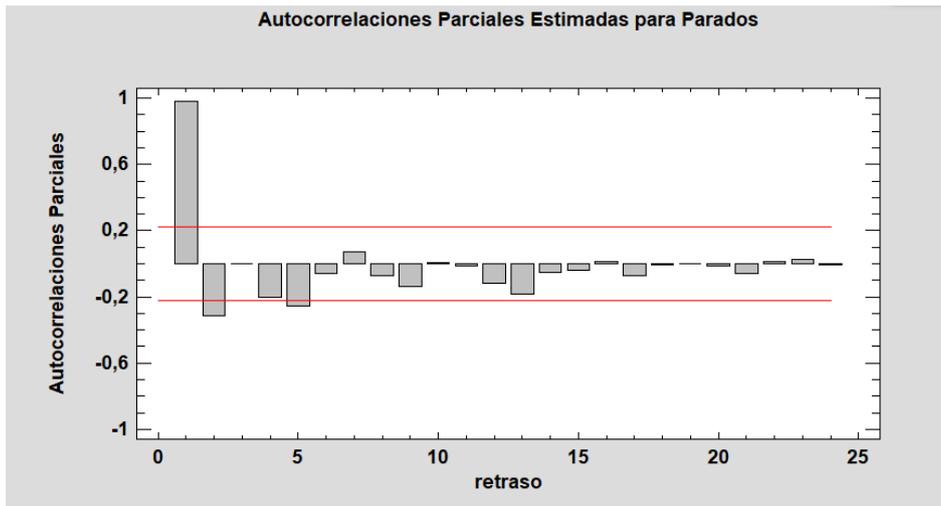
En segundo lugar, no se observan problemas de heteroscedasticidad, ya que no se ve una forma cónica.

En tercer lugar, no se aprecia variaciones estacionales en la serie de tiempo

En cuarto lugar, las autocorrelaciones del FAS Y FAP.



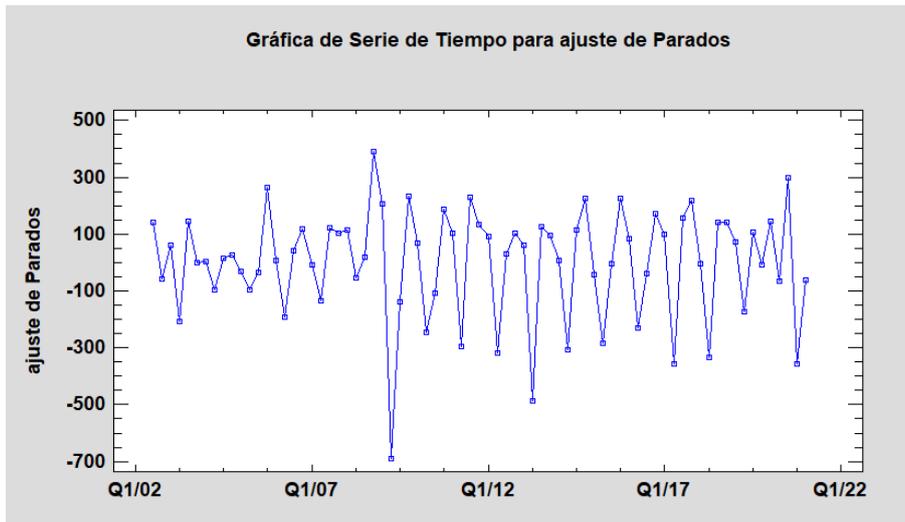
Gráfica 74 FAS Parados.



Gráfica 75 FAP Parados

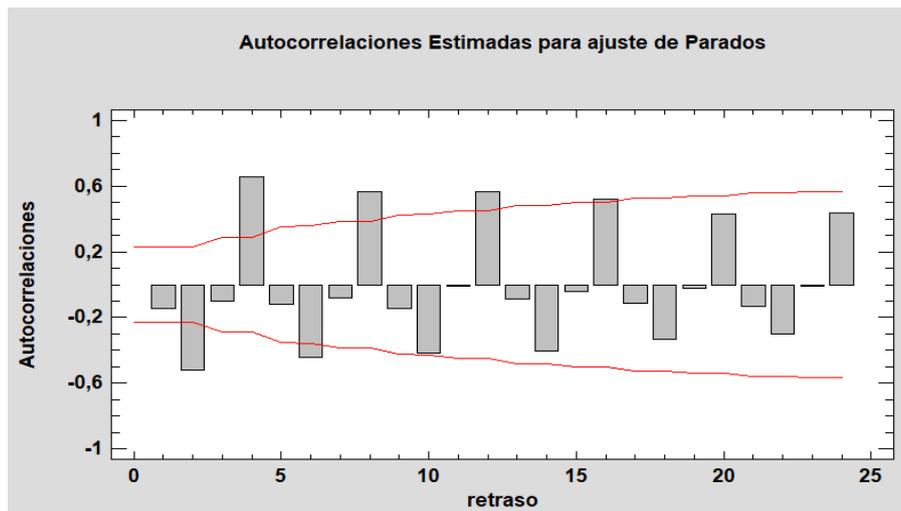
En las gráficas 74 y 75, se aprecia como en el FAS existe una caída rápida de los valores y en el FAP cuando la serie no suele ser estacionaria los primeros ordenes tienen mucha autocorrelación cosa que como se ha visto en la tendencia no existía a simple vista estacionalidad.

Se realiza dos diferenciaciones de orden no estacional y se observa en las siguientes gráficas.

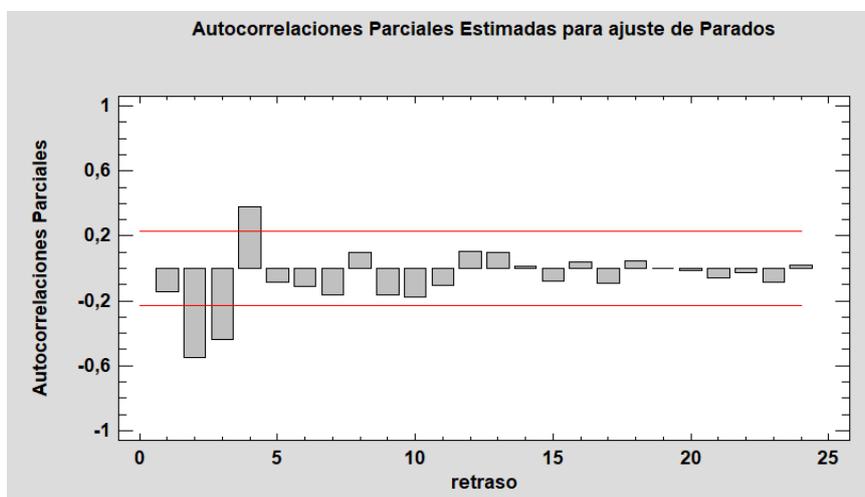


Gráfica 76 Serie Tiempo Ajustes Parados

En la gráfica 76, se observa cómo se ha suavizado la tendencia y se ha conseguido mantener nula.



Gráfica 77 FAS Ajuste Parados



Gráfica 78 FAP Ajuste Parados.

En las gráficas 77 y 78 se observa cómo tanto en el FAP como en el FAS se consiguen unos valores significativos en los dos casos, ya que las autocorrelaciones no son tan grandes y se ajustan más al modelo.

A continuación, se va a pasar la creación y la validación del modelo ARIMA.

Después de un largo análisis, en búsqueda del mejor modelo, se ha decidido utilizar la herramienta de pronósticos automáticos, y sale un ARIMA en el modelo de pronóstico $(2,0,0) \times (1,0,0)_4$.

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	1,6003	0,0930788	17,193	0,000000
AR(2)	-0,614001	0,0937953	-6,54619	0,000000
SAR(1)	0,754657	0,0911874	8,27589	0,000000

Tabla 45 Resumen Modelo ARIMA Parados.

Validación Parámetros.

•Cada parámetro

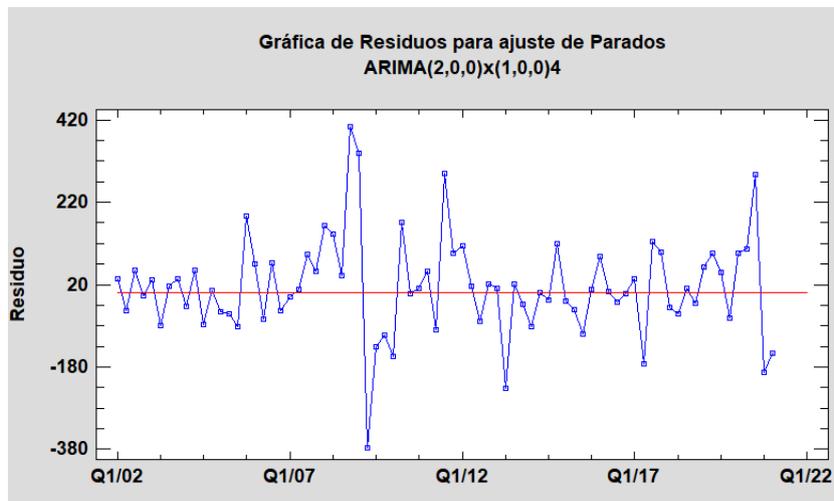
$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor AR1 = 0.0000 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro es significativo.

P- Valor AR2= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SAR1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.



Gráfica 79 Residuos Ajuste Parados.

En el gráfico 79 se muestra que los residuos tienen media nula, y no hay problemas de heterocedasticidad debido a que no presentan una forma cónica.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	19,2857	0,43865
Estadístico W de Shapiro-Wilk	0,95149	0,00521107
Valor-Z para asimetría	0,90012	0,368055
Valor-Z para curtosis	2,81081	0,00494177

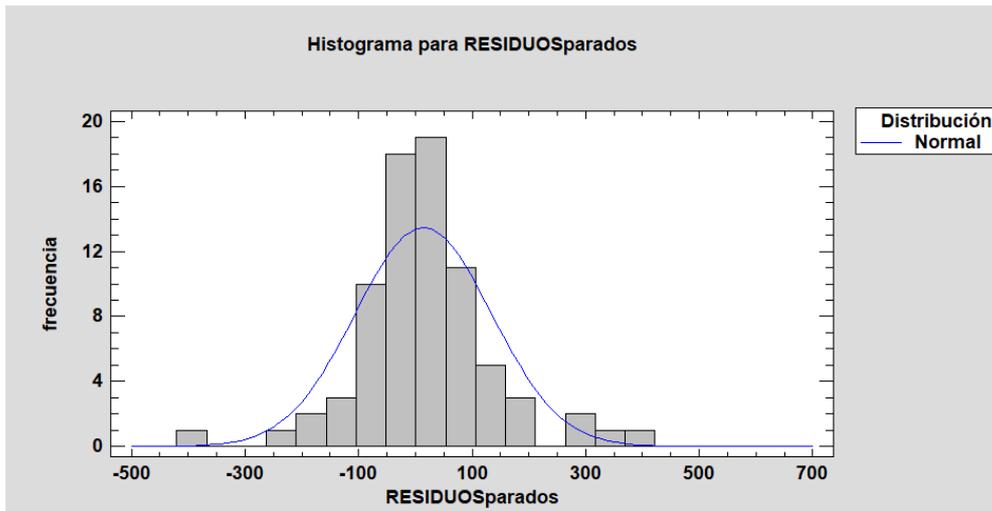
Tabla 46 Pruebas Normalidad Parados

Test normalidad residuos

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente α>0,05

El dato más restrictivo de la prueba de normalidad es el Valor-Z para curtosis cuyo P-Valor es 0.00494177 este valor es inferior a Alfa, por tanto, se rechaza la hipótesis nula, y se concluye que los residuos no se distribuyen de forma normal.



Gráfica 80 Histograma Residuos Parados

En la gráfica 80, en el histograma se puede ver que la distribución parece una distribución normal, debido a que los residuos forman una campana de Gauss no obstante de acuerdo con la teoría aprendida para saber si los residuos se comportan de forma normal se debe tomar el valor del Estadístico W de Shapiro-Wilk y comprobar su P-valor 0.005211 frente al valor de Alfa 0,05 de lo que se concluye que rechaza la hipótesis nula y por tanto los residuos no son normales, por consiguiente el modelo ARIMA no es válido.

9.6.-Modelo serie temporal Viviendas Iniciadas



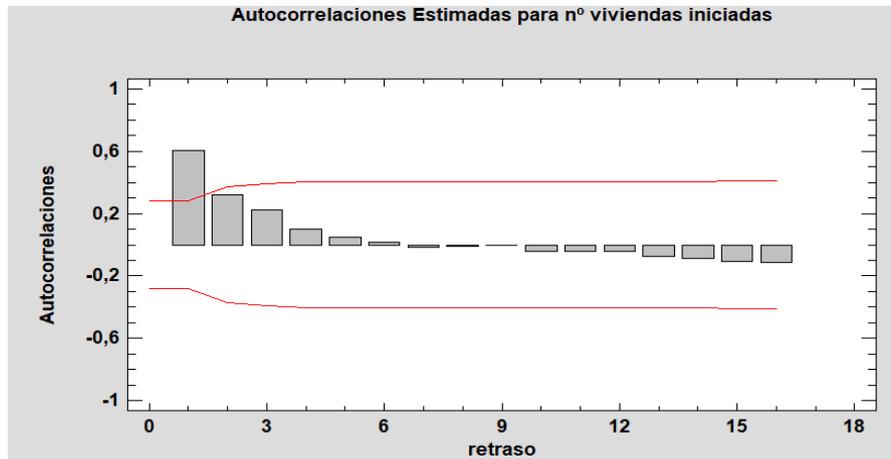
Gráfica 81 Serie Tiempo VivIni

En la gráfica 81, se puede observar en esta grafica que la tendencia está compuesta por 2 tramos, un primer tramo de tendencia descendente con una caída grave de los valores y luego un segundo tramo de tendencia positiva y constante.

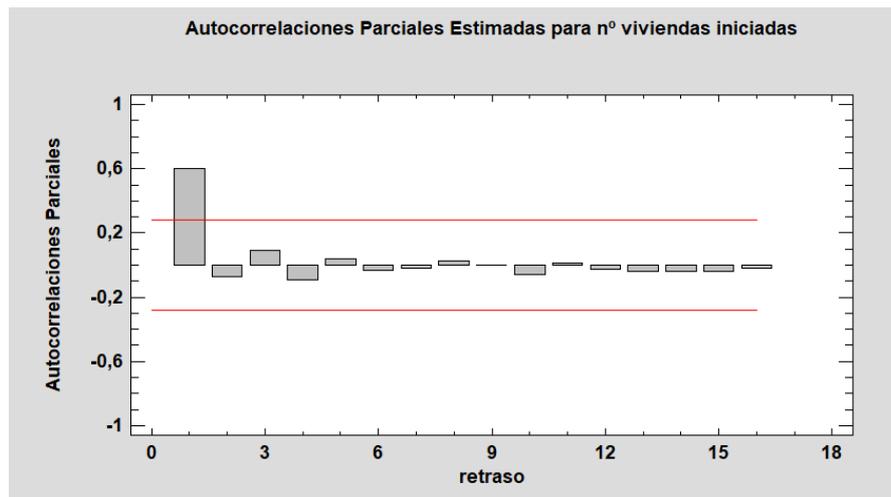
En segundo lugar, se puede observar que no existen problemas de heteroscedasticidad, ya que no se ve una forma cónica, que pueda apreciarse a simple vista.

En tercer lugar, no se aprecia problemas de estacionalidad.

En cuarto lugar, las autocorrelaciones del FAS Y FAP.

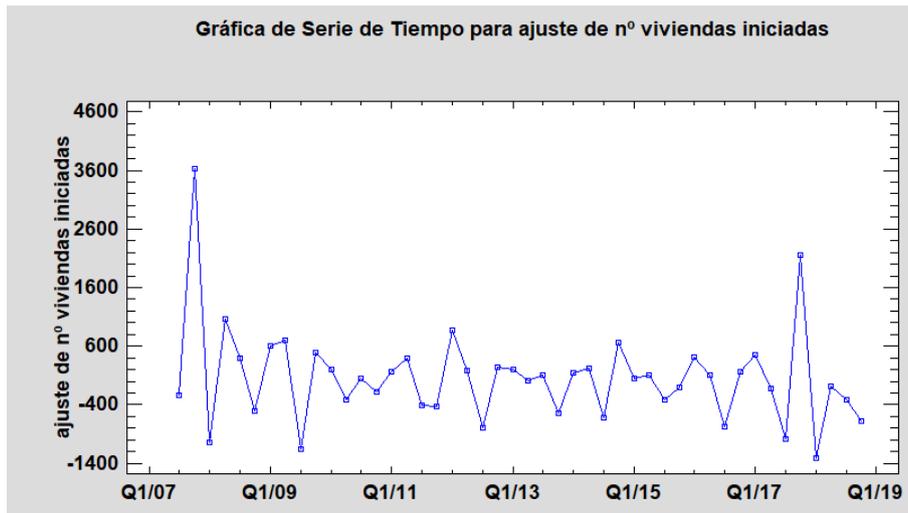


Gráfica 82 FAS Vivini



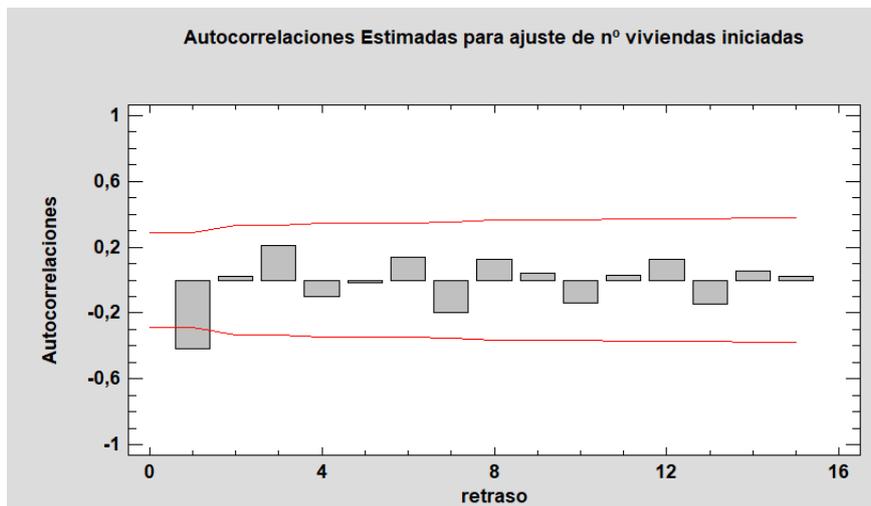
Gráfica 83 FAP Vivini.

En las gráficas 82 y 83, se observa que tanto en el FAS como en el FAP existe autocorrelación. Para transformar la serie de tiempo en estacionaria se realizan 2 diferenciaciones de orden no estacional.

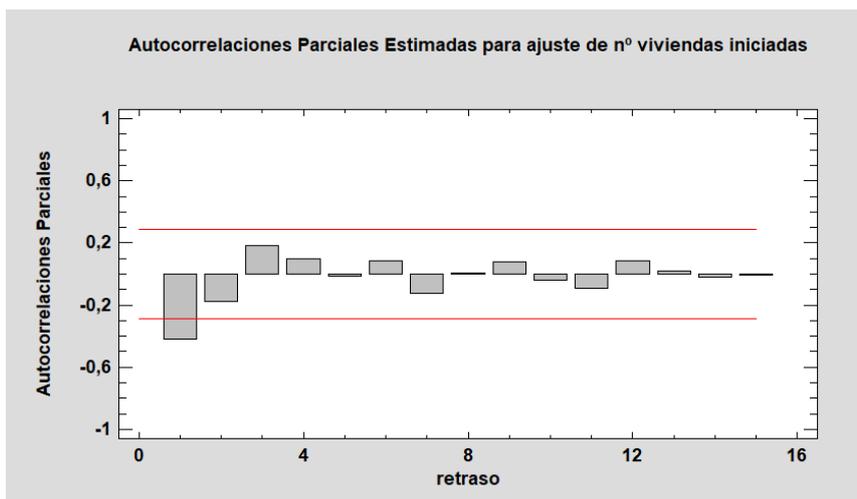


Gráfica 84 Serie Tiempo Ajustado Vivini

En la gráfica 84, se observa como la tendencia de la serie pasa a ser nula o constante.



Gráfica 85 FAS Ajustado Vivini.



Gráfica 86 FAP Ajustado Vivlni.

En las gráficas 85 y 86, se encuentra tanto el FAS como el FAP, y muestran que la serie tiene autocorrelación.

Después de un largo análisis, se ha decidió usar la función de pronósticos automáticos para que el programa muestre el modelo automáticamente; este modelo es un ARIMA (2,2,0) x (1,0,2)4.

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	-0,590238	0,135837	-4,3452	0,000090
AR(2)	-0,43322	0,120801	-3,58623	0,000885
SAR(1)	0,7616	0,05025	15,1562	0,000000
SMA(1)	1,0535	0,0519776	20,2683	0,000000
SMA(2)	-0,85593	0,0531741	-16,0967	0,000000

Tabla 47 Resumen Modelo ARIMA Vivlni

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

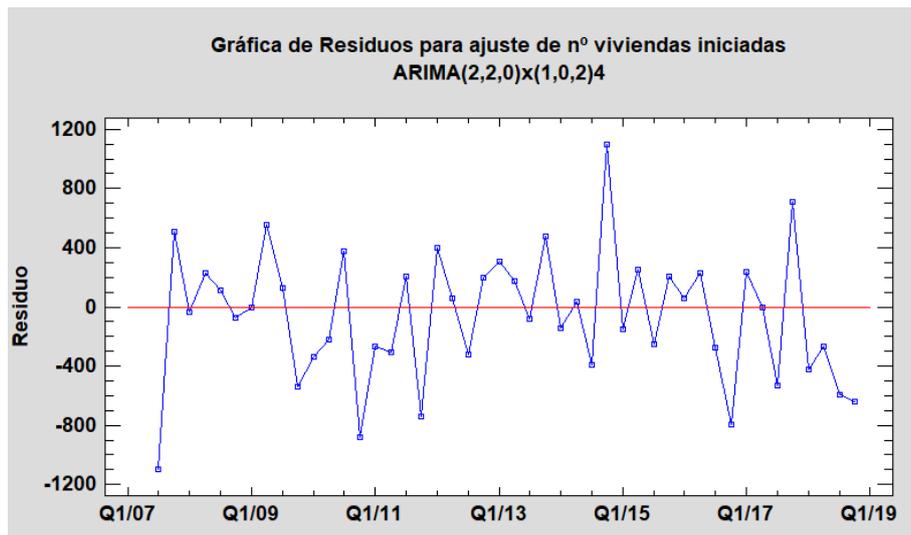
P-Valor AR1 = 0.0000090 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro es significativo.

P- Valor AR2= 0,0000885 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SAR1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA2= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.



Gráfica 87 Residuos Ajuste Vivni

En la gráfica 87, se observa como los residuos tienen media nula y existe homocedasticidad ya que no tiene una forma cónica.

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Chi-Cuadrado	9,56522	0,846162
Estadístico W de Shapiro-Wilk	0,99184	0,985043
Valor-Z para asimetría	0,0557987	0,955497
Valor-Z para curtosis	0,624281	0,532441

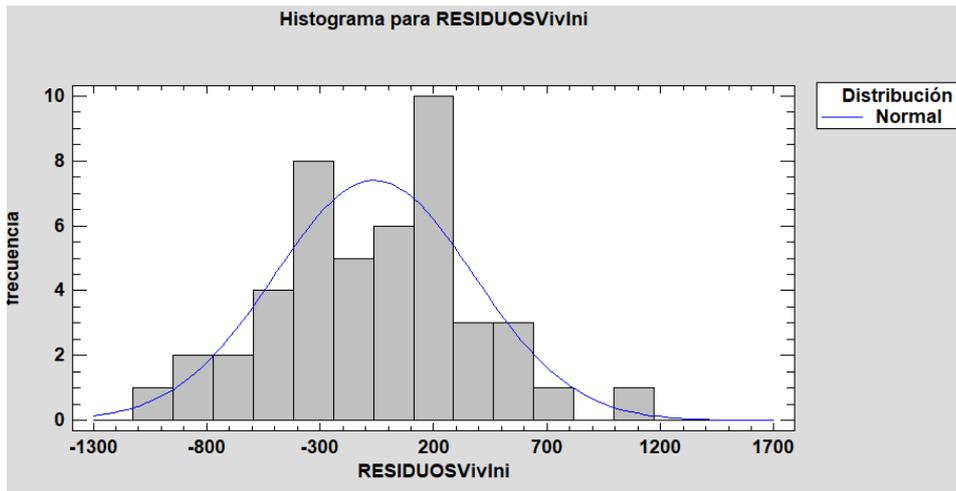
Tabla 48 Normalidad Residuos Vivni.

Test normalidad residuos

H_0 : Los residuos se distribuyen normalmente

H_1 : Los residuos no se distribuyen normalmente $\alpha > 0,05$

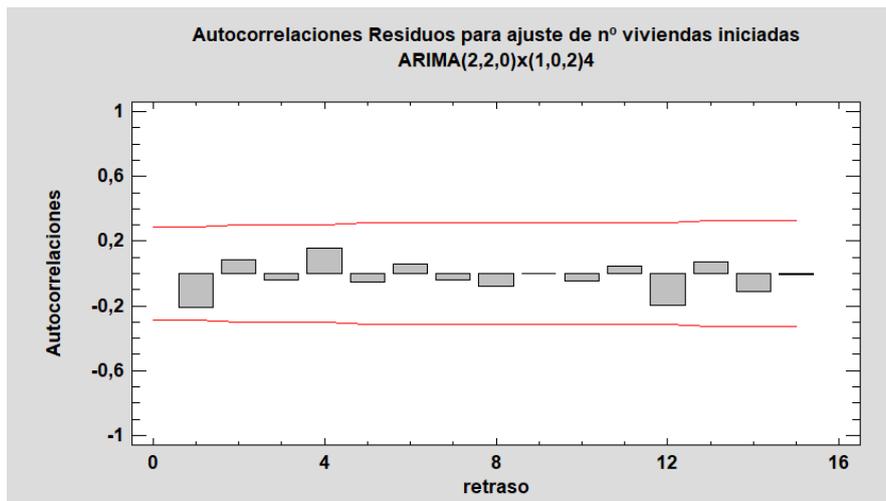
En la tabla 48, se toma el Valor-z para curtosis con P-Valor 0.532441, este valor es superior a Alfa por tanto no existe evidencia para rechazar la hipótesis alternativa y se afirma que los residuos se distribuyen de forma normal.



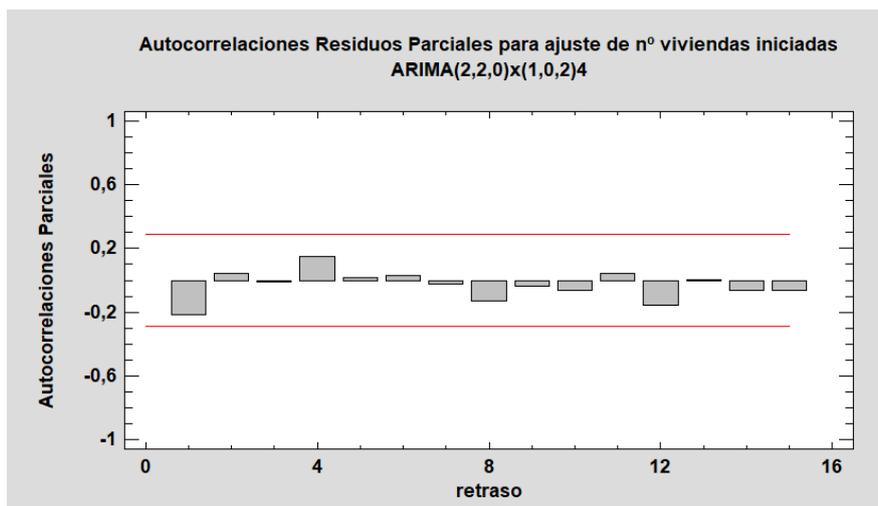
Gráfica 88 Histograma Residuos VivIni

En la gráfica 88. Se analiza el histograma de residuos y se observa la forma de campana de Gauss por tanto se podría afirmar que los residuos son normales.

Una vez se ha cumplido todos los pasos anteriores lo último que faltaría para validar el modelo es observar el FAS Y FAP del modelo ARIMA.



Gráfica 89 FAS Ajustado VivIni



Gráfica 90 FAP Ajustado Vivlni

En las gráficas 89 y 90, muestra que no existen problemas de autocorrelación.

Por tanto el modelo ARIMA (2,2,0)x(1,0,2)4 es un modelo válido para poder hacer una predicción de las Viviendas Iniciadas.

		Límite en 95%	Límite en 95%
Periodo	Pronóstico	Inferior	Superior
Q1/19	1573,93	320,722	2827,14
Q2/19	1542,9	-623,169	3708,96
Q3/19	525,833	-2543,0	3594,67
Q4/19	618,338	-3648,39	4885,06
Q1/20	349,853	-4988,8	5688,51

Tabla 49 Pronósticos Vivlni

Como se observa en la tabla 49, y respecto a lo que se ha comentado anteriormente solo se escogerán 2 o 3 valores debido a que el programa que se está usando arrastra un error y cada vez se hace más grande y deja de ser fiable.

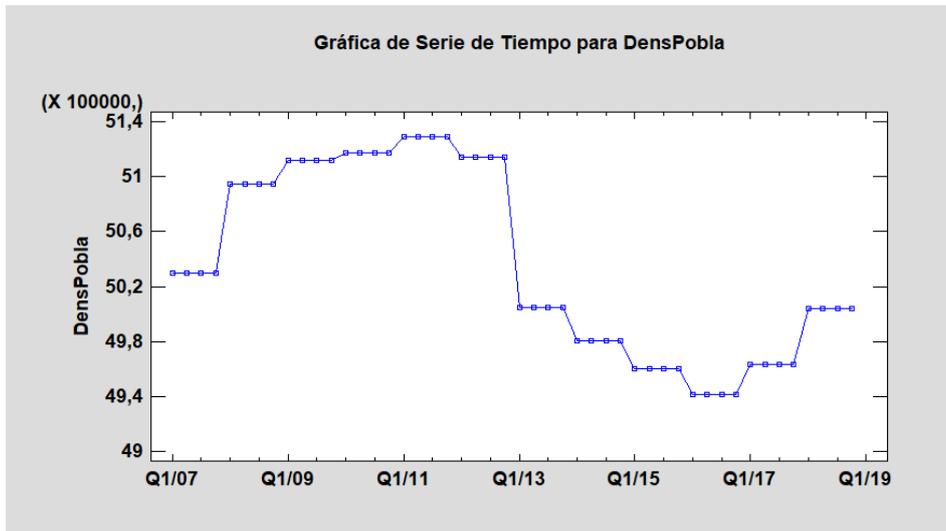
Q1 Predicción puntual: 1573,93

Intervalo confianza (320.722 <IPV< 2827.14) 95% Confianza.

Q2 Predicción puntual: 1542,9

Intervalo confianza (-623.169 <IPV< 3708.96) 95% Confianza.

9.7.-Modelo serie temporal Densidad de Población.



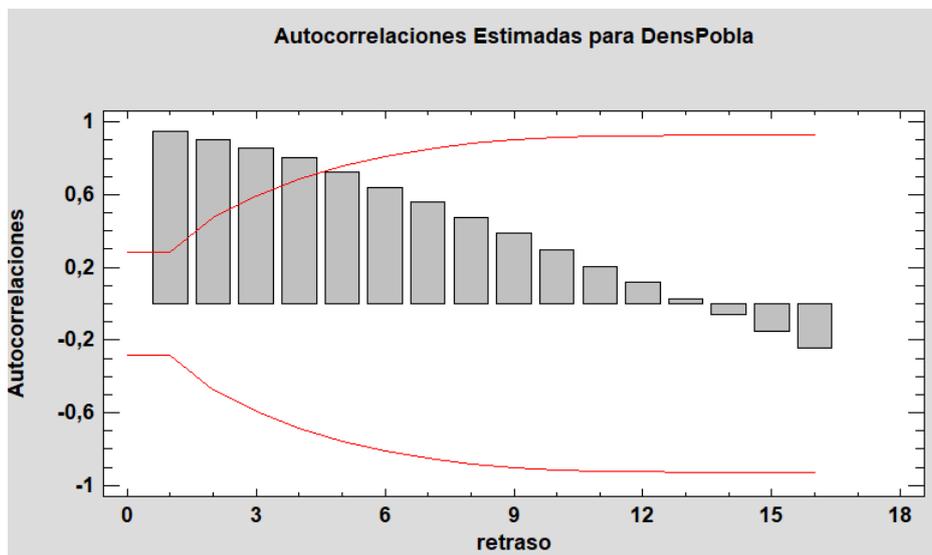
Gráfica 91 Serie Tiempo DensPobla

En la gráfica 91, se puede observar en esta grafica que la tendencia está compuesta por 2 tramos, un primer tramo de tendencia ascendente y un último tramo de tendencia descendente.

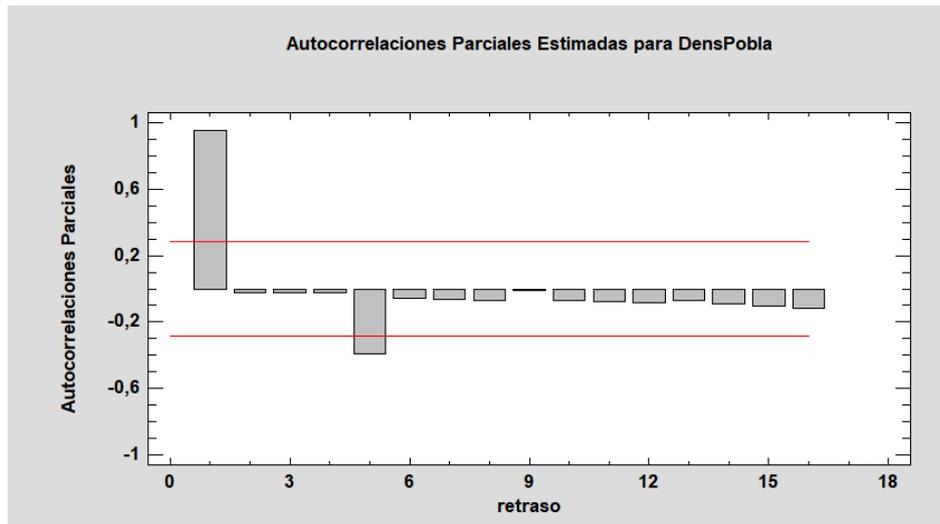
En segundo lugar, se puede observar que no existen problemas de heterocedasticidad ya que no se ve una forma cónica, que pueda apreciarse a simple vista.

En tercer lugar, no se aprecian variaciones estacionales.

En cuarto lugar, las autocorrelaciones del FAS Y FAP.



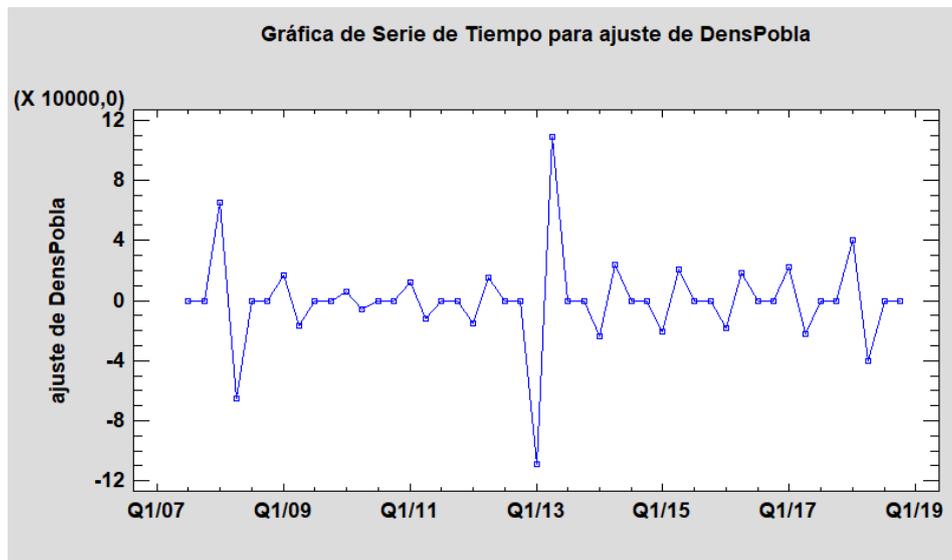
Gráfica 92 FAS DensPobla



Gráfica 93 FAP DensPobla

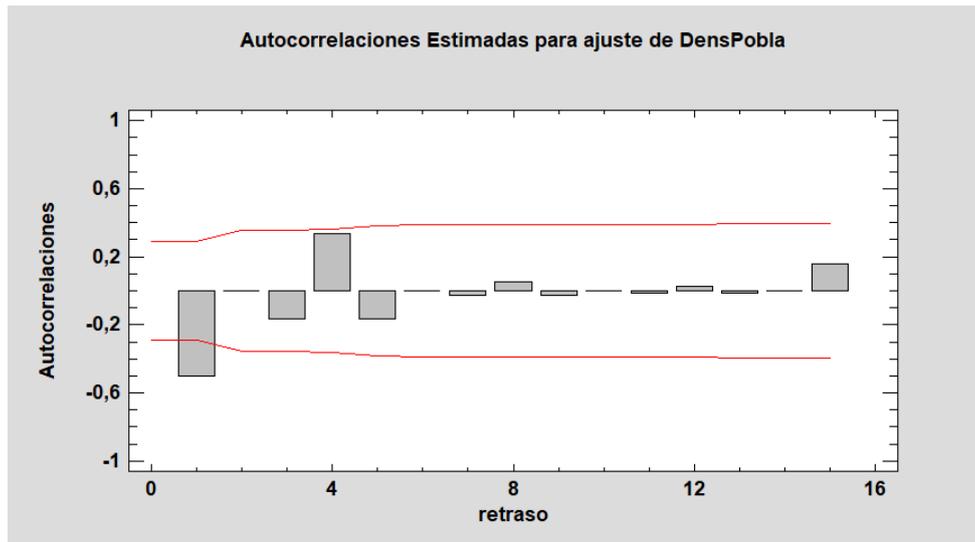
En las gráficas 92 y 93, aparece el FAS y el FAP, en ambos gráficos se observa que existe autocorrelación.

Para conseguir que la serie sea estacionaria se realizan 2 diferenciaciones de orden no estacional.

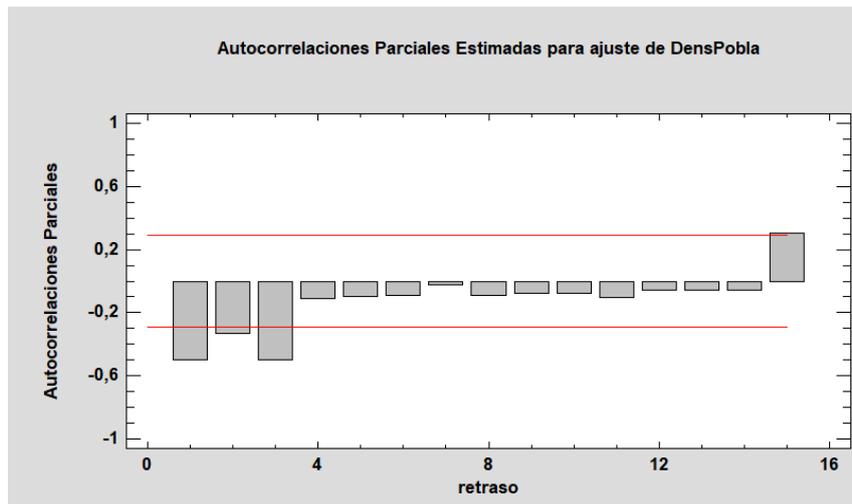


Gráfica 94 Serie Tiempo Ajustada DensPobla.

En la gráfica 94 se observa como la serie tiene una tendencia nula o constante.



Gráfica 95 FAS Ajustado DensPobla.



Gráfica 96 FAP Ajustado DensPobla.

Como se puede apreciar en las gráficas 95 y 96 de FAS y FAP, se observa como siguen existiendo autocorrelación en ambas.

A continuación, se va a proponer a la creación y la validación de un modelo ARIMA.

El modelo que saca el programa con su apartado de pronósticos automáticos, es el pronóstico seleccionado: ARIMA(2,2,2)x(2,0,2)₄

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	-0,505732	2,28357	-0,221465	0,825916
AR(2)	-0,022272	0,16523	-0,134794	0,893486
MA(1)	0,49456	2,28706	0,216243	0,829955
MA(2)	0,505131	2,29835	0,21978	0,827219
SAR(1)	1,66045	0,171355	9,69009	0,000000
SAR(2)	-0,847445	0,131977	-6,42114	0,000000

SMA(1)	1,85023	0,257724	7,17912	0,000000
SMA(2)	-0,788301	0,32733	-2,40827	0,020987

Tabla 50 Modelo ARIMA DensPobla

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor AR1 = 0.825916 , es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto el parámetro no es significativo.

P- Valor AR2= 0,893486 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto el parámetro no es significativo.

P- Valor MA1= 0,829955 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto el parámetro no es significativo.

P- Valor MA2= 0,827219 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto el parámetro no es significativo.

P- Valor SAR1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

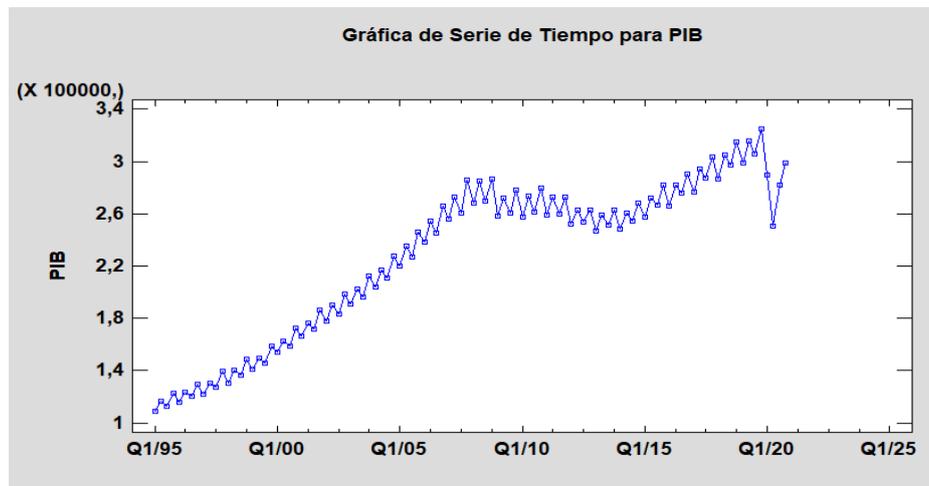
P- Valor SAR2= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor SMA2= 0,020987 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

Por lo que claramente se puede afirmar que no existe modelo ARIMA válido, para la variable Densidad Población debido a que no sale significativo y por lo tanto no se puede continuar con la validación del modelo.

9.8.-Modelo serie temporal PIB

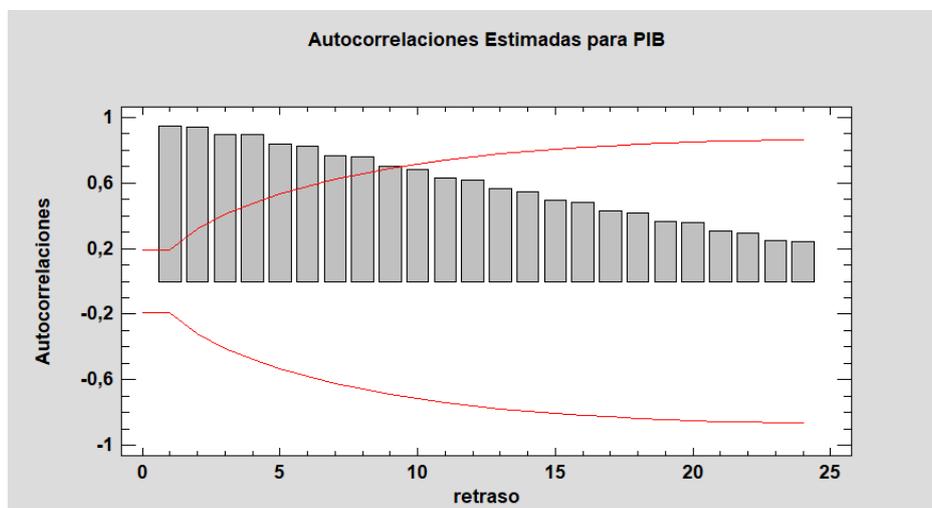


Gráfica 97 Serie Tiempo PIB.

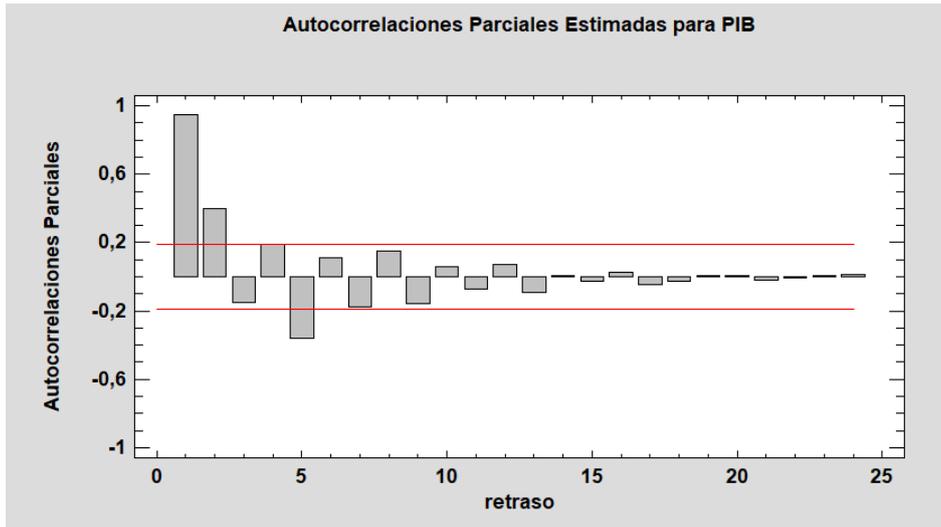
En la gráfica 97, se observa como la serie de tiempo tiene una tendencia positiva.

La varianza no es constante ya que se puede apreciar una forma cónica, por tanto, existe un problema de heterocedasticidad.

Se observan variaciones estacionales, en el segundo y cuarto trimestre.



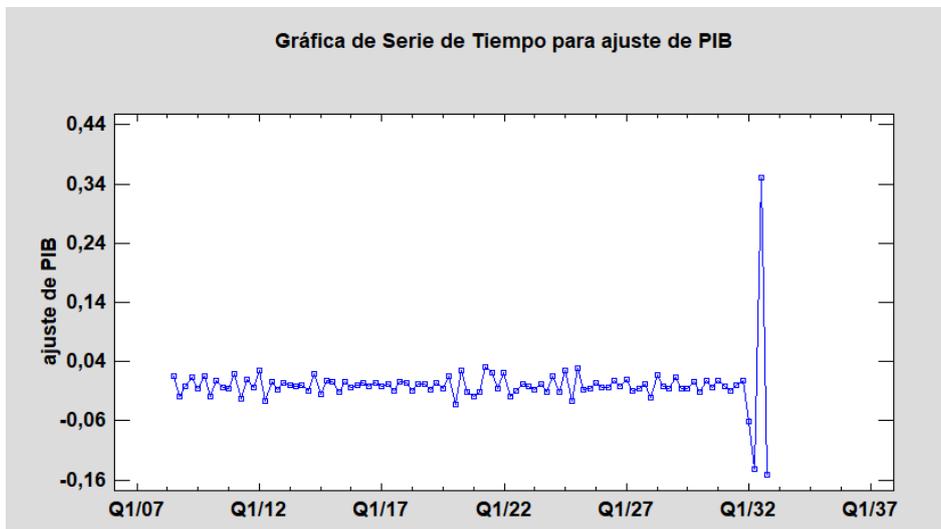
Gráfica 98 FAS PIB.



Gráfica 99 FAP PIB

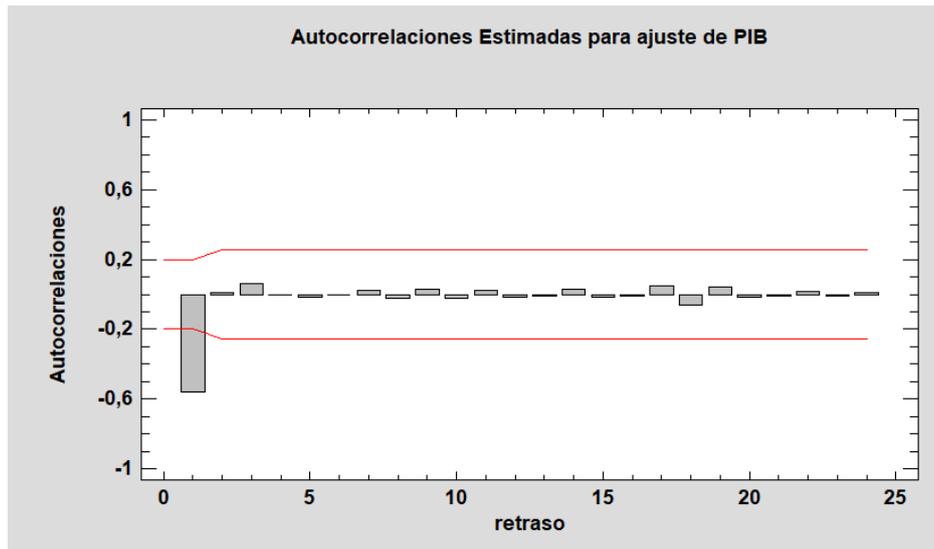
En las gráficas 98 y 99, se puede observar tanto en la FAS como en el FAP que indican que existe autocorrelación.

Para conseguir que la serie sea estacionaria, se debe transformar la serie en logarítmica para solucionar el problema de heteroscedasticidad, posteriormente se realizan 2 diferenciaciones de orden no estacional para mejorar la tendencia, y por último se ha realizado 1 diferenciación de orden estacional para solucionar las variaciones estacionales.

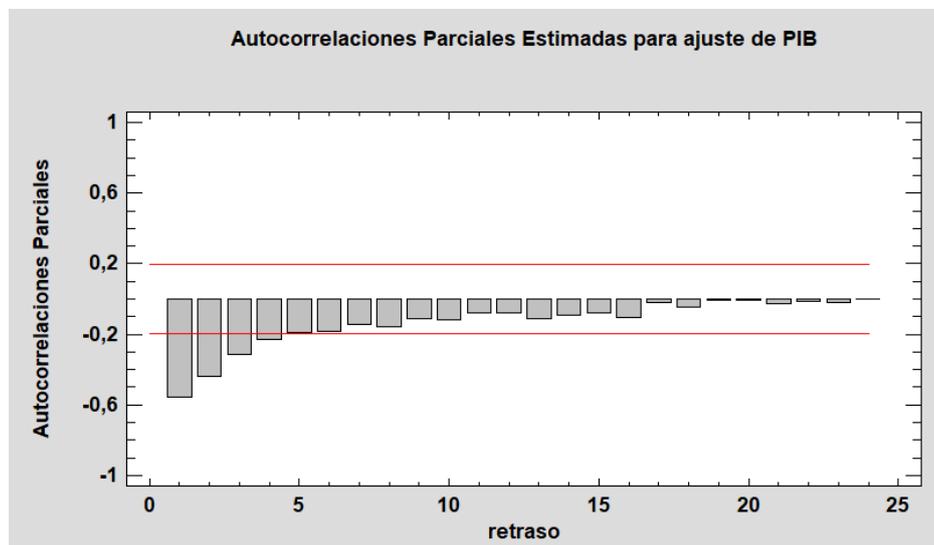


Gráfica 100 Serie Tiempo Ajustada PIB.

En la gráfica 100, se observa que la serie se ha suavizado completamente consiguiendo así una tendencia nula o muy próxima al 0 y no se aprecian problemas de heteroscedasticidad. Se observa que puede existir un punto anómalo.



Gráfica 101 FAS Ajustado PIB



Gráfica 102 FAP Ajustado PIB.

En la gráfica 101 y 102, se puede observar en la FAP Y la FAS se observa que existe autocorrelación.

A continuación, se va a pasar a realizar el modelo ARIMA y su validación.

El Modelo ARIMA es el propuesto por el propio programa, ya que se ha usado la función de pronósticos automáticos para que muestre el mejor modelo.

Este modelo es de pronóstico seleccionado: ARIMA(1,0,1)x(1,0,0)4.

Parámetro	Estimado	Error Estd.	t	Valor-P
AR(1)	0,933753	0,0596591	15,6515	0,000000
MA(1)	0,233603	0,119495	1,95492	0,053357
SAR(1)	0,987989	0,0425375	23,2263	0,000000

Tabla 51 Modelo ARIMA PIB

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

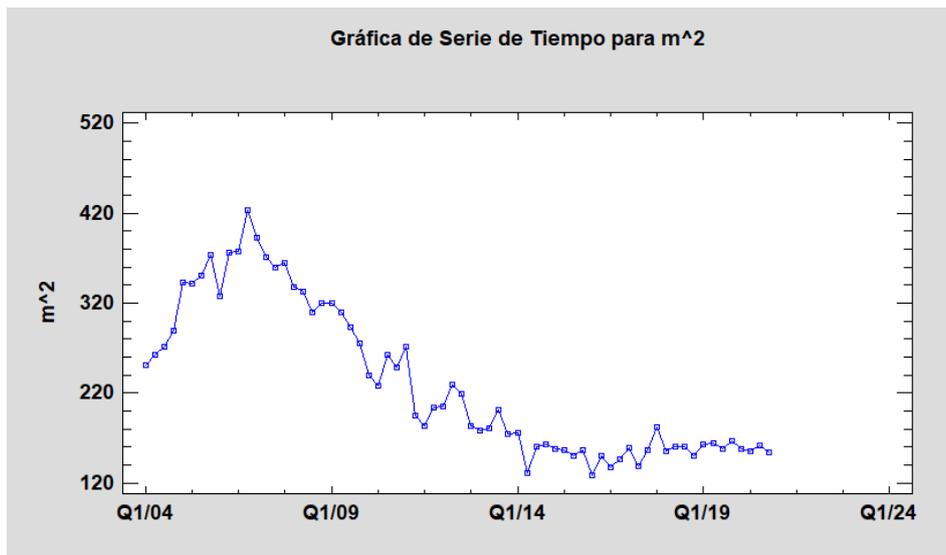
P-Valor AR1 = 0.00000, es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P- Valor MA1= 0,053357 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto, el parámetro no es significativo.

P- Valor SAR1= 0,0000 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

Analizando los parámetros se concluye que el modelo ARIMA (1,0,1)(1,0,0)4 no es un modelo válido.

9.9.-Modelo serie temporal M²



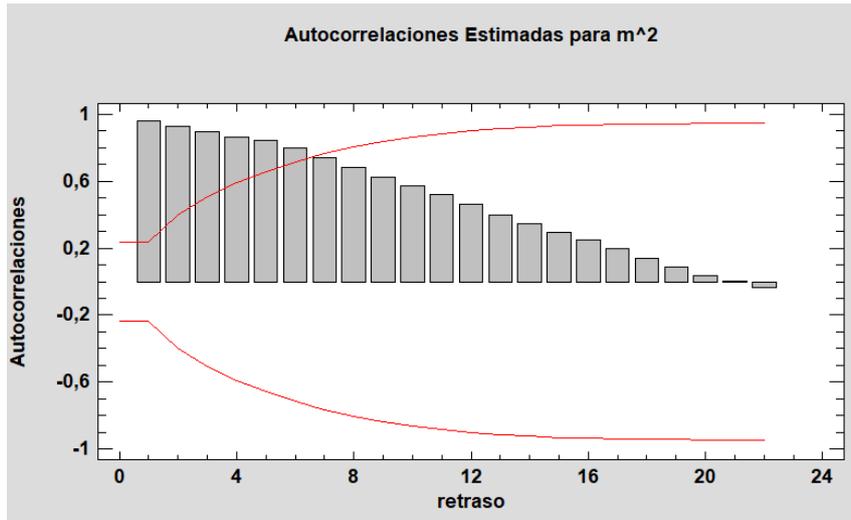
Gráfica 103 Serie Tiempo M²

En la gráfica 103, se observa una tendencia negativa por lo que ya no cumple uno de los criterios para que la serie sea estacionaria, se realizarán diferenciaciones de orden no estacional para convertir la serie en estacionaria.

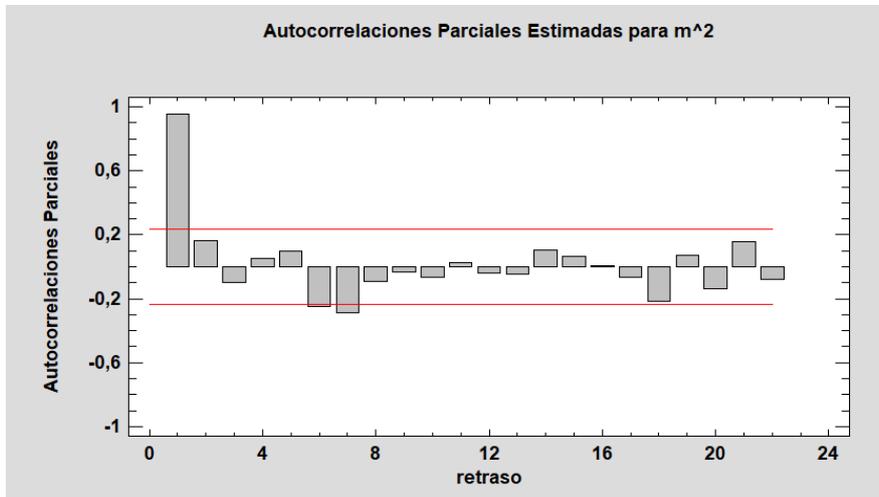
En segundo lugar, no se observan problemas de heteroscedasticidad, ya que no se ve una forma cónica.

En tercer lugar, no se observa variaciones estacionales.

En cuarto lugar, las autocorrelaciones del FAS Y FAP.



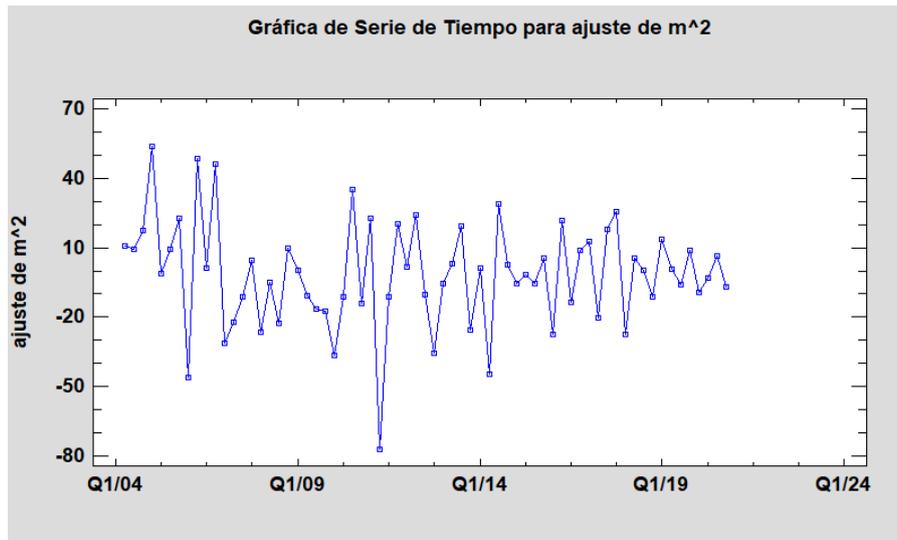
Gráfica 104 FAS M^2 .



Gráfica 105 FAP M^2

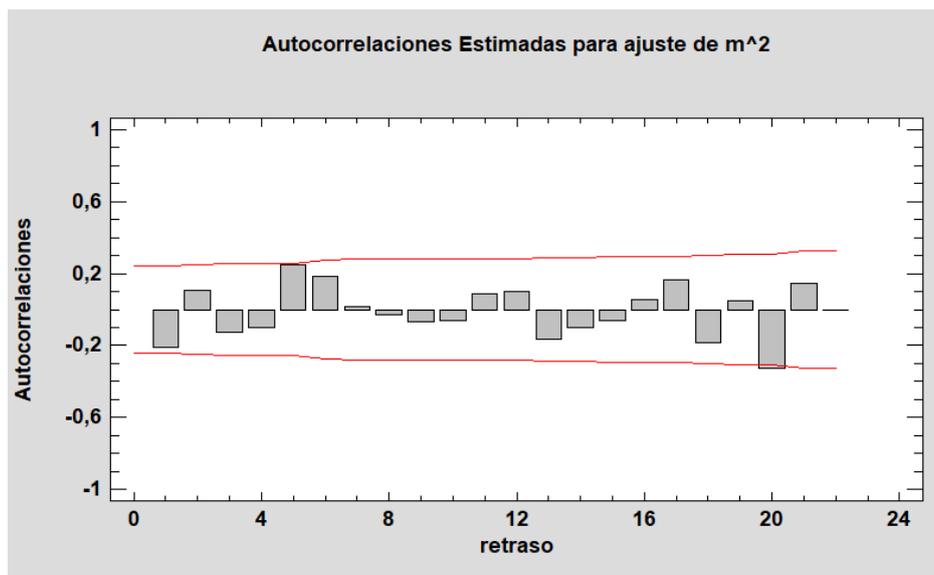
En las gráficas 104 y 105, se observa que tanto en la FAS como en el FAP que indican que existe autocorrelación.

Para conseguir que la serie sea estacionaria, se soluciona el problema de la tendencia aplicando 1 diferenciación de orden no estacional.

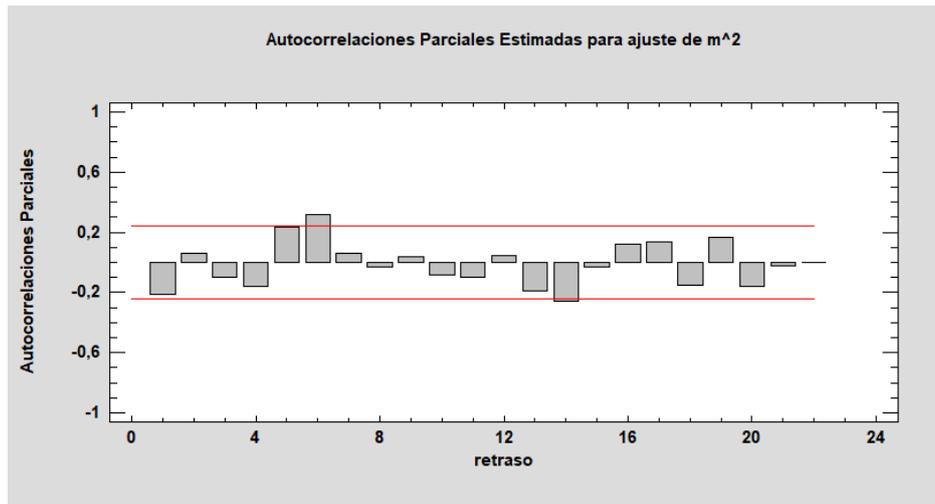


Gráfica 106 Serie Tiempo Ajustada M^2

En la gráfica 106, se puede observar que la tendencia negativa desaparece y se transforma en una tendencia nula o constante debido a que se ha suavizado el modelo.



Gráfica 107 FAS Ajustado M^2



Gráfica 108 FAP Ajustado M²

En la gráfica 107 y 108, se observa los gráficos FAS y FAP que presentan autocorrelación.

A continuación, se va a pasar realizar el modelo ARIMA y su validación.

El modelo que el programa estima que es el óptimo es un ARIMA (0,2,2,) x (2,0,2)4.

Parámetro	Estimado	Error Estd.	t	Valor-P
MA(1)	1,50664	0,129788	11,6085	0,000000
MA(2)	-0,589959	0,129492	-4,55595	0,000048
SAR(1)	-0,32831	0,11093	-2,95962	0,005156
SAR(2)	-0,873787	0,108958	-8,01949	0,000000
SMA(1)	0,0842119	0,129871	0,648427	0,520414
SMA(2)	-0,806054	0,139095	-5,795	0,000001

Tabla 52 Resumen Modelo ARIMA M²

Validación Parámetros.

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

P-Valor MA1 = 0.00000 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro es significativo.

P-Valor MA2 = 0.000048 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro es significativo.

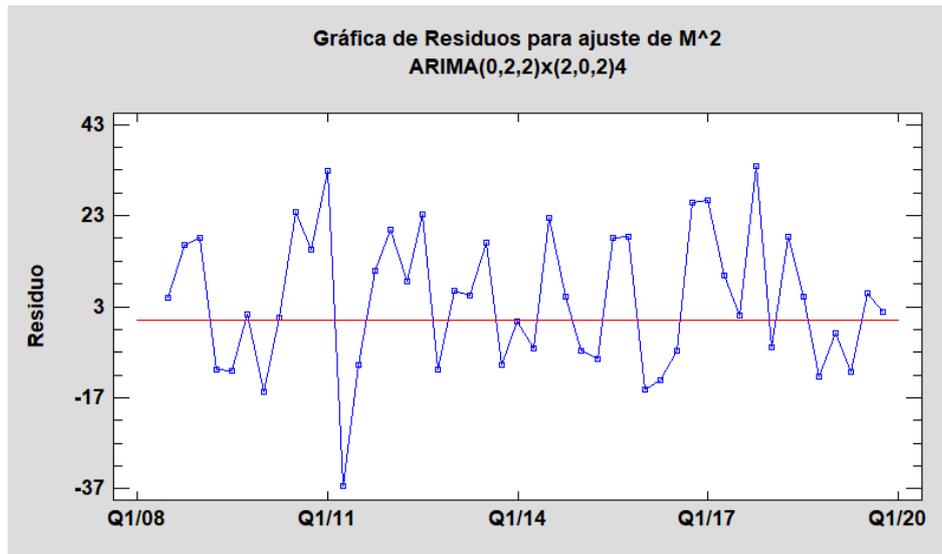
P-Valor SAR1 = 0.005156 es inferior a Alfa, se rechaza la hipótesis nula, por tanto, el parámetro es significativo.

P-Valor SAR2 = 0.00000 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro es significativo.

P- Valor SMA1= 0,520414 es superior a Alfa, no existe evidencia para rechazar la hipótesis alternativa, por tanto el parámetro no es significativo.

P-Valor SMA2 = 0.000001 , es inferior a Alfa, se rechaza la hipótesis nula, por tanto el parámetro es significativo.

De acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior los parámetros SMA1 no se podrían eliminar del modelo.



Gráfica 109 Observación Residuos Ajuste M².

En la gráfica 109, se observa como los residuos tienen media nula y existe homocedasticidad ya que no tiene una forma cónica.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	23,6522	0,0712405
Estadístico W de Shapiro-Wilk	0,969181	0,258203
Valor-Z para asimetría	0,193324	0,846701
Valor-Z para curtosis	-0,244471	0,806862

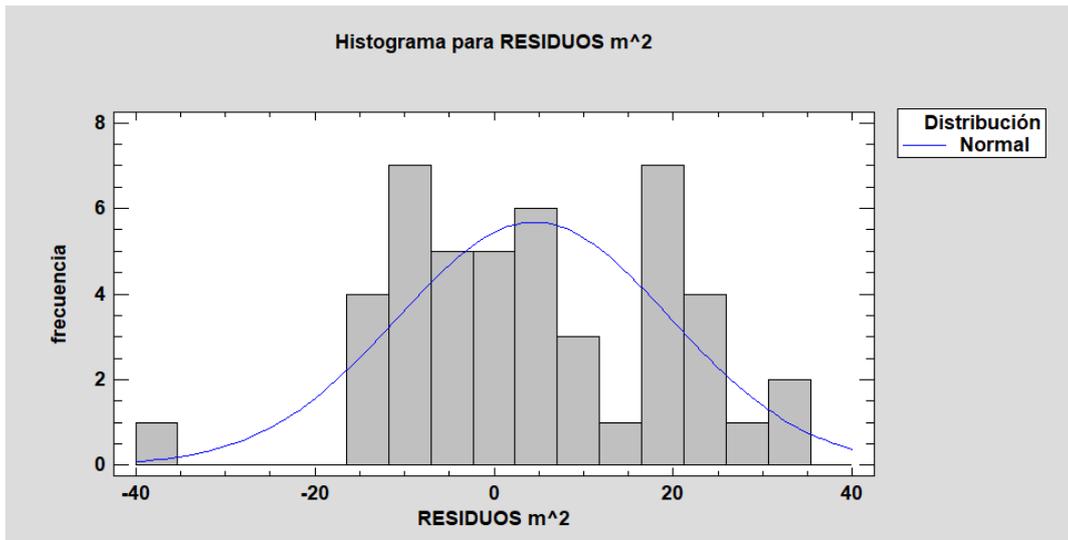
Tabla 53 Prueba Normalidad Residuos M²-

Test normalidad residuos

H0: Los residuos se distribuyen normalmente

H1: Los residuos no se distribuyen normalmente $\alpha > 0,05$

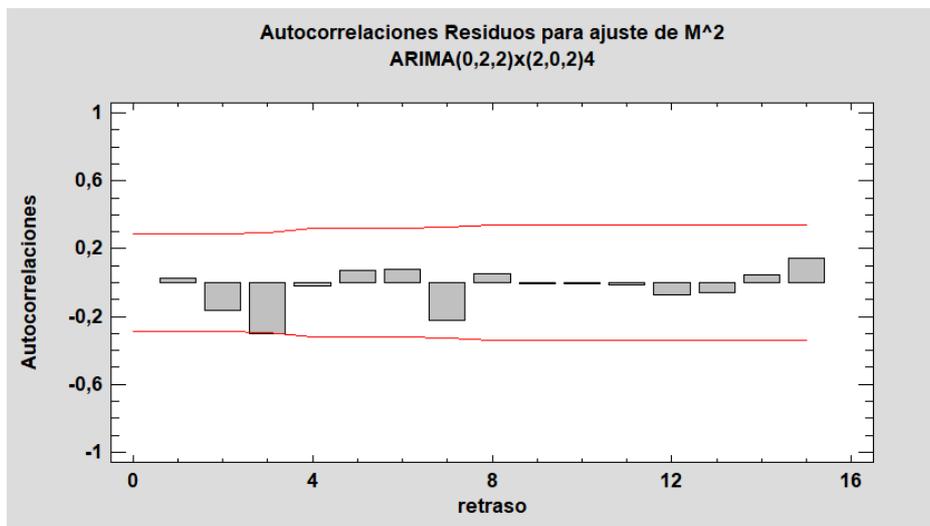
En la tabla 53, se toma el Valor-z para curtosis con P-Valor 0.0712405, este valor es superior a Alfa por tanto no existe evidencia para rechazar la hipótesis alternativa y se afirma que los residuos se distribuyen de forma normal.



Gráfica 110 Histograma Residuos M^2

En la gráfica 110. Se analiza el histograma de residuos y se observa la forma de campana de Gauss por tanto se podría afirmar que los residuos son normales.

Una vez se ha cumplido todos los pasos anteriores lo último que faltaría para validar el modelo es observar el FAS Y FAP del modelo ARIMA



Gráfica 111 FAS Residuos M^2 .

Como se observa en la gráfica 111, existe una autocorrelación de orden 3, por lo que no se puede validar el modelo ARIMA para la variable M^2 .

10.-Análisis de la multicolinealidad

⁵La multicolinealidad aparece cuando dos o más variables explicativas están relacionadas entre sí, o estas toman unos valores semejantes (proporcionales) en la muestra observada, entonces en ese caso se obtiene una situación de multicolinealidad en el modelo.

La multicolinealidad se puede dar de dos maneras diferentes:

- La multicolinealidad se dará cuando existan relaciones lineales, exactas o no, entre las variables explicativas del modelo.
- La multicolinealidad se dará también en el caso de que exista una proporcionalidad entre los valores de las variables explicativas, sin haber realmente una relación entre ellas.

El problema de esta segunda surge de una mala asignación de los valores de las variables explicativas en el experimento a realizar, aunque éste no suele ser el caso más habitual en el análisis económico, dado que los valores utilizados vienen impuestos por la actividad económica. No obstante, en los modelos microeconómicos puede darse este tipo de relaciones.

Si existiendo multicolinealidad se realizara un ajuste, los efectos más destacables y desconcertantes son que las estimaciones de los parámetros obtenidos son globalmente muy significativas, pero ocurre que individualmente no lo son (o viceversa), y también ocurre que las variables son o dejan de ser explicativas según el orden en que se han introducido en el modelo.

La multicolinealidad es un problema porque afecta tanto a las estimaciones de los parámetros como a la varianza de la perturbación, y por lo tanto falsea las pruebas de hipótesis para comprobar si son significativos.

La multicolinealidad hace que las varianzas de los parámetros MCO estimados aumenten de forma drástica con el aumento de la multicolinealidad. Esto puede llevar a rechazar un parámetro como significativo cuando en realidad sí lo es (estadístico t muy pequeño).

Ahora se pasará a ver el problema de la multicolinealidad, se verá si existen relaciones lineales entre las variables explicativas, para esto no hay nada mejor que el coeficiente de correlación lineal para cuantificar dicha relación.

Por lo tanto, la matriz R tiene unos valores en la diagonal principal, y el resto de los elementos de llaman los coeficientes de correlación de cada pareja de las variables explicativas.

La inversa de la matriz de correlación, R^{-1} , sí que tiene en cuenta esa dependencia conjunta. Si en la inversa se obtienen valores por arriba de 10, se puede afirmar que existe una multicolinealidad muy exagerada y se pasará a la eliminación de las variables.

Por último, se realizará el método del índice de acondicionamiento (I.C), y para utilizarlo se necesitan los autovalores de la matriz de correlación, se calcula el I.C y si se dan las siguientes opciones:

⁵ Apuntes Econometría UPV 2021

$1 \leq I.C. \leq 10/20$ no existe multicolinealidad

$10/20 \leq I.C. \leq 20/30$ multicolinealidad moderada

$20/30 \leq I.C.$ existe multicolinealidad elevada

Todas estas pruebas se realizarán hasta que ya no se obtenga más multicolinealidad.

En este apartado se va a tratar de estudiar la multicolinealidad de todas las variables para saber si existen algunas de las variables que tengan relación de dependencia lineal fuerte entre más de dos variables explicativas; en una regresión múltiple que incumple el supuesto de Gauss-Markov cuando es exacta o también conocido cuando la correlación entre 2 o más variables es muy alta.

Para el estudio de la Multicolinealidad se analiza la matriz de correlación, la matriz inversa de correlación y el índice de acondicionamiento, para detectar los posibles problemas de multicolinealidad entre las variables y eliminar aquellas que se presenten problemas.

10.1.-Matriz de correlación 1

	PARADOS	M^2	nºVIVIENDA	PIB	Euribor	DensPobla	Salarios	Nacimientos
Parados	1	-0,314411	-0,703929	-0,709908	-0,294765	0,459725	0,0449301	-0,0390604
M^2	-0,314411	1	0,424791	-0,23894	0,830754	0,523463	-0,307333	0,834338
VivIni	-0,703929	0,424791	1	0,257254	0,528029	-0,263396	-0,277301	0,212461
PIB	-0,709908	-0,23894	0,257254	1	-0,246998	-0,570071	0,551073	-0,460635
EURIBOR	-0,294765	0,830754	0,528029	-0,246998	1	0,414467	-0,295681	0,846073
DENSPOBLA	0,459725	0,523463	-0,263396	-0,570071	0,414467	1	-0,0635796	0,579105
SALARIOS	0,0449301	-0,307333	-0,277301	0,551073	-0,295681	-0,0635796	1	-0,315234
NACIMIENC	-0,0390604	0,834338	0,212461	-0,460635	0,846073	0,579105	-0,315234	1

Tabla 54 Matriz correlación 1 estudio

En la tabla 54, en el análisis de la matriz de correlación analiza la relación que existe entre cada pareja de variables X, y mostrará el porcentaje de autocorrelación que existe entre ambas variables, consideraremos que existe problema de multicolinealidad cuando los datos considerados en valores absolutos sean superiores a 0,7. Se destacan en color naranja aquellos datos que presentan un valor superior a 0,7.

10.2.-Matriz de correlación inversa 1

Parados	30,7490748	9,97483206	3,27490162	28,5952853	2,8829319	-5,50423773	-11,9244601	2,3442556
M^2	9,97483206	9,11588087	-0,24968551	7,50062726	0,37959643	-3,96683202	-2,81485169	-2,61929599
Vivini	3,27490162	-0,24968551	3,64471096	2,39693242	-2,42050925	0,6706971	-0,52636083	2,15958976
PIB	28,5952853	7,50062726	2,39693242	30,2743276	2,66475119	-3,32746704	-13,2244807	3,79861799
EURIBOR	2,8829319	0,37959643	-2,42050925	2,66475119	7,15965123	-1,23205548	-1,45880097	-4,26632354
DENSPOBLA	-5,50423773	-3,96683202	0,6706971	-3,32746704	-1,23205548	4,28227559	1,05494474	0,31451234
SALARIOS	-11,9244601	-2,81485169	-0,52636083	-13,2244807	-1,45880097	1,05494474	7,0540348	-1,25006403
NACIMIENTO	2,3442556	-2,61929599	2,15958976	3,79861799	-4,26632354	0,31451234	-1,25006403	7,60131634

Tabla 55 Matriz correlación inversa 1 estudio

En la tabla 55 se observa la matriz inversa de autocorrelación, en esta matriz se analiza cada variable X, con el resto de las variables y el dato que se debe analizar se encuentra en la diagonal principal, se verificará que existen problemas de multicolinealidad si se encuentran valores superiores a 10 en la diagonal.

Hay 2 variables marcadas de color azul y cuyos números en la diagonal principal son 30.74 y 30.27 que se encuentran marcados en amarillo.

10.3.-Índice de acondicionamiento 1

Componente		Porcentaje d	Porcentaje
Número	Eigenvalor	Varianza	Acumulado
1	3,54776	44,347	44,347
2	2,54499	31,812	76,159
3	1,10777	13,847	90,007
4	0,323138	4,039	94,046
5	0,273107	3,414	97,46
6	0,118808	1,485	98,945
7	0,0698917	0,874	99,818
8	0,0145301	0,182	100

Tabla 56 Índice acondicionamiento 1 estudio.

La tabla 56, se trata de una manera diferente para demostrar si existe o no existe multicolinealidad, para ello se debe realizar la Raíz cuadrada del valor más grande entre el valor más pequeño.

$$\sqrt{\frac{3.5776}{0.0145301}} = 15,6258199$$

Según el IC (Índice de acondicionamiento) al ser mayor que 10 se afirmaría que según este estudio su que existe multicolinealidad.

Por tanto, analizando la matriz de autocorrelación, la matriz inversa de autocorrelación y el índice de acondicionamiento, se concluye que existe problemas de multicolinealidad, para solucionar este problema se debe ir a la matriz inversa de autocorrelación y elegir en la diagonal principal los dos valores más grandes y hacer su diferencia.

Los datos más grandes son: 30.74 correspondiente a la variable Parados y 30.27 correspondiente a la variable PIB.

La diferencia entre las variables parados y PIB es menor a 1 se tendría que analizar dos modelos de regresión, eliminando en el primer modelo una variable (Parados) y en el segundo modelo la otra variable (PIB) y se quedaría con el modelo que mayor R^2 obtenga.

Quitando la variable parados el modelo se queda con una R^2 de 89 por ciento

Quitando la variable de PIB el modelo se queda con una R^2 de 93,68 por ciento

Por lo tanto, se elimina del conjunto de variables independientes la variable PIB y se mantiene la variable parados.

10.4.-Matriz correlación 2

Una vez se ha quitado la variable PIB del modelo se debe demostrar si sigue existiendo multicolinealidad, para ello se vuelve a realizar los mismos análisis, el de matriz de correlación, matriz inversa de correlación e índice de acondicionamiento.

	PARADOS	M^2	nºVIVIENDAS	Euribor	DensPobla	Salarios	Nacimientos
PARADOS	1	-0,314411	-0,703929	-0,294765	0,459725	0,0449301	-0,0390604
M^2	-0,314411	1	0,424791	0,830754	0,523463	-0,307333	0,834338
nºVIVIENDAS	-0,703929	0,424791	1	0,528029	-0,263396	-0,277301	0,212461
Euribor	-0,294765	0,830754	0,528029	1	0,414467	-0,295681	0,846073
DensPobla	0,459725	0,523463	-0,263396	0,414467	1	-0,0635796	0,579105
Salarios	0,0449301	-0,307333	-0,277301	-0,295681	-0,0635796	1	-0,315234
Nacimientos	-0,0390604	0,834338	0,212461	0,846073	0,579105	-0,315234	1

Tabla 57 Matriz correlación 2 estudio

En la tabla 57, se observa los valores que en valor absoluto son superiores a 0,7 y se colorean de color naranja.

10.5.- Matriz inversa correlación 2

PARADOS	3,7397106	2,89019655	1,0109053	0,36597028	-2,36131511	0,56657866	-1,2436875
M^2	2,89019655	7,25756018	-0,84353839	-0,28060999	-3,1424342	0,46158443	-3,560424
nºVIVIENDAS	1,0109053	-0,84353839	3,4549368	-2,63148763	0,93414518	0,52067107	1,85883889
Euribor	0,36597028	-0,28060999	-2,63148763	6,9250994	-0,9391713	-0,29478005	-4,60067851
DensPobla	-2,36131511	-3,1424342	0,93414518	-0,9391713	3,91655197	-0,39856479	0,73202041
Salarios	0,56657866	0,46158443	0,52067107	-0,29478005	-0,39856479	1,27729576	0,40925442
Nacimientos	-1,2436875	-3,560424	1,85883889	-4,60067851	0,73202041	0,40925442	7,12469144

Tabla 58 Matriz inversa correlación 2 estudio

En la tabla 58, no se observan valores superiores a 10 en la diagonal principal, por tanto, en este análisis no se observa problemas de multicolinealidad.

Se observa que haciendo la matriz inversa se ve como existen 2 variables marcadas de color azul y cuyos números de la matriz principal (la diagonal de la matriz) están marcados en amarillos debido a que la cantidad de autocorrelación es muy elevada. Entre estas dos variables al ser la diferencia menor a 1 habría que analizar dos modelos de regresión, eliminando en el primer modelo una variable (M^2) y en el segundo modelo la otra variable (Nacimientos) y se quedaría con el modelo que mayor R^2 obtenga.

Si se quita la variable nacimientos se obtiene un R^2 de 93,3002 por ciento.

Si se quita la variable M^2 se obtiene una R^2 de 91,8842 por ciento

Según el criterio habría que quitar la variable de número de nacimientos debido a que la variable M^2 mejora más el modelo que la otra variable.

10.6.-Índice de acondicionamiento 2

Componente		Porcentaje d	Porcentaje
Número	Eigenvalor	Varianza	Acumulado
1	3,3878	48,397	48,397
2	1,9998	28,569	76,966
3	0,885792	12,654	89,62
4	0,289042	4,129	93,749
5	0,265987	3,8	97,549
6	0,105063	1,501	99,05
7	0,0665142	0,95	100

Tabla 59 Índice acondicionamiento 2 estudio

En la tabla 59, sirve para comprobar de otra manera si existe o no existe multicolinealidad, para ello se debe realizar la Raíz cuadrada del valor más grande entre el valor más pequeño.

$$\sqrt{\frac{3.3878}{0.0665142}} = 7,136769919$$

Según el IC (Índice de acondicionamiento) al ser menor que 10 diríamos que no existe multicolinealidad.

Por tanto, dado que la matriz de correlación si tiene valores superiores a 0,7 a pesar de que las otras pruebas determinen que no existe problema de multicolinealidad, se afirma que sí que hay problema de multicolinealidad y se procede a su solución partiendo de la matriz inversa de autocorrelación.

Se observa en la matriz inversa 2 variables marcadas de color azul y cuyos números en la diagonal principal son 7,25 y 7,12 la diferencia entre estas dos variables al ser menor a 1 se debe analizar dos modelos de regresión, eliminando en el primer modelo una variable (M^2) y en el segundo modelo la otra variable (Nacimientos) y se quedaría con el modelo que mayor R^2 obtenga.

Si se quita la variable nacimientos se obtiene un R^2 de 93,3002 por ciento.

Si se quita la variable M^2 se obtiene una R^2 de 91,8842 por ciento

De acuerdo con el criterio anterior se debe eliminar del modelo la variable nacimientos puestos que la variable M^2 mejora el modelo.

10.7.-Matriz correlación 3

Una vez se han eliminado ya la segunda variable del modelo se debe demostrar que no existe multicolinealidad, para ello se va a volver a realizar los mismos análisis, pero quitando ya del modelo la variable PIB y la variable Nacimientos.

	PARADOS	M ²	nºVIVIENDAS	Euribor	DensPobla	Salarios
PARADOS	1	-0,314411	-0,703929	-0,294765	0,459725	0,0449301
M ²	-0,314411	1	0,424791	0,830754	0,523463	-0,307333
nºVIVIENDAS	-0,703929	0,424791	1	0,528029	-0,263396	-0,277301
Euribor	-0,294765	0,830754	0,528029	1	0,414467	-0,295681
DensPobla	0,459725	0,523463	-0,263396	0,414467	1	-0,0635796
Salarios	0,0449301	-0,307333	-0,277301	-0,295681	-0,0635796	1

Tabla 60 Matriz correlación 3 estudio.

En la tabla 60, se observa cómo se marcan todos los superiores a 0,7 tanto en símbolo negativo como en símbolo positivo, se observa cómo están marcados de color naranja cada uno de ellos.

10.8.-Matriz inversa correlación 3

PARADOS	3,52261227	2,26868826	1,33538457	-0,43712504	-2,23353349	0,63801819
M ²	2,26868826	5,47830852	0,08537967	-2,57970832	-2,77662144	0,66610125
nºVIVIENDAS	1,33538457	0,08537967	2,96996394	-1,43116616	0,74316035	0,41389619
Euribor	-0,43712504	-2,57970832	-1,43116616	3,95427001	-0,46647847	-0,03050924
DensPobla	-2,23353349	-2,77662144	0,74316035	-0,46647847	3,84134115	-0,44061329
Salarios	0,63801819	0,66610125	0,41389619	-0,03050924	-0,44061329	1,25378749

Tabla 61 Matriz inversa correlación 3 estudio.

En la tabla 61, no se observan valores superiores a 10 en la diagonal principal, por tanto, en este análisis no se observa problemas de multicolinealidad.

10.9.- Índice acondicionamiento 3

Número	Eigenvalor	Varianza	Acumulado
1	2,71909	45,318	45,318
2	1,82923	30,487	75,805
3	0,882925	14,715	90,521
4	0,284665	4,744	95,265
5	0,184492	3,075	98,34
6	0,099601	1,66	100

Tabla 62 Índice acondicionamiento 3 estudio.

La tabla 62, es otra manera de estudiar la multicolinealidad con la siguiente formula:

$$\sqrt{\frac{2.71909}{0.099601}} = 5,22492357$$

Mediante este estudio de multicolinealidad al ser menor que 10 se podría decir que no existe multicolinealidad.

Por tanto, dado que la matriz de correlación si tiene valores superiores a 0,7 a pesar de que las otras pruebas determinen que no existe problema de multicolinealidad, se afirma que sí que hay problema de multicolinealidad y se procede a su solución partiendo de la matriz inversa de autocorrelación.

Se observa existen 2 variables marcadas de color azul y valores en la diagonal principal son 5,47 y 3,95 se encuentran marcados en amarillo. Entre estas dos variables la diferencia ya no es menor a 1 y según la teoría habría que eliminar directamente la variable de M^2 , pero si se hace los dos modelos de regresión se observa como si quitamos la variable M^2 se obtiene peor modelo que si quitamos la variable Euribor. Es por eso por lo que sí o sí habrá que quedarse con la variable M^2 debido a que nos mejora el modelo bastante más.

Eliminando la variable Euribor R^2 de 92,3283 por ciento.

Eliminando la variable M^2 obtenemos una R^2 de 89,6827 por ciento.

En este caso se acabaría quitando la variable del Euribor, debido a que la variable M^2 mejora el modelo.

Una vez se ha terminado con todos los estudios de multicolinealidad, se plantea un modelo de regresión con las variables independientes:

1. Parados
2. M^2
3. Nº Viviendas Iniciadas.
4. Densidad Población.
5. Salarios.

11.-Realización de un modelo de regresión

Después de hacer un análisis de la multicolinealidad y descartar las variables que daban problemas debido a la gran relación que tenían, se va a realizar el análisis con un modelo de regresión.

11.1.-Modelo teórico y ajustado

$$\text{IPV} = B_0 + B_1 * \text{Parados} + B_2 * M^2 + B_3 * \text{N}^\circ \text{VivInici} + B_4 * \text{DesPobla} + B_5 * \text{Salario} + U$$

11.2.-Análisis de la significatividad

En primer lugar, se va a tratar de analizar la significatividad del modelo, para saber si tiene una buena R^2 o no.

11.2.1.-Contraste de significación global del modelo

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	638190,	5	127638,	5117,24	0,0000
Residuo	1072,54	43	24,9428		
Total	639263,	48			

Tabla 63 Análisis Varianza Modelo Regresión

R-cuadrada = 99,8322 por ciento

R-cuadrado (ajustado para g.l.) = 99,8166 por ciento

Error estándar del est. = 4,99427

Error absoluto medio = 3,52722

Estadístico Durbin-Watson = 0,967369

Autocorrelación de residuos en retraso 1 = 0,50553

Validación del modelo

$H_0: B_1 = B_2 = B_3 = B_4 = B_5 = 0$

$H_1: \text{Al menos uno es } \neq 0 \quad \alpha=0,05$

P-Valor del Modelo = 0,0000

En la tabla 63, como el p-valor del modelo es 0,0000 por lo que es menor que alfa, se rechaza la Hipótesis nula por lo que se puede concluir que el modelo sí es significativo.

A continuación, se va a realizar un contraste significación individual (parámetros)

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
Parados	-0,00858787	0,00108407	-7,92188	0,0000
M^2	0,193674	0,0154473	12,5378	0,0000
nº viviendas iniciadas	-0,00181805	0,000545029	-3,3357	0,0018
DensPobla	0,0000251484	0,00000444177	5,6618	0,0000
Salario	-0,00531331	0,0107682	-0,493426	0,6242

Tabla 64 Análisis Varianza Individua Modelo Regresión

Validación Parámetros

$H_0: B_0 = 0$

$H_1: B_0 \neq 0 \quad \alpha=0,05$

11.2.1.1.-Parados

P-Valor – 0,0000

Como el P-Valor es 0,0000 la constante es significativa puesto que el p-valor es menor a alfa, por lo que se rechaza la H_0 .

11.2.1.2.- M^2

P-Valor – 0,0000

Como el P-Valor es 0,0000 la constante es significativa puesto que el p-valor es menor a alfa, por lo que se rechaza la H_0 .

11.2.1.3.- N° Viviendas Iniciadas

P-Valor – 0,0018

Como el P-Valor es 0,0019 la constante es significativa puesto que el p-valor es menor a alfa, por lo que se rechaza la H_0 .

11.2.1.4.-Densidad Población

P-Valor – 0,0000

Como el P-Valor es 0,0000 la constante es significativa puesto que el p-valor es menor a alfa, por lo que se rechaza la H_0 .

11.2.1.5.-Salarios.

P-Valor- 0,6242

Como el P-Valor es 0,6242 la constante no es significativa puesto que el p-valor es mayor a alfa, por lo que no existe evidencia para rechazar la hipótesis alternativa.

11.3.-Detección gráfica de posibles problemas.

Una vez se ha estimado el modelo de regresión⁶, se debe comprobar si se cumplen las hipótesis referentes al error o perturbación, a las variables explicativas y explicadas, a los parámetros del modelo.

Los gráficos más importantes para el análisis de los residuos y que se van a tratar de analizar en el siguiente apartado son:

1. Gráfico de los residuos frente a los valores estimados. (Vs \hat{Y}).
2. Gráfico de los residuos frente a las variables explicativas. (Vs X)

1. Respecto a los gráficos de los residuos frente a la estimación de la variable explicada pueden indicar problemas de heterocedasticidad ($\sigma^2(U_i) = \sigma_i^2$), falta de linealidad, valores atípicos o autocorrelación ($cov(U_i, U_j) \neq 0$). Si existe alguno de estos problemas en el modelo, entonces los gráficos de residuos frente a la estimación lo mostrarán.

Se va a explicar casos donde se presentan problemas:

Heterocedasticidad. (Varianza del error o perturbación no se mantiene constante, lo que provoca que no tenga altura constante)

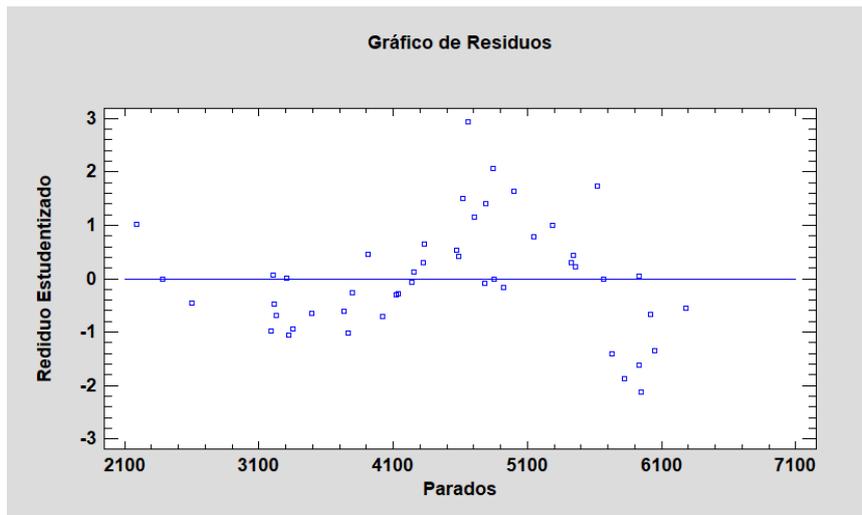
2. Los gráficos de los residuos frente a las variables explicativas ayudan a identificar los problemas del modelo como debidos a alguna variable explicativa en particular.

⁶ Apuntes Econometría UPV 2021

Los problemas que se pueden presentar en el estudio de los gráficos son los mismos que los que se han visto anteriormente

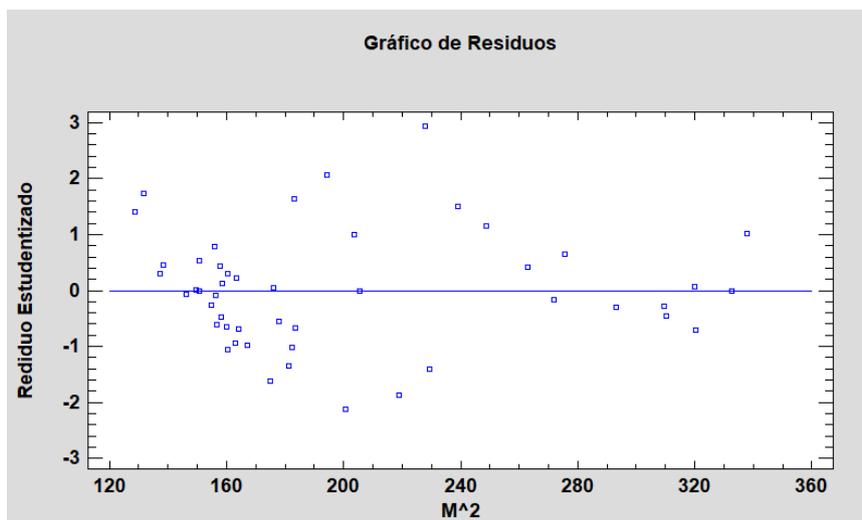
11.3.1.-Residuos Vs X

Se va a realizar el modelo de regresión con todas las variables y se observan todos los gráficos vs Residuos.



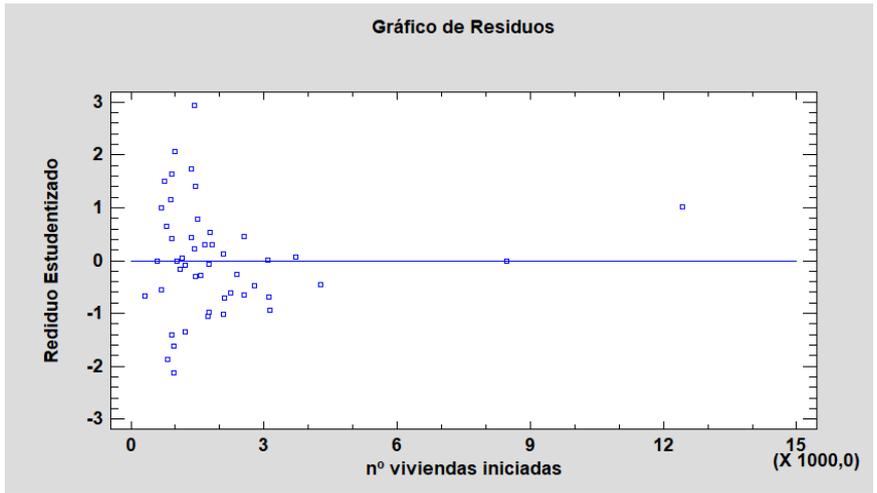
Gráfica 112 Residuos vs Parados Modelo Regresión

En la gráfica 112, se observa una dispersión aleatoria de puntos que no se asocia a ningún problema, se puede apreciar la existencia de algún punto anómalo ya que se acerca al 3.



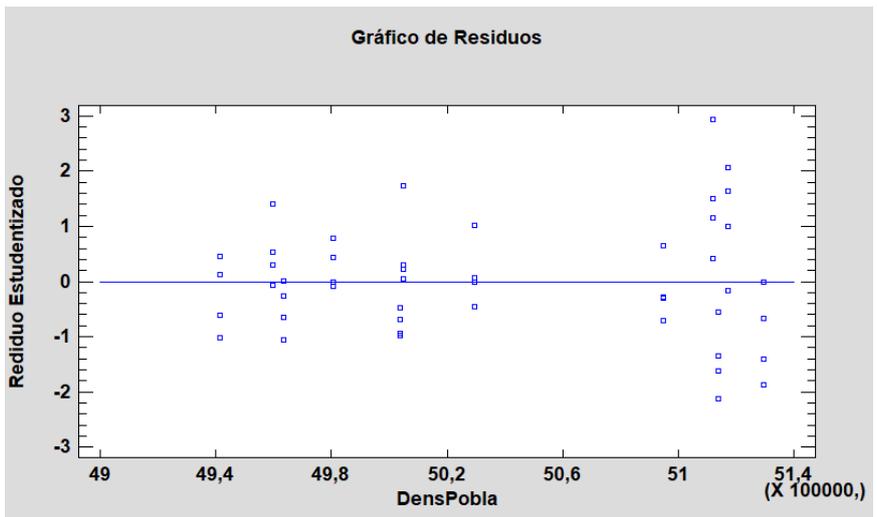
Gráfica 113 Residuos vs M^2 Modelo Regresión

En la gráfica 113, se observa una dispersión aleatoria de puntos que no se asocia a ningún problema y se observa un punto anómalo cerca del 3.



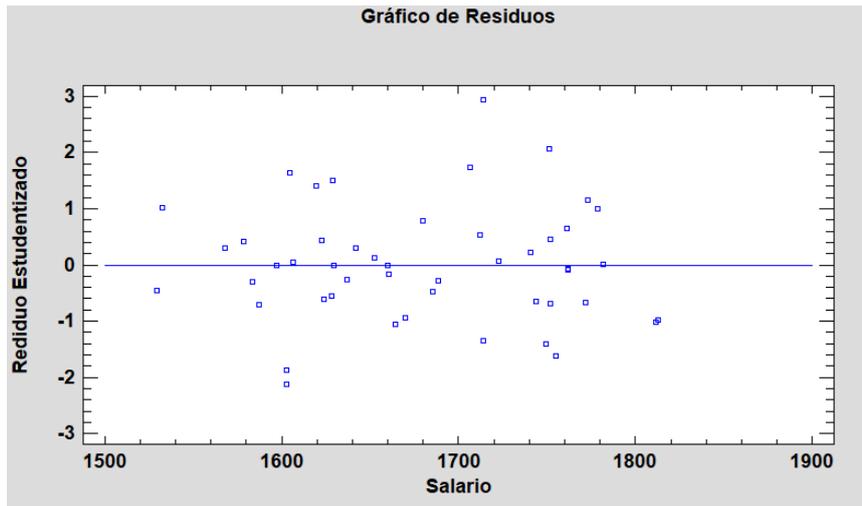
Gráfica 114 Residuos vs Vivlni Modelo Regresión

En la gráfica 114, se observa una dispersión aleatoria de puntos que no se asocia a ningún problema además se aprecian 2 puntos anómalos.



Gráfica 115 Residuos vs DensPobla Modelo Regresión

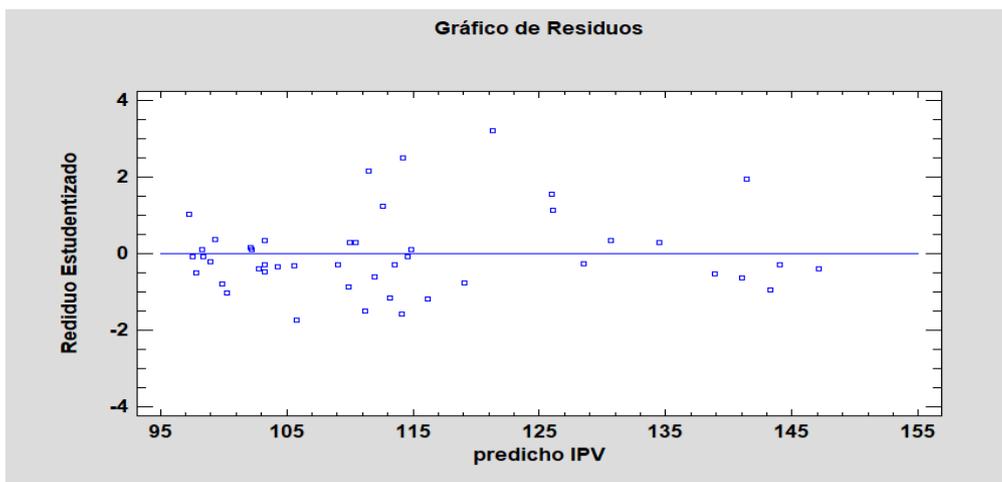
En la gráfica 115, se observa una dispersión aleatoria de puntos que no se asocia a ningún problema, además se observan dos puntos anómalos.



Gráfica 116 Residuos vs Salario Modelo Regresión

En la gráfica 116, se observa una dispersión aleatoria de puntos que no se asocia a ningún problema, salvo un punto anómalo en la parte superior.

11.3.2.-Residuos VS Y predicho



Gráfica 117 Residuos vs Predicho IPV Modelo Regresión.

Se observa una dispersión aleatoria de puntos que no se asocia a ningún problema.

Antes de estudiar la normalidad de residuos hemos eliminado la variable Salarios de nuestro modelo de regresión debido a que no era una variable significativa.

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	638184,	4	159546,	6508,39	0,0000
Residuo	1078,61	44	24,5139		
Total	639263,	48			

Tabla 65 Análisis Varianza Modelo Regresión 2

R-cuadrada = 99,8313 por ciento

R-cuadrado (ajustado para g.l.) = 99,8198 por ciento

Error estándar del est. = 4,95115

Error absoluto medio = 3,53539

Estadístico Durbin-Watson = 0,928746

Autocorrelación de residuos en retraso 1 = 0,52366

En la tabla 65, vuelve a ser todo significativo y el R^2 del modelo es de casi un 100%, ahora se pasará a realizar el análisis individual de las variables.

Parámetro	Estimación	Error		Estadístico T	Valor-P
		Estándar			
Parados	-0,00839385	0,00100153		-8,38107	0,0000
M ²	0,197126	0,0136535		14,4378	0,0000
nº viviendas iniciadas	-0,00175409	0,000524818		-3,34228	0,0017
DensPobla	0,0000230407	0,00000120698		19,0895	0,0000

Tabla 66 Análisis Varianza Individual Modelo Regresión 2

En la tabla 66, se observa como ahora todas son significativas y se puede seguir adelante con los siguientes estudios.

11.4.-Estudio de normalidad de residuos

Para seguir con el análisis de regresión, el siguiente paso es estudiar si los residuos del modelo son normales o no.

11.4.1.-Contraste de hipótesis

Se realizará el contraste de hipótesis para conocer la normalidad de los residuos. Para ello se escogerá el P-Valor más pequeño, es decir, el más restrictivo.

Prueba	Estadístico	Valor-P
Chi-Cuadrado	9,75	0,835164
Estadístico W de Shapiro-Wilk	0,980308	0,591569
Valor-Z para asimetría	0,864677	0,387214
Valor-Z para curtosis	0,691301	0,489374

Tabla 67 Normalidad Residuos Modelo Regresión

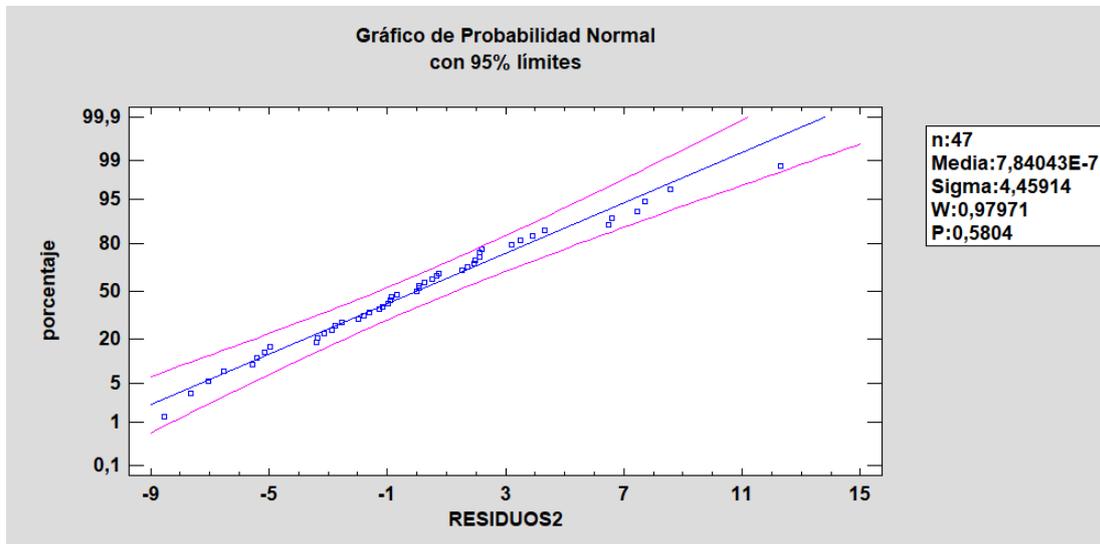
Test normalidad residuos

H₀: Los residuos se distribuyen normalmente

H₁: Los residuos no se distribuyen normalmente $\alpha > 0,05$

En la tabla 67, se escoge el Valor-z para asimetría por resultar el dato más restrictivo con un P-Valor de 0,387214, este valor es mayor a alfa, 0,05, por tanto, no existe evidencia para rechazar la hipótesis alternativa, y se afirma que los residuos son normales.

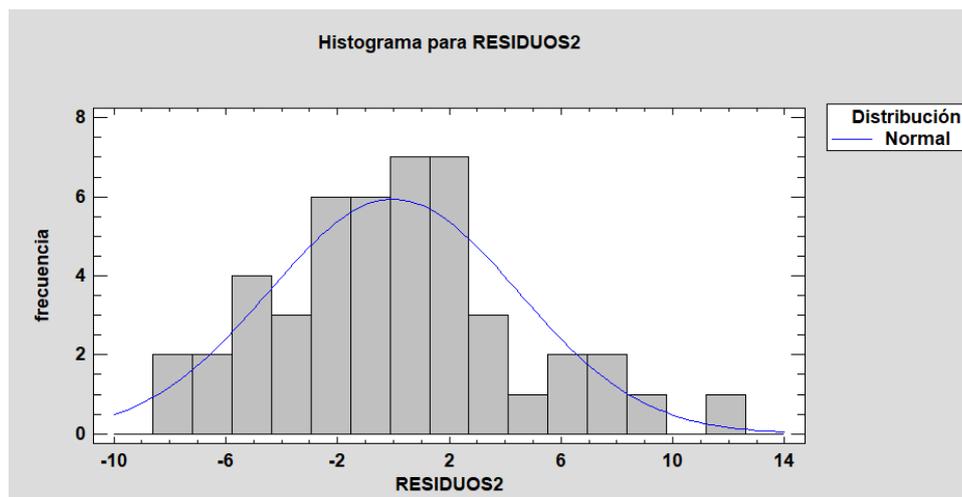
11.4.2.-Gráfico probabilístico



Gráfica 118 Probabilidad Normal Residuos Modelo Regresión.

En la gráfica 118, se observa que los residuos se distribuyen a lo largo de la recta o muy próximos a esta, por tanto, se podría afirmar que los residuos se comportan de forma normal.

11.4.3.-Histograma de residuos

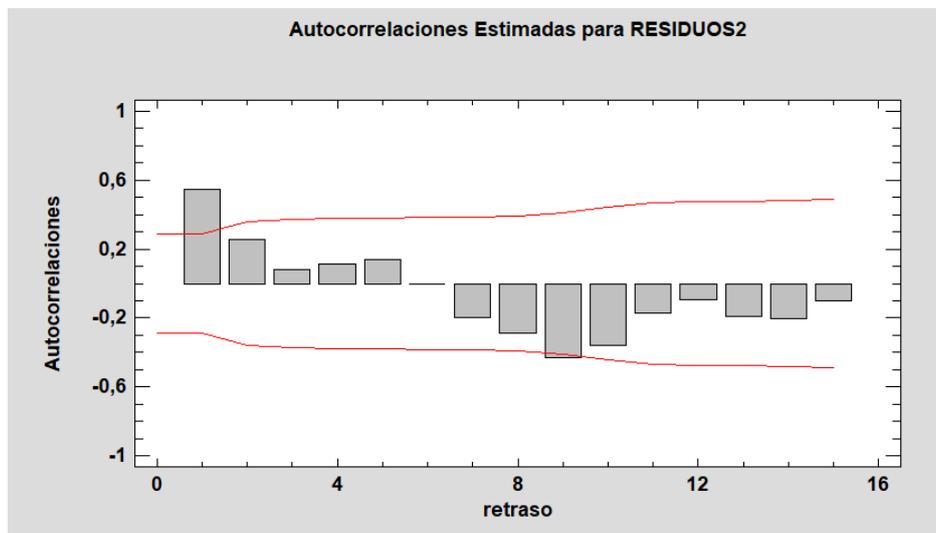


Gráfica 119 Histograma Modelo Regresión

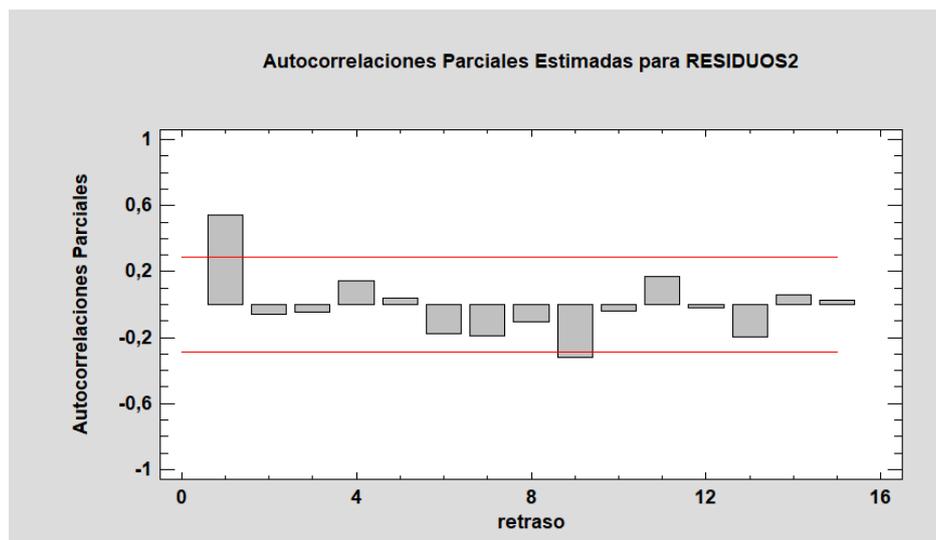
En la gráfica 119, Gráficamente se observaría que los residuos se distribuyen en forma de campana de Gauss. Por tanto, se podría afirmar que los residuos se distribuyen de forma normal. Finalmente, una vez realizado los 3 estudios de normalidad de residuos se concluirá el presente análisis afirmando que los residuos son normales.

11.5.-Estudio autocorrelación

En cuanto al estudio de la autocorrelación para los Residuos del modelo de regresión se muestran los gráficos FAS Y FAP para su análisis.



Gráfica 120 FAS Modelo Regresión



Gráfica 121 FAP Modelo Regresión

En las gráficas 120 y 121, se puede observar que existen problemas de autocorrelación de primer orden positivo.

11.6.-Estudio de la heterocedasticidad

La hipótesis de varianza ⁷ constante se utilizaba en la estimación de la varianza de la perturbación. Las perturbaciones (Econometría EPSA 2021) U_i tienen distribución normal de media cero y desviación típica $\sigma_i = \sigma$ (constante), y de esta forma la desviación típica del error, σ , puede estimarse como la desviación típica de todos los residuos.

Ahora bien, si las desviaciones típicas no son constantes, no tiene sentido hablar de una varianza de la perturbación, ni del error cometido en las estimaciones de los parámetros, ni de las predicciones del modelo y mucho menos de las pruebas de hipótesis t y F.

La heterocedasticidad puede aparecer en el modelo por varios motivos:

1. Alguna variable explicativa utilizada en el modelo es heteroscedasticidad.
2. Se ha introducido artificialmente la heterocedasticidad en los datos.
3. Se ha especificado incorrectamente el modelo.
4. Hay valores atípicos entre los datos con los que se ha hecho el ajuste.

Se tendrá que proponer el modelo que explica el promedio del cuadrado del residuo.

$$e^2 = Y_0 + Y_1 X \text{ Salarios} + U$$

Después de esto se plantea la prueba de hipótesis de la heterocedasticidad.

$H_0: Y_1 X = 0$ y por lo tanto no existe heterocedasticidad

$H_1: Y_1 X \neq 0$ y por lo tanto existe heterocedasticidad

		<i>Error</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimación</i>	<i>Estándar</i>	<i>T</i>	<i>Valor-P</i>
CONSTANTE	-1682,53	489,55	-3,43688	0,0013
Parados	-0,00447303	0,00629918	-0,710098	0,4815
M ²	-0,280075	0,115045	-2,43448	0,0191
nº viviendas iniciadas	0,00187713	0,00283586	0,661927	0,5116
DensPobla	0,000352294	0,000103844	3,39254	0,0015

Tabla 68 Heterocedasticidad

11.6.1.-Parados

$H_0: \sigma^2 = \text{constante} \rightarrow$ no existe heterocedasticidad

$H_1: \sigma^2 \neq \text{constante} \rightarrow$ existe heterocedasticidad $\alpha = 0,05$

Como P-Valor 0.4815, es mayor que alfa, 0,05, no existe evidencia para rechazar la hipótesis alternativa y se concluye que la variable Parados no genera problemas de heterocedasticidad.

11.6.2.-M²

$H_0: \sigma^2 = \text{constante} \rightarrow$ no existe heterocedasticidad

$H_1: \sigma^2 \neq \text{constante} \rightarrow$ existe heterocedasticidad $\alpha = 0,05$

⁷ Fuente: Econometría, EPSA 2021

Como P-Valor, 0,0191, es menor que alfa, 0,05, se rechaza la hipótesis nula y se concluye que la variable M^2 genera problemas de heterocedasticidad.

11.6.3.-Nº Viviendas Iniciadas

$H_0: \sigma^2 = \text{constante} \rightarrow$ no existe heterocedasticidad

$H_1: \sigma^2 \neq \text{constante} \rightarrow$ existe heterocedasticidad $\alpha = 0,05$

Como P-Valor, 0,5116, es mayor que alfa, 0,05, no existe evidencia para rechazar la hipótesis alternativa y se concluye en que la variable Nº viviendas iniciadas no genera problemas de heterocedasticidad

11.6.4.-Densidad Población

$H_0: \sigma^2 = \text{constante} \rightarrow$ no existe heterocedasticidad

$H_1: \sigma^2 \neq \text{constante} \rightarrow$ existe heterocedasticidad $\alpha = 0,05$

Como P-Valor, 0,0015, es menor que alfa, 0,05, se rechaza la hipótesis nula y se concluye que la variable densidad población genera problemas de heterocedasticidad.

Debido a que existen problemas de heterocedasticidad, se debería plantear un nuevo modelo que, para solucionar este problema, no obstante, se conoce que el modelo plantea problemas de autocorrelación en los residuos, se decide dejar el modelo de regresión normal para plantear un nuevo modelo de regresión penalizado.

Para realizar dicho modelo de regresión se va a pasar a otro programa, debido a que el Statgraphics se encuentra limitado para realizar algunos cálculos, es por esa razón que los cálculos se realizarán en el programa R-studio.

12.- Nuevo modelo de regresión penalizado basado en el método de Elastic Net.

Después del análisis anterior se concluye que el modelo de regresión no es válido por presentar problemas de autocorrelación. Llegados a este punto se vuelve a empezar un nuevo modelo, esta vez como modelo de Regresión Penalizado, para elegir de manera más consistente las variables a considerar.

En el anexo II se adjunta los comandos y fórmulas que se han introducido en el R Studio para el desarrollo del nuevo modelo de regresión penalizado.

El Método de Regresión Penalizado es un método que ajusta el modelo de regresión imponiendo una restricción o penalización sobre los coeficientes de regresión sesgando dichos coeficientes hacia cero, mediante este sesgo se produce una reducción en la complejidad del modelo y por lo tanto se controla el riesgo de sobreajuste.

Entre los métodos penalizados que podemos estudiar encontramos entre otros:

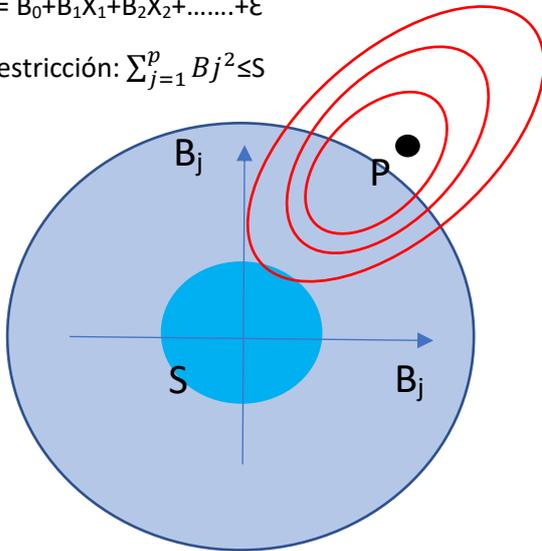
- Método Regresión Ridge
- Método LASSO (LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR)

- Método Elastic Net

12.1.-Método de regresión Ridge.⁸

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + \epsilon$$

$$\text{Restricción: } \sum_{j=1}^p B_j^2 \leq S$$



Se añade una restricción al modelo que controla la complejidad: Cuando s es grande la solución es el modelo lineal normal, cuando s es pequeña la solución reduce todos los Betas hacia cero.

Esto nos permite ajustar modelos con miles de variables.

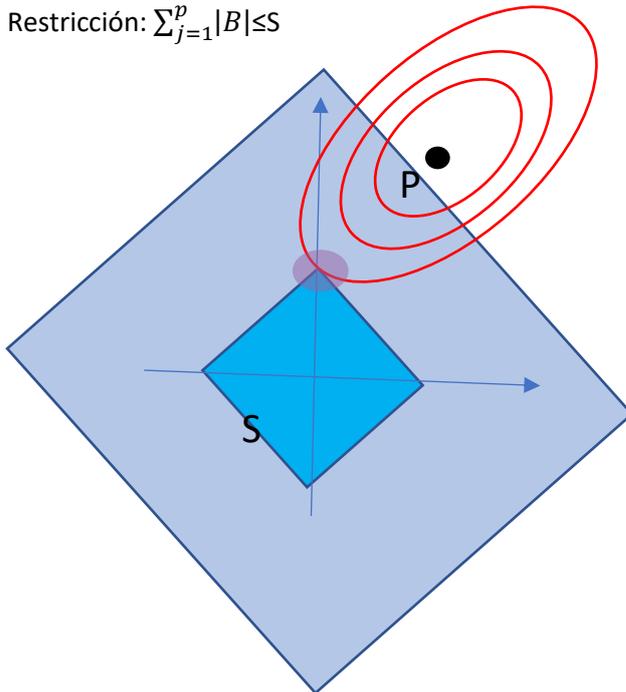
Ridge se caracteriza por ser capaz de ajustar el modelo en presencia de multicolinealidad y por no realizar selección de variables, es decir no se elimina ninguna variable del modelo, por tanto, este método es descartado en la elaboración del presente TFG.

⁸ Apuntes profesor David Hervás

12.2.-Metodo de regresión Lasso.⁹

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + \epsilon$$

$$\text{Restricción: } \sum_{j=1}^p |B_j| \leq S$$



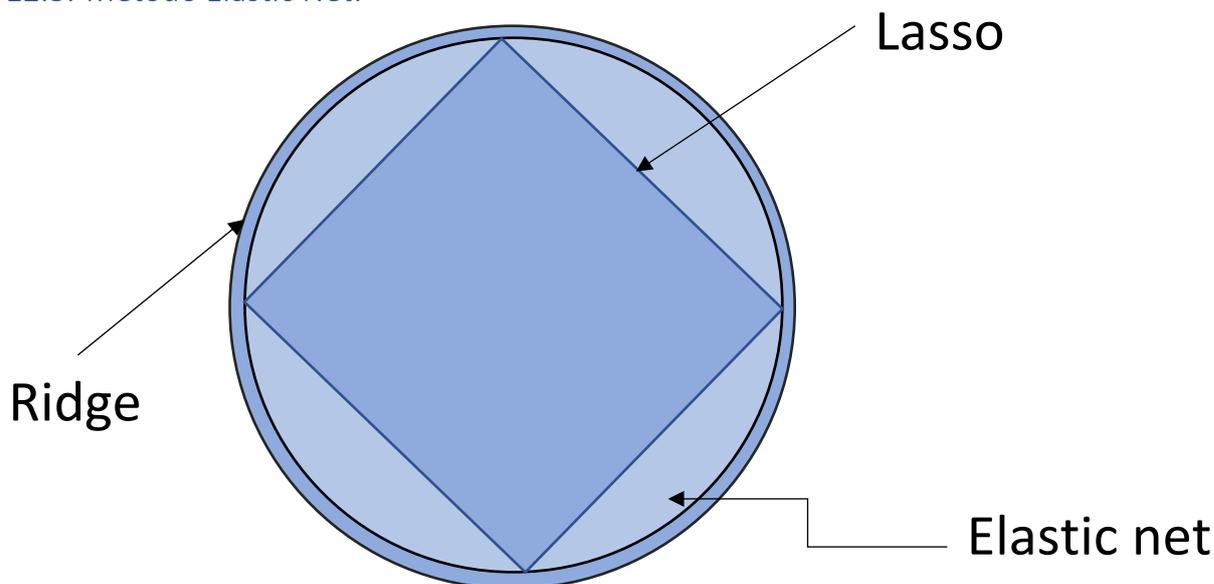
En este caso la restricción es capaz de forzar que la estimación de muchos de los coeficientes sea cero, por lo que se realiza una selección de variables al mismo tiempo que se ajusta el modelo.

Esta característica de penalizar hacia cero simplifica más aún los modelos.

El sistema Lasso sí que permite eliminar variables, pero tiene un problema cuando se tienen dos o más variables muy correlacionadas, es decir, en presencia de multicolinealidad es inestable a la hora de seleccionar las variables, por tanto, este método no se utiliza en el presente TFG.

⁹ Apuntes profesor David Hervás

12.3.-Metodo Elastic Net.¹⁰



El método Elastic Net permite combinar ambos tipos de penalización (Lasso y Ridge), a través de esta combinación se conservan características importantes de cada una de las aproximaciones (de ridge nos quedamos con su capacidad de ajustar el modelo en presencia de multicolinealidad y de lasso con la capacidad de selección de variables), siendo este método el elegido para la realización del nuevo modelo de Regresión.

En este apartado del trabajo, se utiliza como herramienta de trabajo el programa de R Studio para la utilización de comandos automáticos que ayuden al mejor análisis.

12.3.1.- Formulación teórica del modelo penalizado Elastic net

$$\hat{\beta}^{\text{enet}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j)^2 \quad \text{Subject to: } \alpha \sum_{j=1}^p \beta_j^2 + (1 - \alpha) \sum_{j=1}^p |\beta_j| \leq s$$

Esta fórmula muestra por un lado la ecuación de mínimos cuadrados sujeta a las restricciones de Ridge y Lasso, ambas restricciones se suman de forma ponderada en función de un Alfa que se escoge de forma arbitraria en función de la multicolinealidad que haya en la muestra, este alfa se encuentra siempre comprendido entre los valores 0 y 1, en el caso que decidamos un Alfa de 1, significa que se escoge por el modelo Ridge, y en el caso de escoger un Alfa de Cero se escoge por un modelo Lasso. Cabe destacar que la penalización se encuentra sujeta a un factor de penalización s es decir indica cuanto se desea restringir el modelo (para el cálculo de esta restricción es necesario la utilización de la validación cruzada).

A continuación, se identifica los diferentes apartados de la formulación principal

¹⁰ Apuntes profesor David Hervás

12.3.1.1-Fórmula para minimizar la suma de cuadrados (regresión lineal)

$$\hat{\beta}^{\text{enet}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j)^2$$

12.3.1.2.- Fórmula de las restricciones de Ridge y Lasso ponderadas en función de Alfa

$$\text{Subject to: } \alpha \sum_{j=1}^p \beta_j^2 + (1 - \alpha) \sum_{j=1}^p |\beta_j| \leq s$$

12.3.1.3-Fórmula de la restricción de Ridge

$$\alpha \sum_{j=1}^p \beta_j^2 -$$

12.3.1.4.- Fórmula restricción Lasso

$$(1 - \alpha) \sum_{j=1}^p |\beta_j| \leq s$$

Siendo S factor penalización.

Para la realización del cálculo del modelo de regresión penalizado Elastic Net se decide un valor a Alfa de 0,7 para evitar problemas de multicolinealidad.

```
Call: glmnet(x = X, y = Y, family = "gaussian", alpha = 0.7, relax = TRUE)
Relaxed
```

	Df	%Dev	%Dev R	Lambda
1	0	0.00	0.00	19.9400
2	1	9.84	76.90	18.1700
3	1	18.61	76.90	16.5600
4	1	26.40	76.90	15.0900
5	1	33.27	76.90	13.7500
6	1	39.32	76.90	12.5300
7	2	44.80	77.29	11.4100
8	3	50.39	87.44	10.4000
9	3	56.09	87.44	9.4750
10	3	60.95	87.44	8.6330
11	3	65.08	87.44	7.8660
12	3	68.58	87.44	7.1680
13	3	71.55	87.44	6.5310
14	3	74.07	87.44	5.9510
15	4	76.25	88.44	5.4220
16	4	78.16	88.44	4.9400
17	5	79.94	92.48	4.5010
18	5	81.52	92.48	4.1020
19	5	82.89	92.48	3.7370
20	5	84.08	92.48	3.4050
21	5	85.11	92.48	3.1030
22	5	86.01	92.48	2.8270
23	5	86.88	94.80	2.5760
24	5	87.89	94.80	2.3470
25	5	88.78	94.80	2.1390
26	5	89.57	94.80	1.9490
27	5	90.25	94.80	1.7750
28	5	90.85	94.80	1.6180
29	5	91.38	94.80	1.4740
30	5	91.84	94.80	1.3430

Imagen 2 Modelo de regresión penalizado Elastic Net. Fuente: R Studio

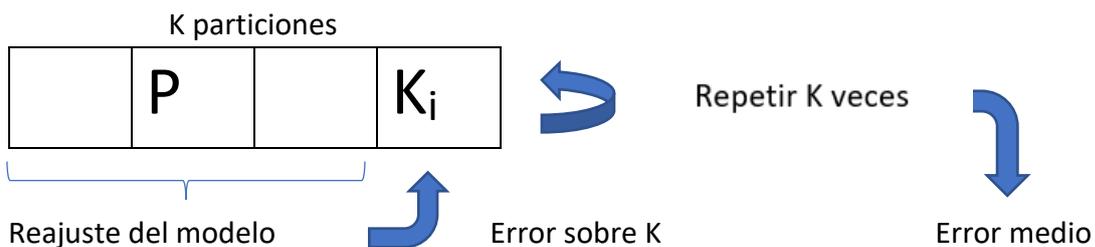
12.3.2.-Validación cruzada y su ejecución

La validación cruzada es una técnica que se utiliza para buscar el valor óptimo de lambda en nuestro modelo penalizado, que será el que produzca el menor error posible en una estimación sobre datos independientes. Se procede a la utilización de un ejemplo sencillo para explicar en qué consiste la presente técnica.

Se tienen 100 observaciones y se ajusta un modelo, a continuación, se puede probar el modelo creado sobre las 100 observaciones, y observar el error que existe entre las 100 observaciones y el modelo, ese error (error aparente) que se obtiene no se puede generalizar y utilizar como referencia cuando aparecen nuevos datos (error real), ya que se tendría un problema de sobreajuste, que se soluciona con la validación cruzada.

Utilizando el ejemplo anterior, la técnica de la validación cruzada consiste en coger las 100 observaciones y dividir las en partes iguales, como por ejemplo en 10 partes iguales, de esas 10 partes se escogerían 9 partes y se procedería a ajustar un nuevo modelo. Ese modelo ajustado se compararía con las 10 observaciones que no se han tenido en cuenta y se obtiene un error. Este proceso se debe repetir teniendo en cuenta las diferentes combinaciones, consiguiendo así los diferentes errores.

Validación cruzada



	lambda.min	lambda.1se	lambda.min.relaxed	lambda.1se.relaxed	gamma.min	gamma.1se
[1,]	0.007336291	0.11956326	0.008836588	0.2516692	0.00	0.25
[2,]	0.024588332	0.25166920	0.173465818	0.7002845	0.00	0.25
[3,]	0.008051570	0.19037852	0.062340383	0.5813885	0.00	0.00
[4,]	0.008836588	0.17346582	0.131220530	0.2516692	0.00	0.50
[5,]	0.007336291	0.11956326	0.190378521	0.2762066	0.00	0.00
[6,]	0.007336291	0.17346582	0.208940192	0.2762066	0.00	0.25
[7,]	0.008051570	0.17346582	0.173465818	0.3031364	0.00	0.25
[8,]	0.007336291	0.19037852	0.190378521	0.2516692	0.00	0.50
[9,]	0.022403972	0.19037852	0.190378521	0.5813885	0.00	0.00
[10,]	0.082410317	0.20894019	0.190378521	0.2516692	0.00	0.50
[11,]	0.047158213	0.19037852	0.190378521	0.2293116	0.00	0.50
[12,]	0.008836588	0.22931160	0.208940192	0.6380732	0.00	0.00
[13,]	0.007336291	0.19037852	0.007336291	0.7002845	0.75	0.00
[14,]	0.008836588	0.22931160	0.173465818	0.6380732	0.00	0.25
[15,]	0.007336291	0.15805559	0.173465818	0.2516692	0.00	0.25
[16,]	0.009698145	0.14401437	0.190378521	0.2293116	0.00	0.25
[17,]	0.007336291	0.20894019	0.190378521	0.3031364	0.00	0.25
[18,]	0.008051570	0.17346582	0.190378521	0.2516692	0.00	0.25
[19,]	0.047158213	0.22931160	0.208940192	0.7002845	0.25	0.00
[20,]	0.008051570	0.19037852	0.208940192	0.3031364	0.00	0.25
[21,]	0.008836588	0.15805559	0.173465818	0.2516692	0.00	0.25
[22,]	0.008836588	0.15805559	0.208940192	0.3031364	0.00	0.00
[23,]	0.007336291	0.17346582	0.229311603	0.2762066	0.00	0.50
[24,]	0.009698145	0.14401437	0.056802235	0.7002845	0.00	0.00
[25,]	0.062340383	0.20894019	0.173465818	0.3031364	0.00	0.25
[26,]	0.008836588	0.19037852	0.208940192	0.2516692	0.00	0.25
[27,]	0.008836588	0.20894019	0.016947783	0.5297395	0.00	0.00
[28,]	0.007336291	0.15805559	0.062340383	0.2762066	0.00	0.25
[29,]	0.008836588	0.17346582	0.158055593	0.2089402	0.00	0.50
[30,]	0.010643702	0.15805559	0.173465818	0.2293116	0.00	0.25
[31,]	0.047158213	0.19037852	0.190378521	0.2762066	0.00	0.00
[32,]	0.007336291	0.15805559	0.190378521	0.2516692	0.00	0.50
[33,]	0.008051570	0.15805559	0.075089212	0.2293116	0.00	0.50
[34,]	0.008051570	0.19037852	0.190378521	0.2293116	0.00	0.50
[35,]	0.016947783	0.17346582	0.208940192	0.2516692	0.00	0.25
[36,]	0.007336291	0.15805559	0.208940192	0.3326919	0.00	0.00
[37,]	0.009698145	0.25166920	0.131220530	0.5813885	0.00	0.25
[38,]	0.009698145	0.25166920	0.190378521	0.5297395	0.00	0.00
[39,]	0.032504327	0.19037852	0.190378521	0.2293116	0.00	0.25
[40,]	0.007336291	0.19037852	0.158055593	0.6380732	0.00	0.00
[41,]	0.090445221	0.22931160	0.229311603	0.2516692	0.00	0.25
[42,]	0.007336291	0.15805559	0.208940192	0.2516692	0.00	0.25
[43,]	0.032504327	0.20894019	0.229311603	0.2762066	0.00	0.25
[44,]	0.010643702	0.19037852	0.190378521	0.2762066	0.00	0.25
[45,]	0.016947783	0.20894019	0.158055593	0.2516692	0.00	0.25
[46,]	0.008836588	0.15805559	0.082410317	0.2516692	0.00	0.50

Imagen 3 Validación Cruzada. Fuente: R studio

Con la validación cruzada se estima los errores que saldrían cuando aplicamos al modelo los diferentes Lambdas.

En esta tabla se aprecian los diferentes lambdas que se pueden aplicar al modelo, a la hora de escoger que lambda es el más apropiado, es decir, el lambda que aporta menos error. Se debe tener en cuenta el principio de Parsimonia o Navaja de Ockhan según el cual «en igualdad de condiciones, la explicación más sencilla suele ser la más probable».

En el desarrollo de este apartado de acuerdo con el principio de Parsimonia se coge la mediana como Lambda para la ejecución el modelo de regresión penalizado.

12.3.3.- Validación modelo regresión penalización Elastic Net

```
Call:
lm(formula = ipv ~ ns(tiempo, 3) + log_viviendas_f + m_2 + parados_f +
    pib + euribor, data = datos)

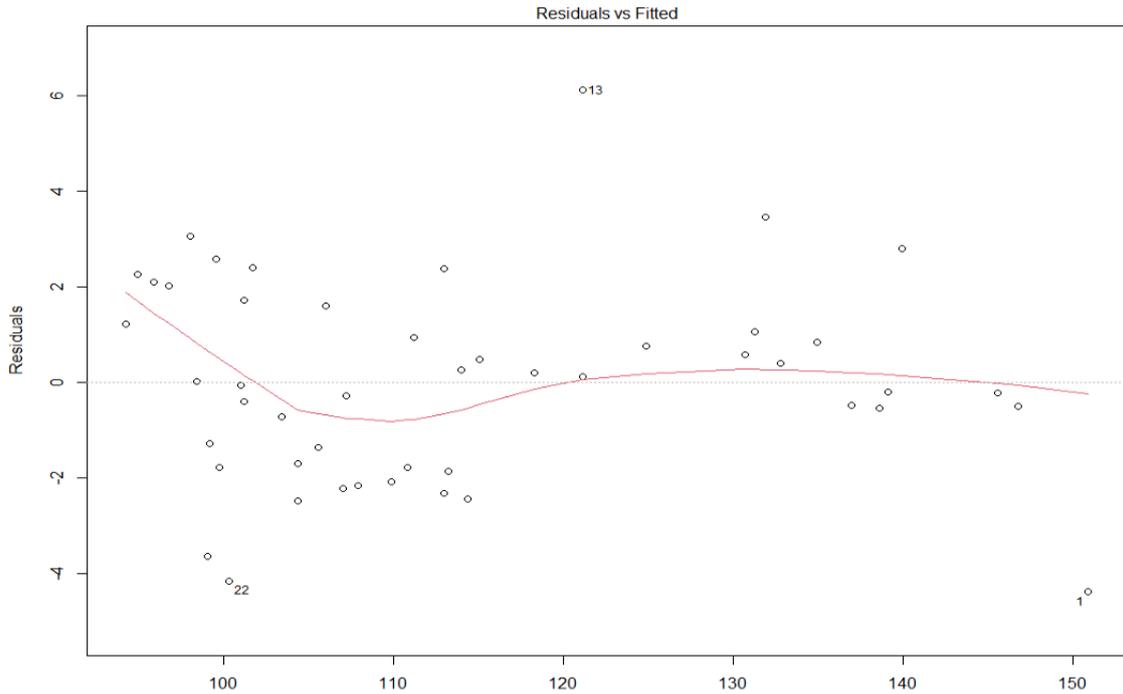
Residuals:
    Min       1Q   Median       3Q      Max
-4.3917 -1.7330 -0.0259  1.3036  6.1225

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.096e+02  2.064e+01  10.156 1.65e-12 ***
ns(tiempo, 3)1 -4.400e+01  5.101e+00  -8.625 1.41e-10 ***
ns(tiempo, 3)2 -5.871e+01  1.165e+01  -5.040 1.10e-05 ***
ns(tiempo, 3)3 -4.631e+01  4.929e+00  -9.396 1.45e-11 ***
log_viviendas_f -1.237e+01  6.127e+00  -2.019  0.0504 .
m_2          -3.614e-02  2.179e-02  -1.659  0.1052
parados_f    -6.753e+00  8.986e-01  -7.516 4.24e-09 ***
pib           7.603e-05  4.607e-05   1.650  0.1069
euribor      -3.275e+02  6.468e+01  -5.063 1.03e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.313 on 39 degrees of freedom
Multiple R-squared:  0.9828,    Adjusted R-squared:  0.9793
F-statistic: 279.3 on 8 and 39 DF,  p-value: < 2.2e-16
```

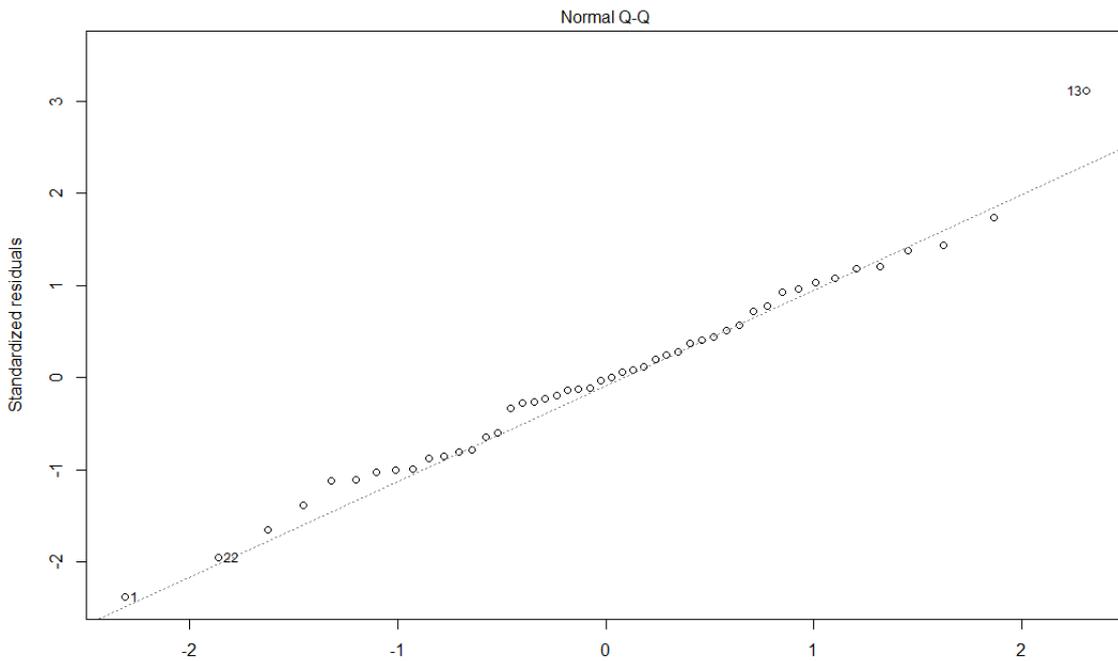
Imagen 4 Resultado Modelo Regresión Penalizado. Fuente R Studio.

En el resumen del análisis de R Studio se observa, que la variable independiente, tiempo, parados y Euribor son muy significativas, su P-valor es muy próximo a cero, mientras que el PIB y los metros cuadrados no serían significativos. En este análisis se puede apreciar una R cuadrada ajustada del 97,93%.



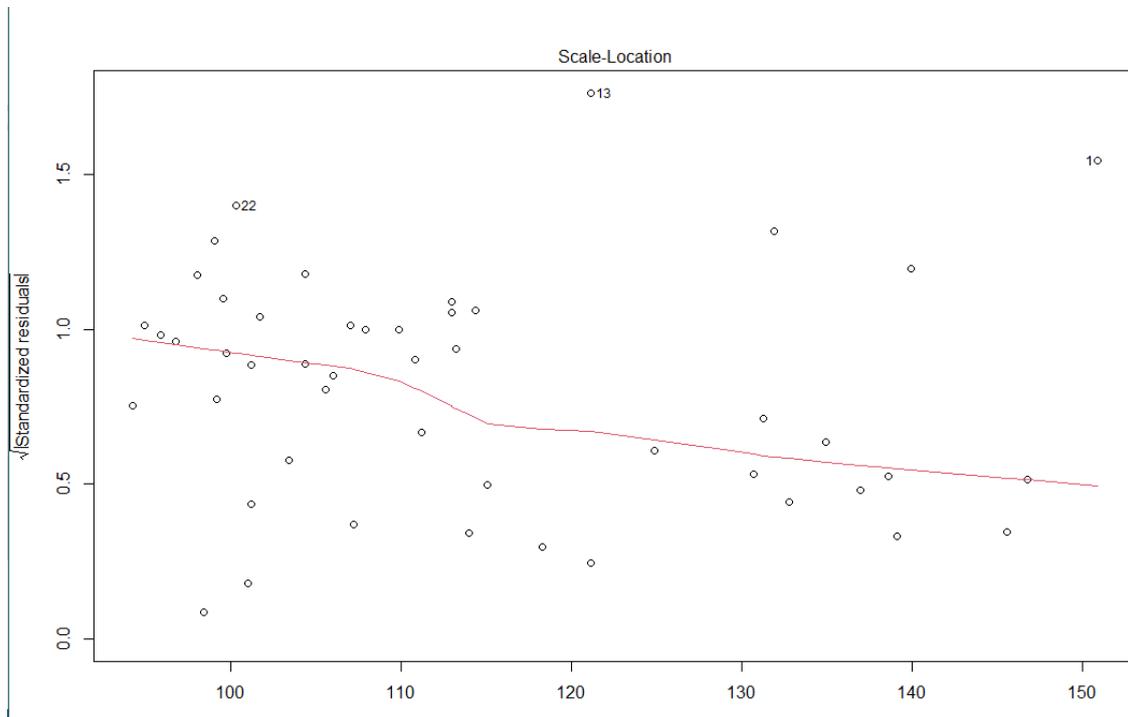
Gráfica 122 Residuos VS IPV. Fuente: R Studio

En el análisis de la gráfica 122, se observa que los residuos se distribuyen de forma normal, no se detectan problemas de heterocedasticidad.



Gráfica 123 Probabilístico normal de los residuos. Fuente: R Studio

Analizando el gráfico 123, se verifica que los residuos se distribuyen de forma normal y una gran mayoría de estos se posan sobre la recta y el resto se encuentra muy próximos a esta, mirando el presente gráfico se podría llegar a afirmar que los residuos son normales.



Gráfica 124 Residuos número de filas. Fuente: R Studio

En el análisis del gráfico 124, se observa forma sinusoidal, evidencia de que existe un problema de autocorrelación.

Una vez analizado el modelo de regresión penalizado y habiendo evidencias de autocorrelación, se concluye que el modelo no es válido y se propone realizar un nuevo modelo ARIMAX.

13.- Nuevo modelo ARIMAX.

Los modelos ARIMAX ¹¹son una extensión de las series temporales a las que se les va a añadir variables explicativas X.

El nombre ARIMAX nace de la combinación de un modelo ARIMA con variables exógenas es decir variables explicativas.

¹¹ Apuntes Econometría UPV 2021

13.1.-Concepto ARIMAX

Los modelos ARIMAX pueden trabajarse como una serie temporal o como un modelo de regresión.

Resulta coherente pensar que si las variables que se están empleando durante un modelo de regresión tienen un comportamiento en el tiempo (Estructura temporal) se puedan trabajar como una serie temporal.

Del mismo modo, si los residuos de ese modelo tienen un comportamiento en el tiempo (Autocorrelación) también deberían poder tenerse en cuenta en la modelización como si de una serie temporal se tratase.

La estructura de un modelo ARIMAX tiene la siguiente composición:

$$\underbrace{\varphi(B)Y_t}_{\text{Polinomio de la Y}} = \underbrace{\gamma(B)X_t}_{\text{Polinomio de la X}} + \underbrace{\theta(B)\varepsilon_t}_{\text{Polinomio de los Residuos}}$$

Imagen 5 Apuntes Econometría. Fuente: UPV EPSA

$$\underbrace{Y_t - \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} - \dots - \varphi_k Y_{t-k}}_{\substack{\text{AUTOREGRESIVO} \\ \mathbf{AR}}} = \underbrace{\beta_0 + \beta_1 X_t - \beta_2 X_{t-1} - \beta_3 X_{t-2} + \dots}_{\text{Media móvil}} + \underbrace{\theta_1 \text{Res}_{t-1} + \theta_2 \text{Res}_{t-2} + U}_{\mathbf{MA}}$$

Imagen 6 Apuntes Econometría. Fuente: UPV EPSA

Como se puede observar en la matriz donde aparece la variable Y, es un modelo de serie temporal autorregresivo, la matriz donde aparece la X son las variables del modelo de regresión, y la matriz del error es una serie de tiempo de media móvil. Para la resolución del presente TFG, se decide pasar con signo negativo el modelo de regresión al otro lado de la igualdad para conseguir un modelo ARIMAX con regresores.

13.2.- Modelo ARIMAX con regresores.

$$\phi(B)(Y_t - (\beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt})) = \theta(B)\varepsilon_t$$

Imagen 7 ARIMAX con Regresores. Fuente: UPV EPSA

13.2.1.-Validación modelo ARIMAX con los regresores: Viviendas, M², Parados, PIB y Euribor

```
Call:
auto.arima(y = datos$ipv, max.p = 10, max.q = 10, stepwise = FALSE, approximation = FALSE,
xreg = X_reg, x = structure(list(x = structure(c(146.519, 146.262, 145.333,
142.711, 138.894, 138.074, 136.444, 135.781, 133.17, 135.343, 132.307, 131.259,
127.294, 125.625, 121.236, 118.46, 111.9, 107.77, 104.218, 101.888, 97.954,
96.157, 97.859, 95.391, 95.453, 98.408, 98.762, 98.015, 97.208, 101.074,
100.8, 100.918, 102.167, 102.913, 102.7, 102.668, 104.081, 104.811, 105.721,
106.925, 107.643, 109.036, 110.662, 111.364, 112.128, 114.209, 115.325,
115.562), .Tsp = c(1, 48, 1), class = "ts")), class = "data.frame", row.names = c(NA,
-48L)))

Coefficients:
      ar1      ar2      ar3      ar4      ar5 intercept ns(tiempo, 4)1 ns(tiempo, 4)2 ns(tiempo, 4)3 ns(tiempo, 4)4
0.4789 0.3095 0.2524 -0.0118 -0.5085 147.6720 -51.4693 -35.1380 -41.8116 -28.6287
s.e. 0.1259 0.1494 0.1422 0.1420 0.1194 8.4781 4.6275 2.5122 6.4024 2.5271
log_viviendas_f m_2 parados_f pib euribor
1.7700 -0.0113 -1.4010 0 -6.5266
s.e. 2.5096 0.0060 0.5874 0 26.3339

sigma^2 estimated as 1.197: log likelihood = -64.9, aic = 161.79
```

Imagen 8 ARIMAX. Fuente R Studio

En este primer modelo ARIMAX se ha decidido introducir como variables independientes las variables viviendas, metros cuadrados, número de parados, PIB y EURIBOR obteniéndose los resultados de la imagen anterior que se van a analizar.

Se puede apreciar que el programa facilita una solución de quinto grado de Regresores, esto quiere decir que el modelo solución es ARIMAX (5, 0, 0)₄. Donde se tiene un modelo Regresivo de quinto nivel, no hay diferenciaciones de orden no estacionaria y un modelo de Medias Móviles de cero, con una estacionalidad trimestral.

Analizando cada variable X que se han utilizado en el modelo ARIMAX aparecen los siguientes datos que evidencian que existen variables no significativas.

13.2.2.-Análisis de las variables.

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ar1	4.7892e-01	1.2589e-01	3.8043	0.0001422	***
ar2	3.0954e-01	1.4936e-01	2.0724	0.0382292	*
ar3	2.5243e-01	1.4224e-01	1.7747	0.0759456	.
ar4	-1.1764e-02	1.4197e-01	-0.0829	0.9339602	
ar5	-5.0846e-01	1.1943e-01	-4.2575	2.067e-05	***
intercept	1.4767e+02	8.4781e+00	17.4180	< 2.2e-16	***
ns(tiempo, 4)1	-5.1469e+01	4.6275e+00	-11.1225	< 2.2e-16	***
ns(tiempo, 4)2	-3.5138e+01	2.5122e+00	-13.9868	< 2.2e-16	***
ns(tiempo, 4)3	-4.1812e+01	6.4024e+00	-6.5306	6.550e-11	***
ns(tiempo, 4)4	-2.8629e+01	2.5271e+00	-11.3286	< 2.2e-16	***
log_viviendas_f	1.7700e+00	2.5096e+00	0.7053	0.4806323	
m_2	-1.1266e-02	6.0260e-03	-1.8696	0.0615381	.
parados_f	-1.4010e+00	5.8740e-01	-2.3851	0.0170743	*
pib	2.5133e-05	1.6184e-05	1.5529	0.1204433	
euribor	-6.5266e+00	2.6334e+01	-0.2478	0.8042573	

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

Imagen 9 P-Valor Modelo ARIMAX. Fuente: R Studio

PRUEBAS DE HIPÓTESIS

•Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{cale} = \frac{\beta_i}{S_{\beta_i}}$$

VIVIENDAS P-VALOR= 0,4806323 dato mayor que Alfa=0,05 por tanto la variable es no significativa.

METROS CUADRADOS P-VALOR= 0,0615381 dato mayor que Alfa=0,05 por tanto la variable es no significativa.

PARADOS P-VALOR= 0.0170743 dato menor que Alfa=0,05 por tanto la variable es significativa.

PIB P-VALOR= 0,1204433 dato mayor que Alfa=0,05 por tanto la variable es no significativa.

EURIBOR P-VALOR = 0.8042573 dato mayor que Alfa=0,05 por tanto la variable es no significativa.

AR5 P-VALOR = $2.067e^{-05}$ dato menor que Alfa=0,05 por tanto la variable es significativa.

AR4 P-VALOR= 0.93396 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AR3 P-VALOR = 0.759456 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AR2 P-VALOR= 0.0382292 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AR1 P-VALOR= 0.0001422 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AIC

EL AIC (criterio de información de Akaike) es una medida de la calidad relativa de un modelo estadístico que nace de una muestra, frente a la población estudiada, proporcionando un medio para la selección del modelo. Es decir, El criterio de Información de Akaike mide la diferencia que hay entre la muestra y la población en valores absolutos por tanto siempre interesará que esa diferencia sea lo más pequeña posible.

Fórmula

$$AIC = 2k - 2 \ln(L)$$

Siendo k el número de parámetros y L la verosimilitud.

El AIC (criterio de información de Akaike) muestra un dato de 161,79

El criterio de Información de Akaike mide la diferencia que hay entre la muestra y la población en valores absolutos por tanto siempre interesará que esa diferencia sea lo más pequeña posible.

13.2.3.-Validación modelo ARIMAX con los regresores: m², parados y PIB

Se plantea un nuevo modelo ARIMAX, esta vez introduciendo las variables X metros cuadrados, parados y PIB.

```
Call:
auto.arima(y = datos$ipv, max.p = 10, max.q = 10, stepwise = FALSE, approximation = FALSE,
  xreg = X_reg2, x = structure(list(x = structure(c(146.519, 146.262, 145.333,
142.711, 138.894, 138.074, 136.444, 135.781, 133.17, 135.343, 132.307, 131.259,
127.294, 125.625, 121.236, 118.46, 111.9, 107.77, 104.218, 101.888, 97.954,
96.157, 97.859, 95.391, 95.453, 98.408, 98.762, 98.015, 97.208, 101.074,
100.8, 100.918, 102.167, 102.913, 102.7, 102.668, 104.081, 104.811, 105.721,
106.925, 107.643, 109.036, 110.662, 111.364, 112.128, 114.209, 115.325,
115.562), .Tsp = c(1, 48, 1), class = "ts")), class = "data.frame", row.names = c(NA,
-48L)))

Coefficients:
      ar1      ar2      ar3      ar4      ar5  intercept ns(tiempo, 4)1 ns(tiempo, 4)2 ns(tiempo, 4)3 ns(tiempo, 4)4
s.e.  0.4965  0.2905  0.2217  0.0063 -0.5080  150.1173      -52.7215      -34.8169      -43.3599      -28.1170
      m_2  parados_f  pib
s.e. -0.0099  -1.3030  0
      m_2  parados_f  pib
s.e.  0.0057  0.5235  0

sigma^2 estimated as 1.143: log likelihood = -65.18, aic = 158.37
```

Imagen 10 ARIMAX 2. Fuente: R Studio.

Nuevamente aparece un modelo ARIMAX (5, 0, 0) ₄. Donde se tiene un modelo Regresivo de quinto nivel, no hay diferenciaciones de orden no estaciona y un modelo de Medias Móviles de igual a cero, con una estacionalidad trimestral.

Analizando las variables X, aparece que todas las variables son significativas.

13.2.4.-Análisis y validación de las variables

```
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1      4.9653e-01  1.2244e-01  4.0552 5.009e-05 ***
ar2      2.9050e-01  1.5083e-01  1.9259  0.05411 .
ar3      2.2175e-01  1.3936e-01  1.5912  0.11157
ar4      6.2614e-03  1.4227e-01  0.0440  0.96490
ar5     -5.0800e-01  1.2040e-01  -4.2192 2.452e-05 ***
intercept  1.5012e+02  6.7822e+00  22.1341 < 2.2e-16 ***
ns(tiempo, 4)1 -5.2721e+01  4.3726e+00 -12.0572 < 2.2e-16 ***
ns(tiempo, 4)2 -3.4817e+01  2.4011e+00 -14.5006 < 2.2e-16 ***
ns(tiempo, 4)3 -4.3360e+01  6.1482e+00 -7.0525 1.757e-12 ***
ns(tiempo, 4)4 -2.8117e+01  2.0057e+00 -14.0183 < 2.2e-16 ***
m_2      -9.9047e-03  5.7349e-03  -1.7271  0.08415 .
parados_f -1.3030e+00  5.2345e-01  -2.4893  0.01280 *
pib      2.3850e-05  1.6014e-05  1.4893  0.13640
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Imagen 11 P-Valor Modelo ARIMAX 2. Fuente: R Studio

PRUEBAS DE HIPÓTESIS

- Cada parámetro

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad t_{calc} = \frac{\beta_i}{S_{\beta_i}}$$

METROS CUADRADOS P-VALOR= 0,08415 dato mayor que Alfa=0,05 por tanto la variable es no significativa.

PARADOS P-VALOR= 0.01280 dato menor que Alfa=0,05 por tanto la variable es significativa.

PIB P-VALOR= 0,13640 dato mayor que Alfa=0,05 por tanto la variable es no significativa.

AR5 P-VALOR = $2.452e^{-05}$ dato menor que Alfa=0,05 por tanto la variable es significativa.

AR4 P-VALOR= 0.96490 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AR3 P-VALOR = 0.11157 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AR2 P-VALOR= 0.05411 de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AR1 P-VALOR= $5.009e^{-05}$ de acuerdo con el Principio de Marginalidad que afirma si existe términos de orden superior relevantes o significativos nunca se pueden eliminar de los de orden inferior, por tanto, esta variable debe aparecer en el modelo.

AIC

El AIC (criterio de información de Akaike) muestra un dato de 158,37

13.2.5.- Selección mejor modelo ARIMAX

Para seleccionar cuál de los dos modelos ARIMAX calculados es el óptimo, debemos utilizar el dato que aparece en R Studio con nomenclatura AIC (Criterio de información de Akaike).

El AIC se utiliza para la selección de modelos de series temporales, y siempre se escogerá aquel modelo que contenga el AIC más pequeño. A diferencia de la herramienta de R2 ajustado que nos da una evidencia de lo bueno o malo que es el modelo, el AIC no nos facilita esa evidencia, pero si ayuda a elegir entre dos modelos.

13.2.6.-Predicción modelo ARIMAX con AIC= 158,37

```
Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
49             116.1570 114.7870 117.5270 114.0617 118.2523
50             116.0107 114.4810 117.5403 113.6713 118.3501
51             115.4901 113.7927 117.1875 112.8941 118.0860
52             114.8529 112.9470 116.7588 111.9381 117.7677
> load("C:/Users/Acer/Downloads/workspace_final.RData")
```

Imagen 12 Predicción para 4 periodos. Fuente: R Studio.

En la siguiente imagen, se observan las predicciones que realiza el programa R Studio para una predicción de 4 periodos, apareciendo la predicción puntual y las predicciones por intervalos de confianza del 80% y del 90%.

Periodo 49:

Predicción Puntual = 116,1570

[114.787< IPV<117.527] 80% confianza

[114.0617 < IPV < 118.2523] 95% confianza

Periodo 50:

Predicción Puntual = 116,0107

[114.4810 < IPV < 117.5403] 80% confianza

[113.6713 < IPV < 118.3501] 95% confianza

Periodo 51:

Predicción Puntual = 115.4901

[113.7927 < IPV < 117.1875] 80% confianza

[112.8941 < IPV < 118.086] 95% confianza

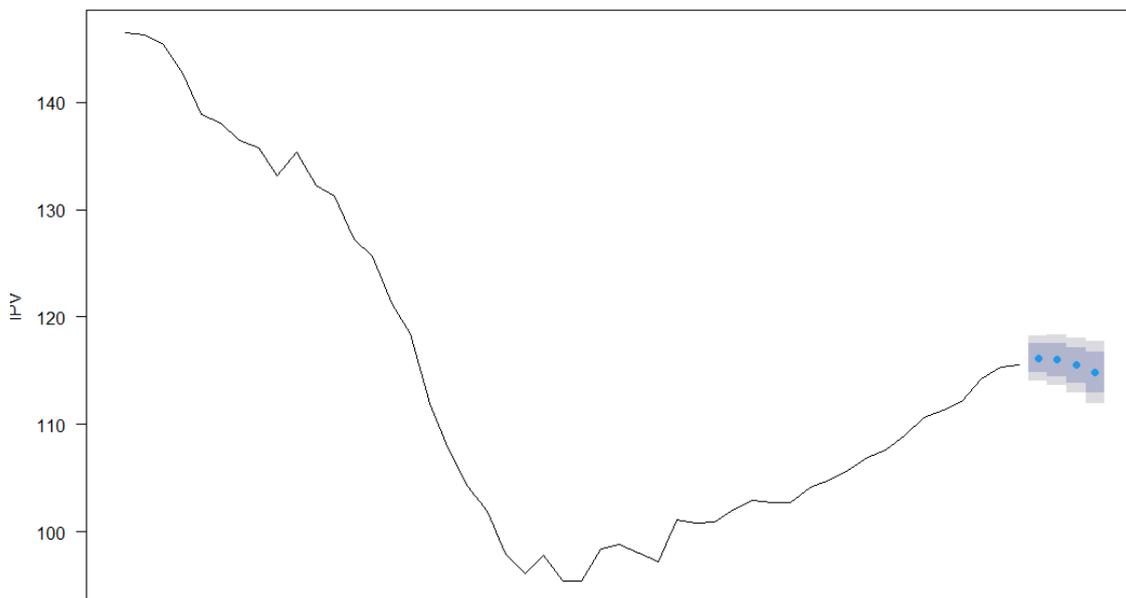
Periodo 52:

Predicción Puntual = 114.8529

[112.9470 < IPV < 116.7588] 80% confianza

[111.9381 < IPV < 117.7677] 95% confianza

Forecasts from Regression with ARIMA(5,0,0) errors



Gráfica 125 Representación predicción para 4 periodos. Fuente: R Studio

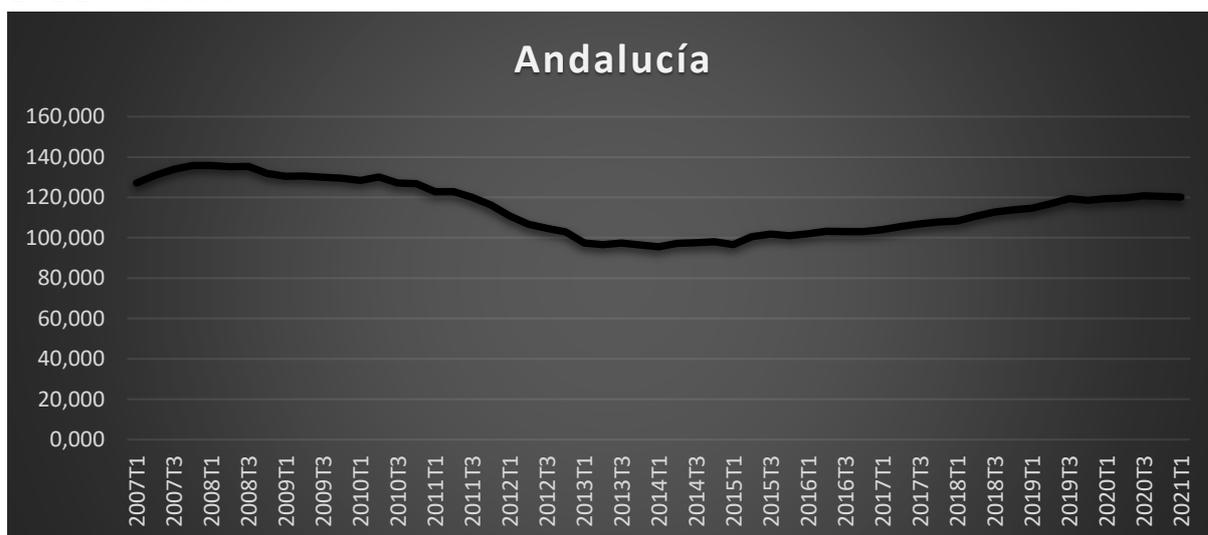
De acuerdo con los datos facilitados por el programa R Studio la previsión del IPV para los próximos 4 periodos, tiende un pequeño incremento durante el primer periodo y una disminución en los 3 periodos restantes.

14.- Objetivo secundario: Comparativa de la evolución IPV entre comunidades.

14.1-Estudio individual de las comunidades.

En este apartado se va a tratar de hacer un mini análisis de cada una de las comunidades autónomas, en su evolución del IPV.

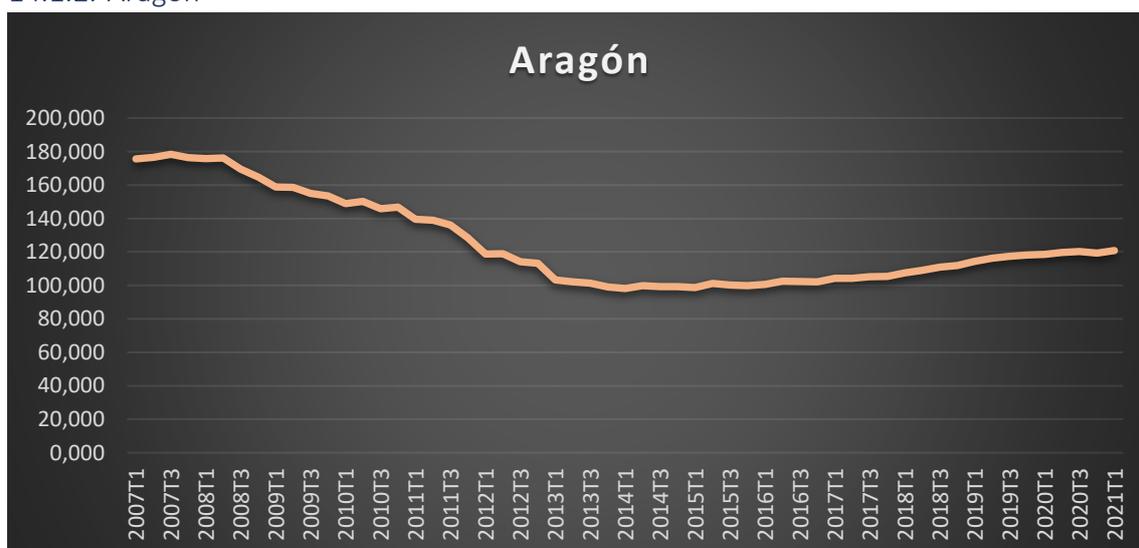
14.1.1.-Andalucía



Gráfica 126 IPV Andalucía

En la gráfica 126, se observa como el IPV en Andalucía se mantiene constante a lo largo de los años, se observa una bajada desde el primer trimestre del 2007 hasta el del 2013, a continuación, se observa un pequeño incremento que llega hasta día de hoy, lo que quiere decir que el IPV está aumentando.

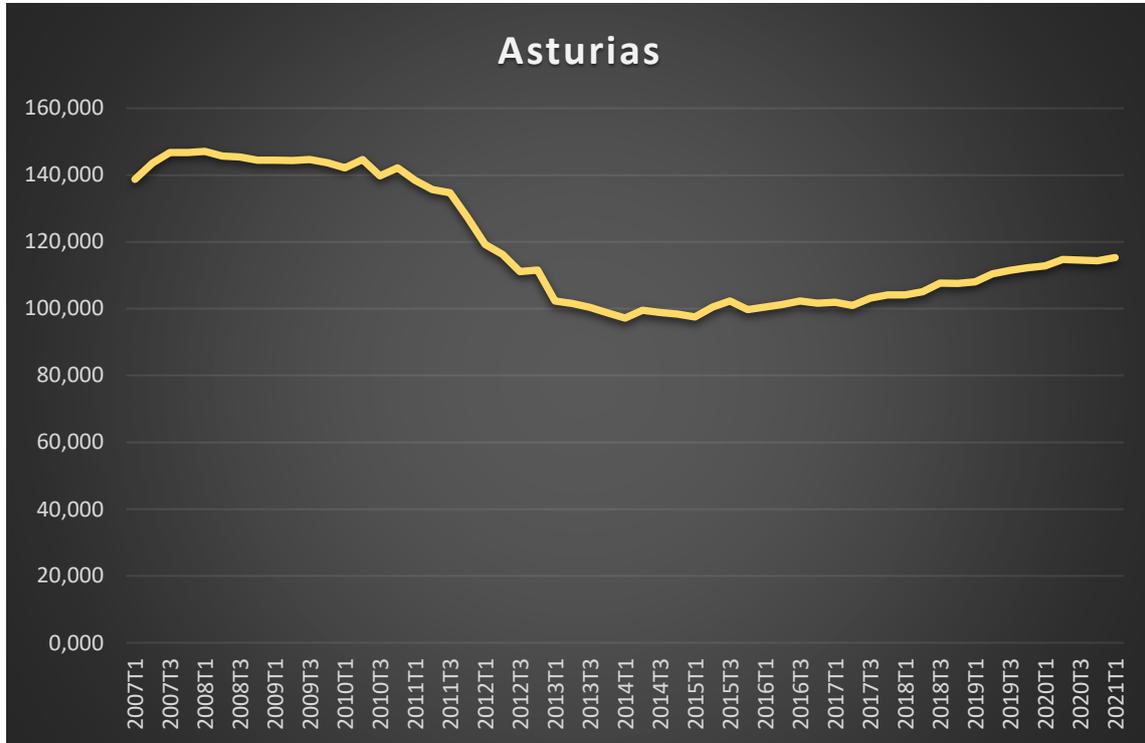
14.1.2.-Aragón



Gráfica 127 IPV Aragón

En la gráfica 127, se observa como el IPV de Aragón tiene una disminución de los valores desde el primer trimestre del 2007 hasta el del 2016, a continuación, se observa un pequeño incremento que llega hasta día de hoy, lo que quiere decir que el IPV estos últimos años está mejorando su valor.

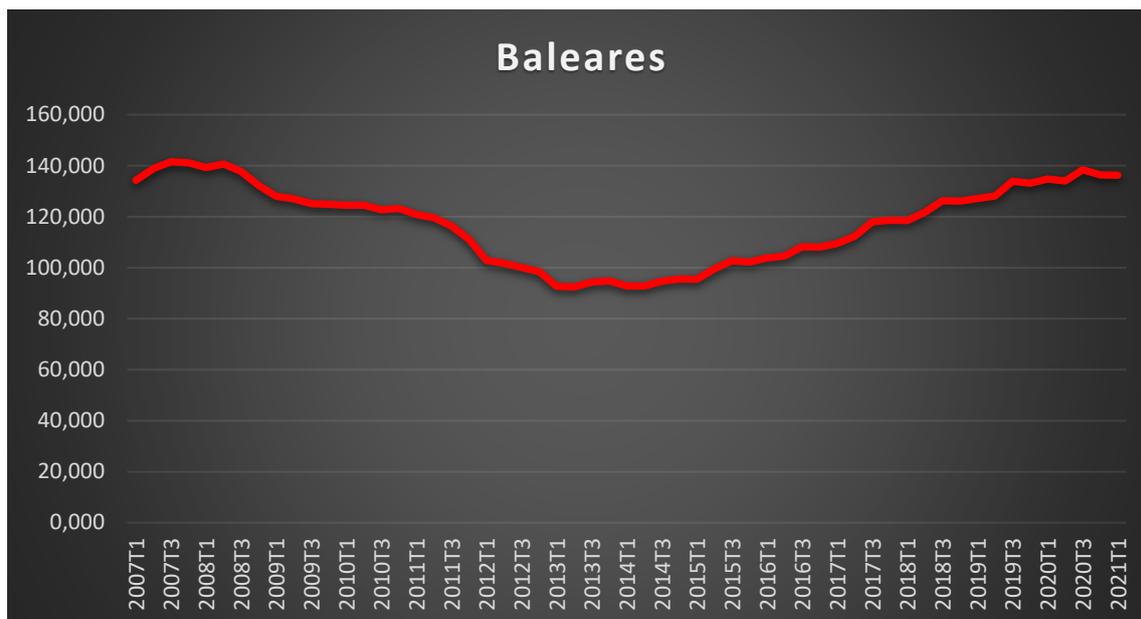
14.1.3.-Asturias



Gráfica 128 IPV Asturias

En la gráfica 128, se observa como el IPV de Asturias, donde se aprecia un descenso desde el primer trimestre del 2010 hasta el del 2016, y a partir del 2016 se intuye un incremento que llega hasta día de hoy, lo que quiere decir que el IPV aumenta con el paso del tiempo.

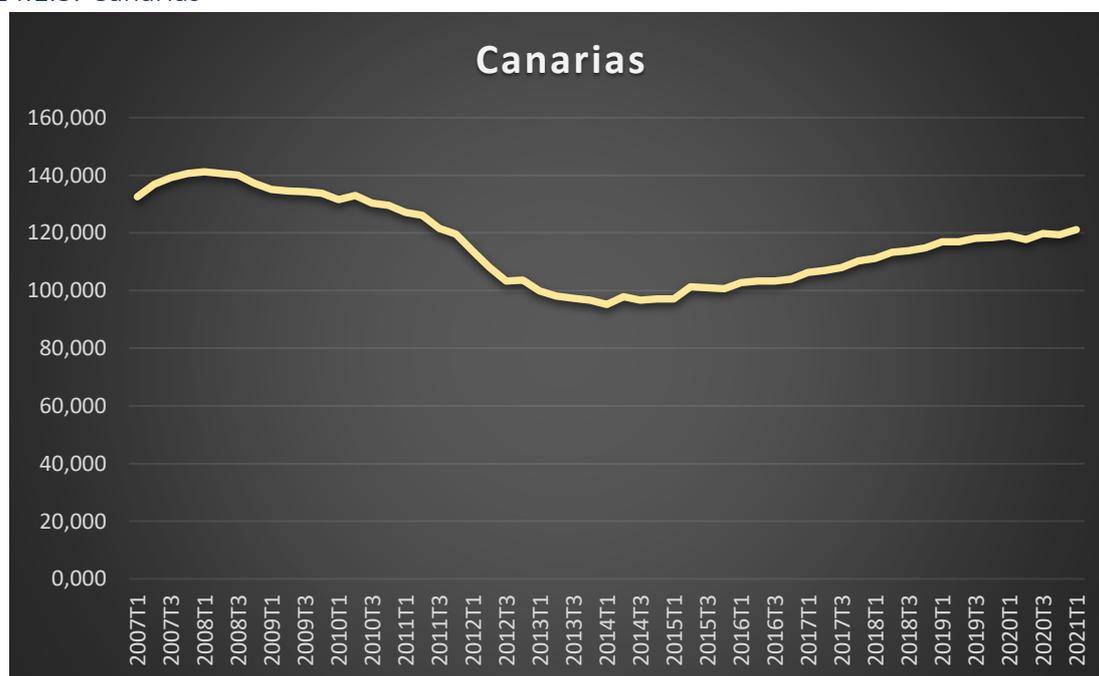
14.1.4.-Baleares



Gráfica 129 IPV Baleares

En la gráfica 129, se observa como el IPV Baleares, se aprecia una bajada desde el primer trimestre del 2007 hasta el cuarto trimestre del 2013, posteriormente se observa un incremento que llega hasta día de hoy, llegando a situarse en los valores conseguidos en el T1 del 2007.

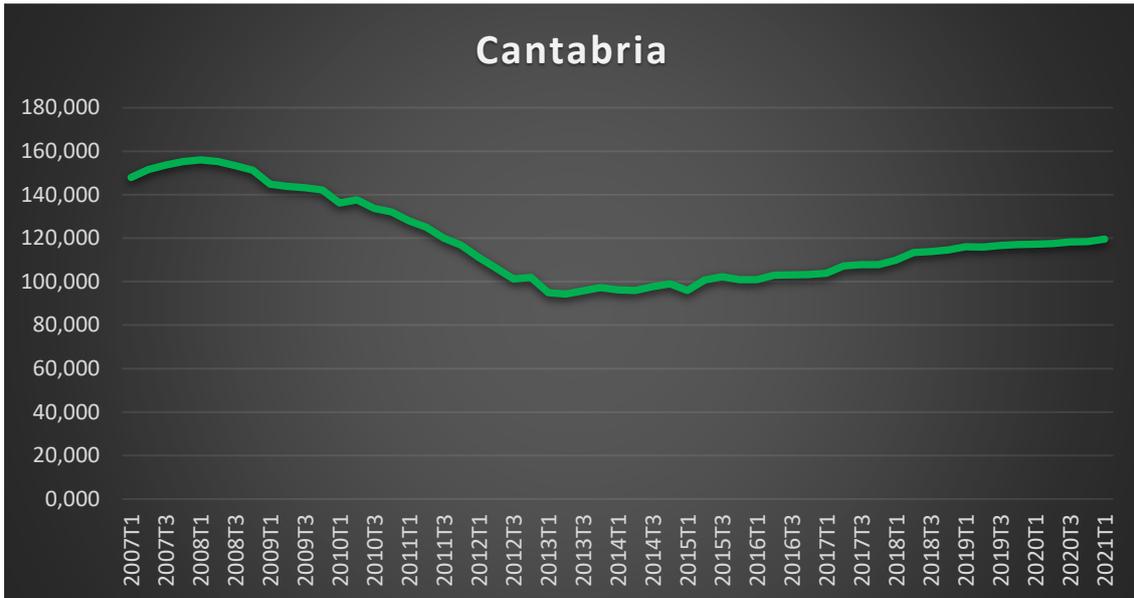
14.1.5.-Canarias



Gráfica 130 IPV Canarias

En la gráfica 130, se observa como el IPV Canarias, tiene una bajada desde el primer trimestre del 2007 hasta el cuarto trimestre del 2013, a continuación, se observa un incremento que llega hasta el 1T de 2021, lo que significa que el IPV aumentará a corto plazo.

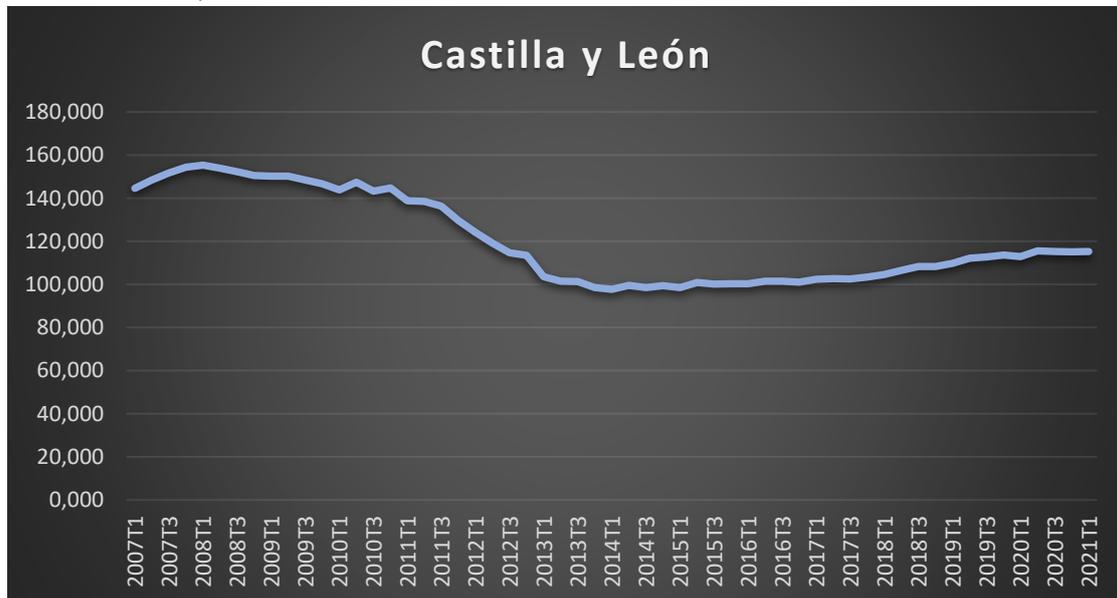
14.1.6.-Cantabria



Gráfica 131 IPV Cantabria

En la gráfica 131, se observa como el IPV de Cantabria, tiene una bajada desde el primer trimestre del 2007 hasta el primer trimestre del 2013, luego a partir de ahí se observa un incremento notable que llega hasta día de hoy, lo que significa que el IPV se está revalorizando.

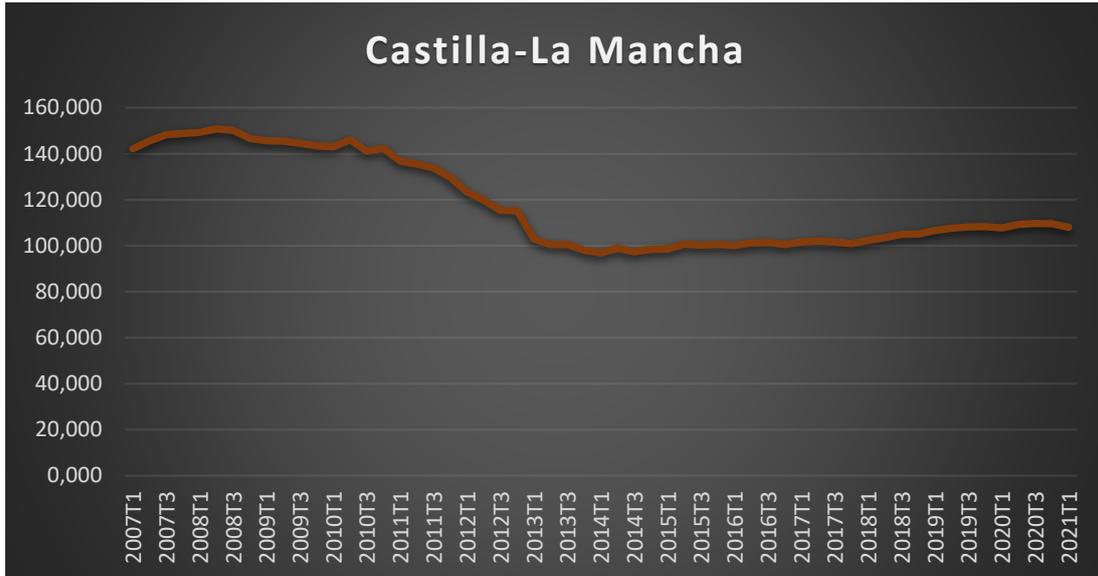
14.1.7.-Castilla y León



Gráfica 132 IPV Castilla y León

En la gráfica 132, se observa como el IPV de Castilla y León, tiene un descenso desde el primer trimestre del 2007 hasta el cuarto trimestre del 2013, a continuación, se observa un incremento escalonado que llega hasta el 1T de 2021, esto significa que el IPV se está revalorizando.

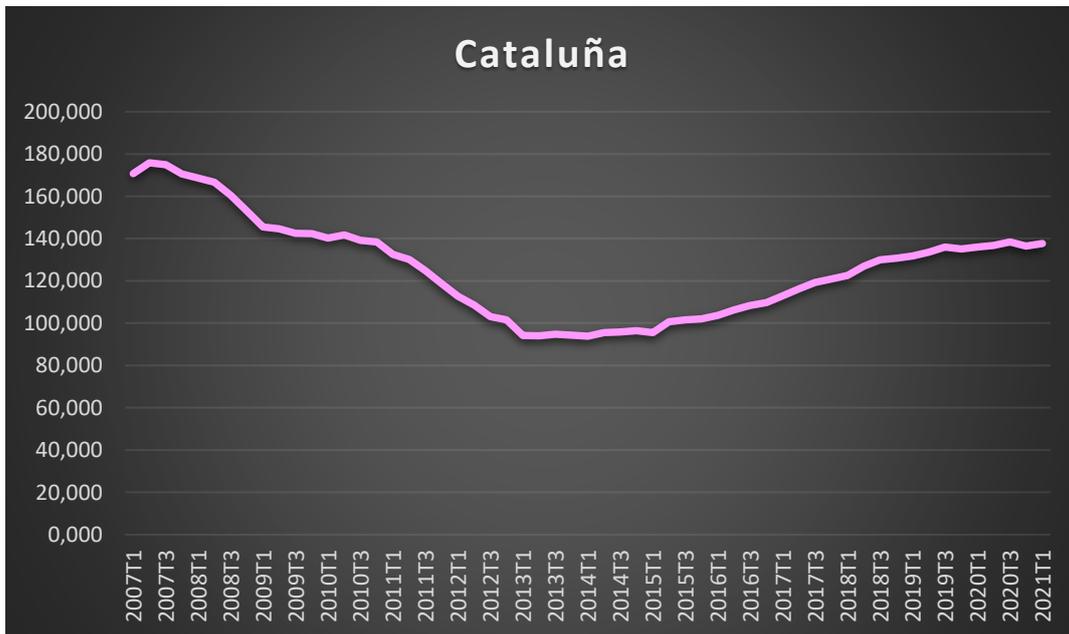
14.1.8.-Castilla -La Mancha



Gráfica 133 IPV Castilla-La Mancha

En la gráfica 133, se observa como el IPV de Castilla la Mancha, tiene una bajada desde el primer trimestre del 2007 hasta el primer trimestre del 2013, seguidamente se observa un pequeño incremento que llega hasta día de hoy, manteniéndose bastante similar en los últimos años.

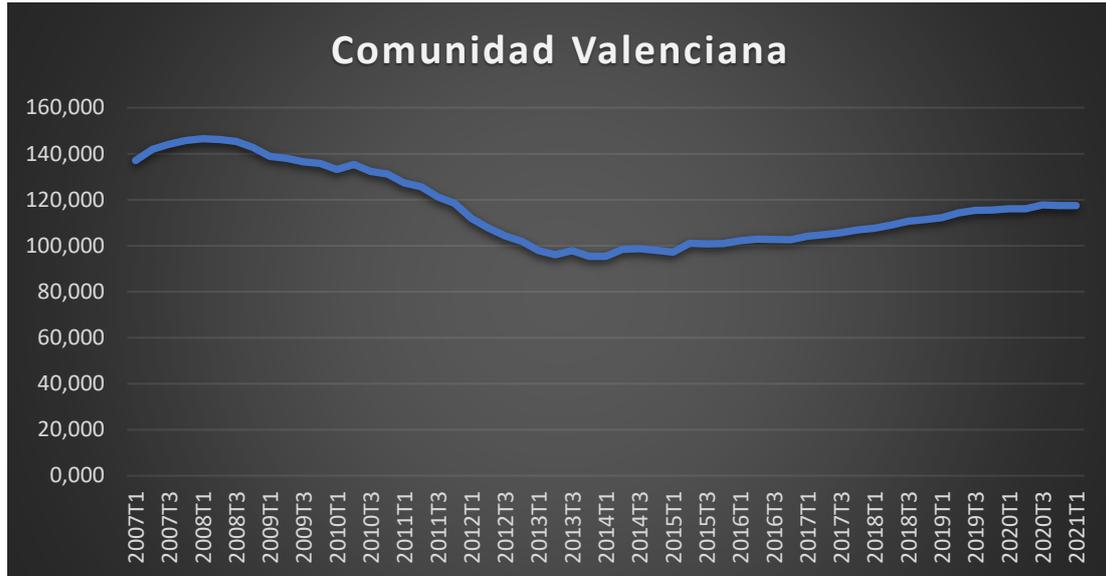
14.1.9.-Cataluña



Gráfica 134 IPV Cataluña

En la gráfica 134, se observa como el IPV de Cataluña tiene forma de V, existe una bajada desde el primer trimestre del 2007 hasta el primer trimestre del 2013, a continuación, se aprecia un IPV constantes hasta el segundo trimestre del 2015 y posteriormente se observa un incremento, que llega hasta día de hoy, lo que puede significar que el IPV se está revalorizando y llegue a sus valores iniciales.

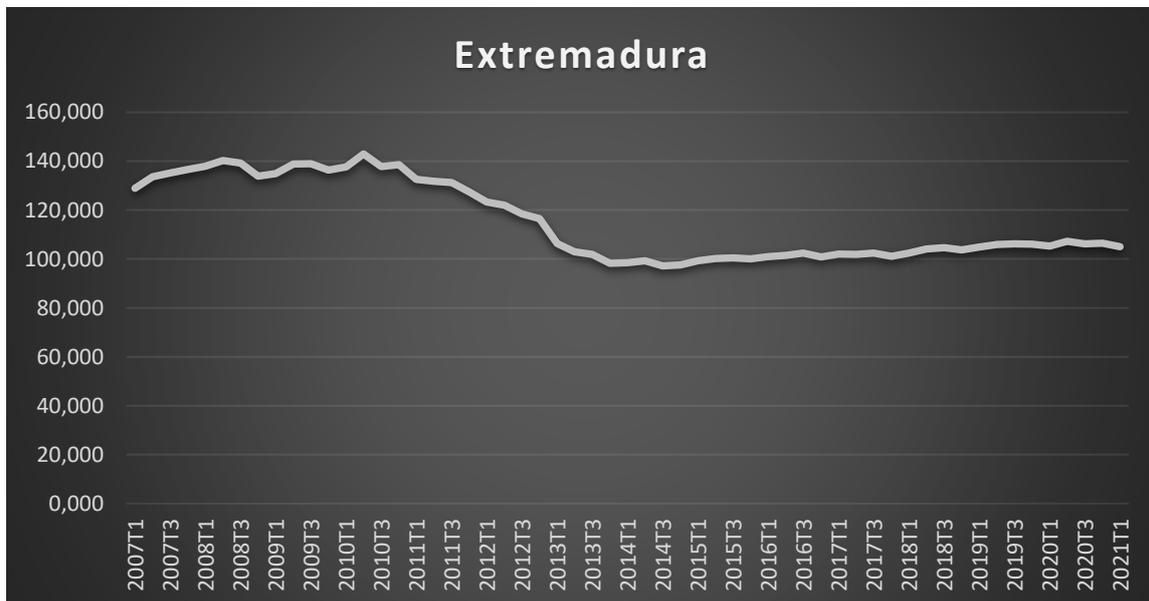
14.1.10.-Comunidad Valenciana



Gráfica 135 IPV Comunidad Valenciana

En la gráfica 135, se observa como el IPV de la Comunidad Valenciana, tiene un descenso desde el primer trimestre del 2007 hasta el primer trimestre del 2013, a continuación, se observa un incremento notable que llega hasta el 1T de 2021, lo que significa que el IPV se está revalorizando y pueda llegar a sus valores iniciales.

14.1.11.-Extremadura

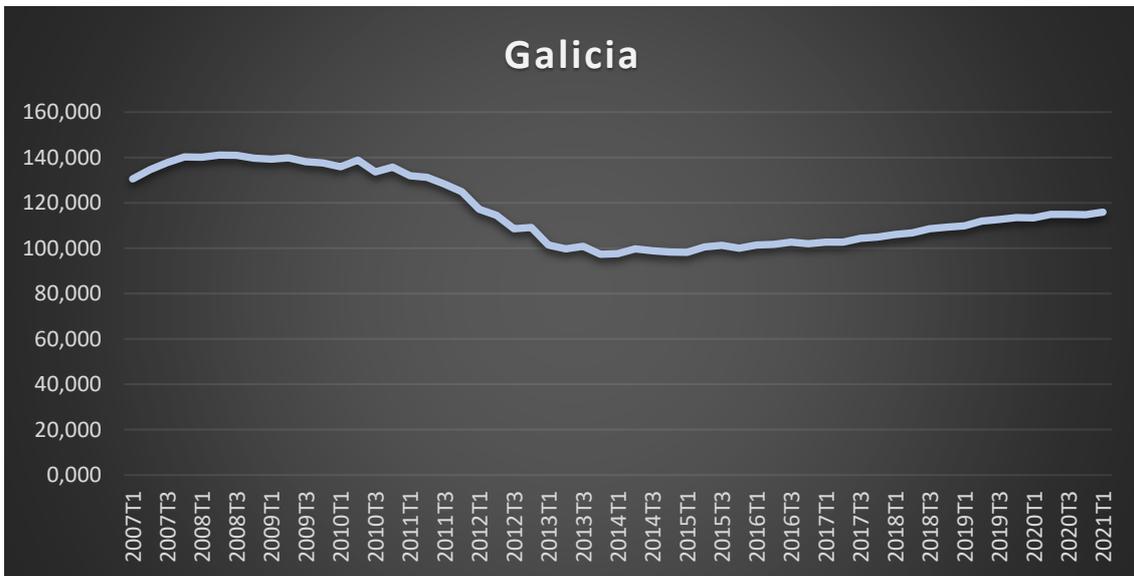


Gráfica 136 IPV Extremadura

En la gráfica 136, se observa como el IPV de Extremadura, tiene un incremento desde el primer trimestre del 2007 hasta el primer trimestre del 2010, luego se observa una bajada desde el primer trimestre de 2010 hasta el tercer trimestre del 2014 y a partir de ahí se observa un IPV

constante, con un ligero incremento a partir del 1T de 2018 hasta el 1T de 2021, pero por el momento el IPV no crece lo suficiente para alcanzar los datos del 2010.

14.1.12.-Galicia



Gráfica 137 IPV Galicia

En la gráfica 137, se observa como el IPV de Galicia, tiene un descenso desde el primer trimestre del 2007 hasta el tercer trimestre del 2014, a continuación, se observa una aumento de los valores del IPV hasta día de hoy, lo que significa que el IPV se está revalorizando y puede llegar a sus valores iniciales.

14.1.13.-Comunidad Madrid

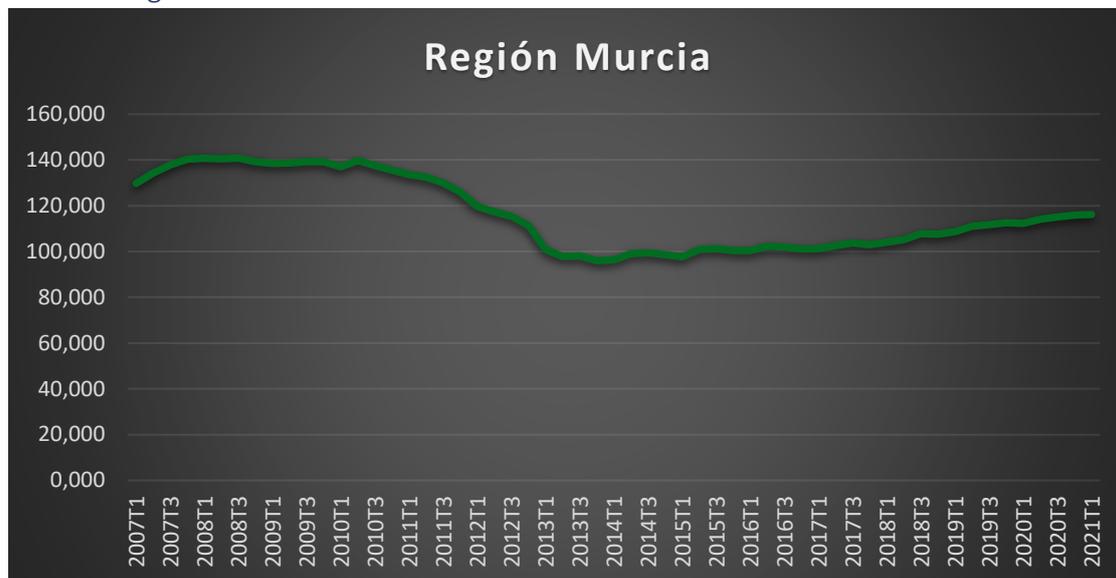


Gráfica 138 IPV Comunidad Madrid

En la gráfica 138, se observa como el IPV de la Comunidad de Madrid tiene forma de V, existe una bajada desde el primer trimestre del 2007 hasta el primer trimestre del 2013, a continuación, el IPV permanece constante hasta el segundo trimestre del 2015 y a partir de ese

momento se observa un incremento, que llega hasta el 1T de 2021, lo que puede significar que el IPV se está revalorizando y podría llegar al valor del año 2007.

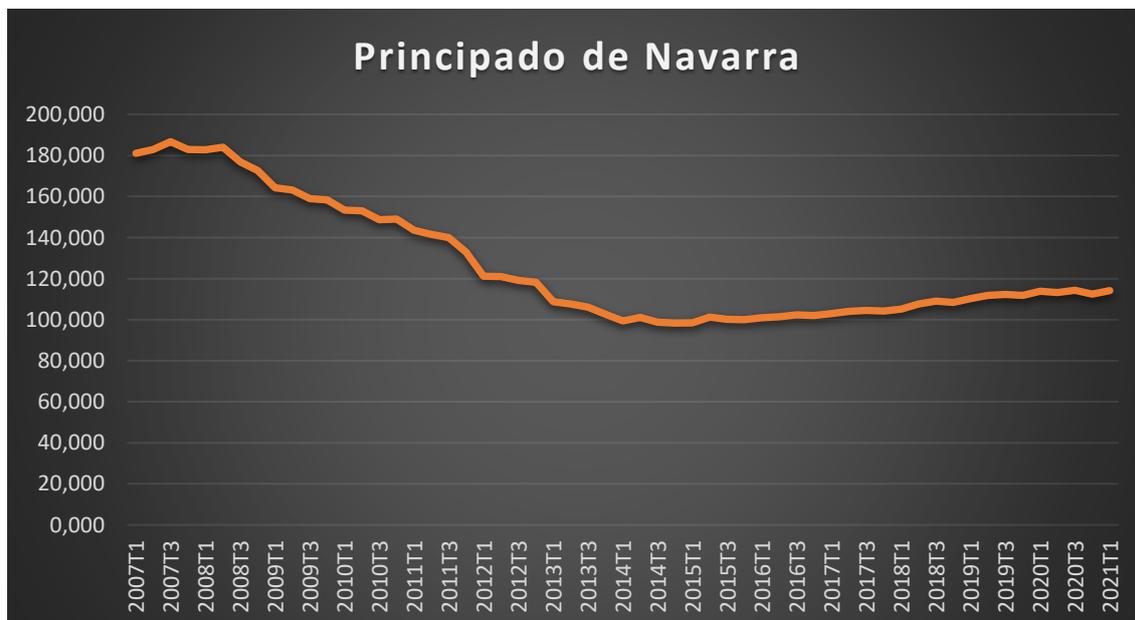
14.1.14.-Región de Murcia



Gráfica 139 IPV Región Murcia

En la gráfica 139, se observa como el IPV de Murcia, se mantiene estable desde el primer trimestre de 2007 hasta el primer trimestre de 2010, a continuación, se observa una disminución desde el primer trimestre del 2010 hasta el tercer trimestre del 2014, finalmente desde el 4T de 2014 hasta el 1T de 2021 se observa un incremento de los valores reflejados, lo que significa que el IPV se está revalorizando y pueda llegar a sus valores iniciales.

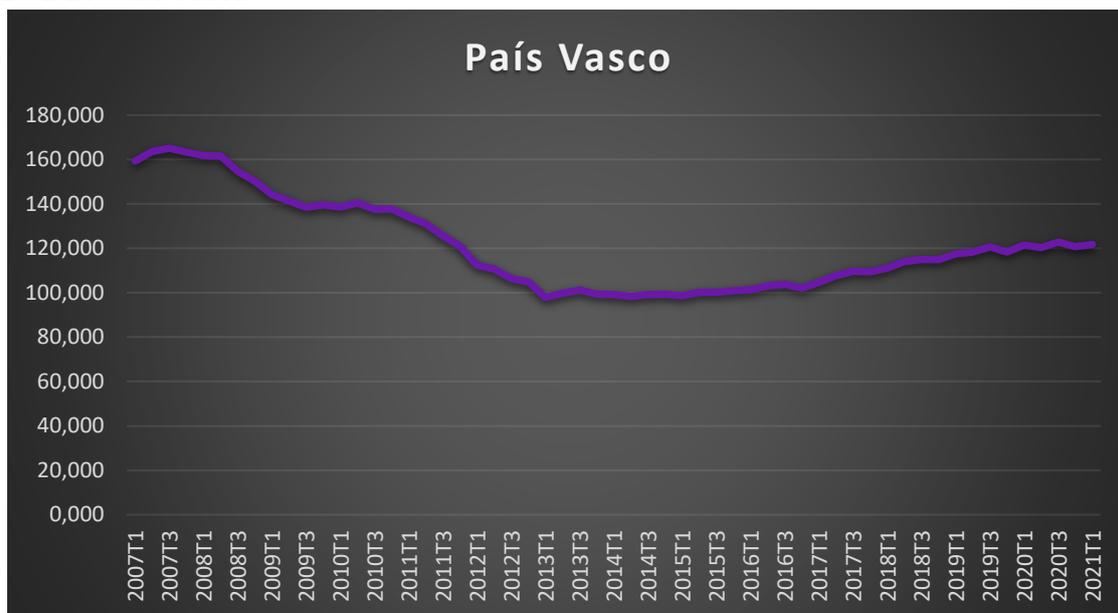
14.1.15.-Navarra



Gráfica 140 IPV Principado Navarra

En la gráfica 140, se observa como el IPV de Navarra, se observa el mayor descenso del IPV desde el primer trimestre del 2007 hasta el tercer trimestre del 2014, a continuación se observa una mejoría de los valores del IPV ya que sufren un ligero incremento hasta el 1T de 2021. Lo que significa que el IPV se está revalorizando ligeramente y no se espera un incremento pronunciado en el corto plazo, hay que destacar que el principado de Navarra ha sufrido el mayor descenso de IPV frente al resto de comunidades autónomas.

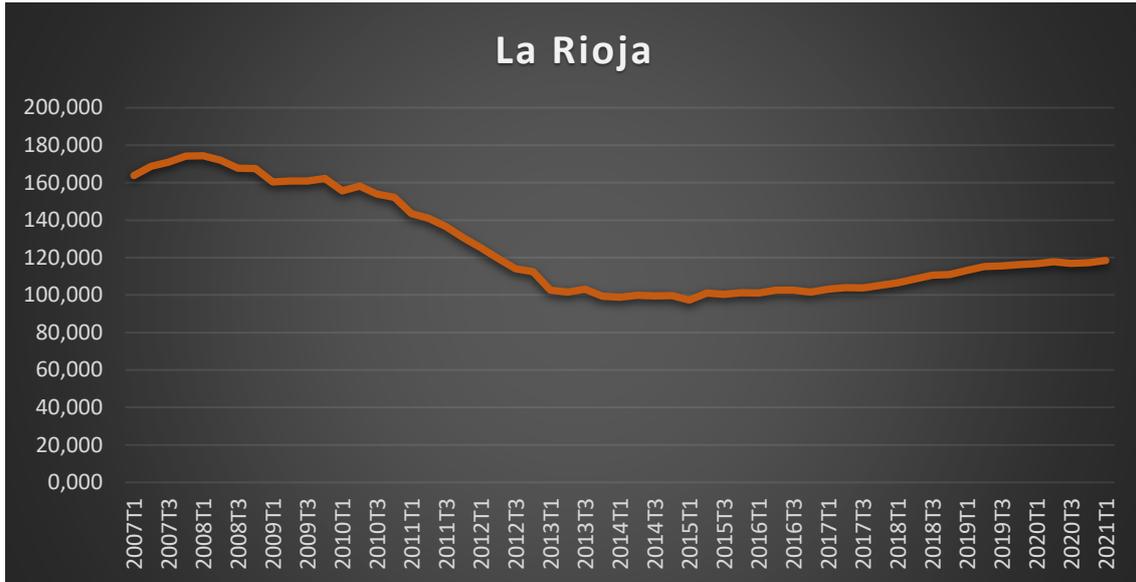
14.1.16.-País Vasco



Gráfica 141 IPV País Vasco

En la gráfica 141, se observa como el IPV del País Vasco, tiene un descenso desde el primer trimestre del 2007 hasta el primer trimestre del 2013, a continuación, se observa un ligero incremento del IPV que llega hasta el 1T de 2021, lo que significa que el IPV se está revalorizando y pueda llegar a sus valores iniciales.

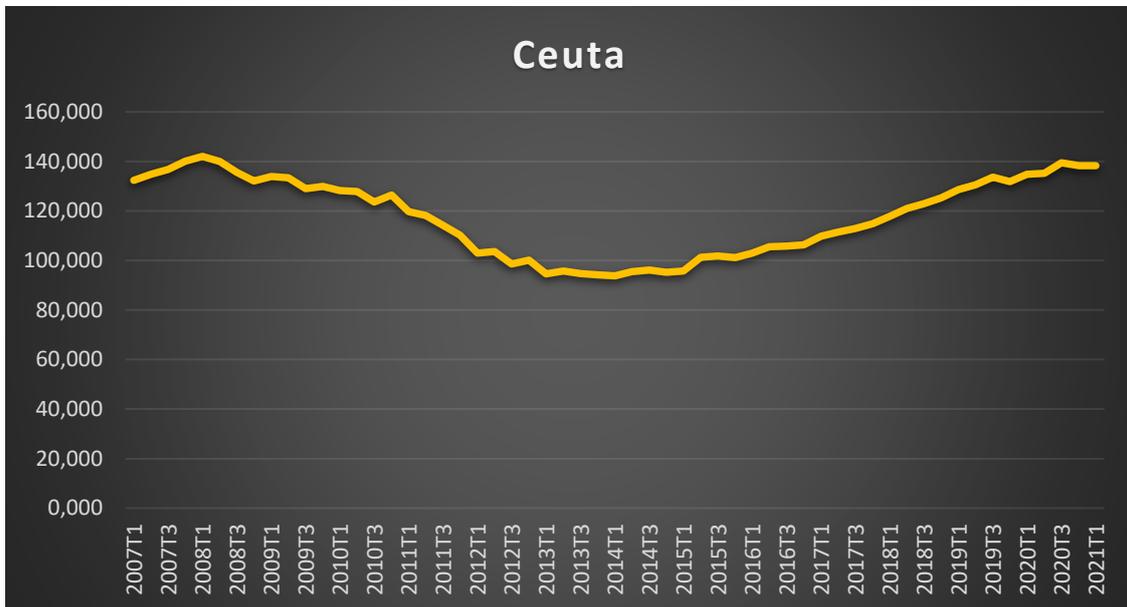
14.1.17-La Rioja



Gráfica 142 IPV La Rioja

En la gráfica 142, se observa como el IPV de La Rioja, se observa un descenso del IPV desde el primer trimestre del 2007 hasta el primer trimestre del 2013, a continuación, el IPV permanece constante hasta el primer trimestre de 2016, a partir del 2T de 2016 se observa un ligero aumento en el IPV, pero no se espera que se llegue al IPV de 2007 en el corto plazo.

14.1.18.-Ceuta

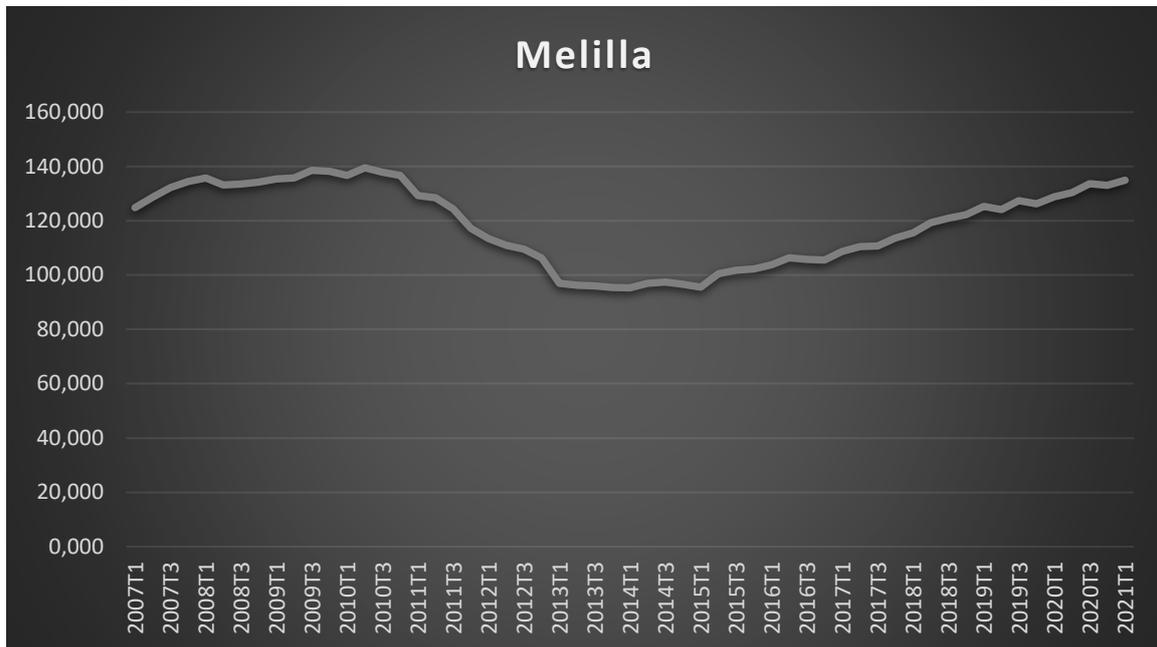


Gráfica 143 IPV Ceuta

En la gráfica 143, se observa como el IPV de Ceuta, tiene forma de V, existe un descenso desde el primer trimestre del 2007 hasta el primer trimestre del 2013, a continuación, se observa como el IPV permanece constante hasta el segundo trimestre del 2015 y a partir de ahí se observa un gran incremento, que llega hasta el 1T de 2021. Se espera que el valor del IPV sobrepase los

datos del IPV de 2007, siendo la única comunidad autónoma que puede alcanzar y sobrepasar sus datos históricos.

14.1.19.-Melilla

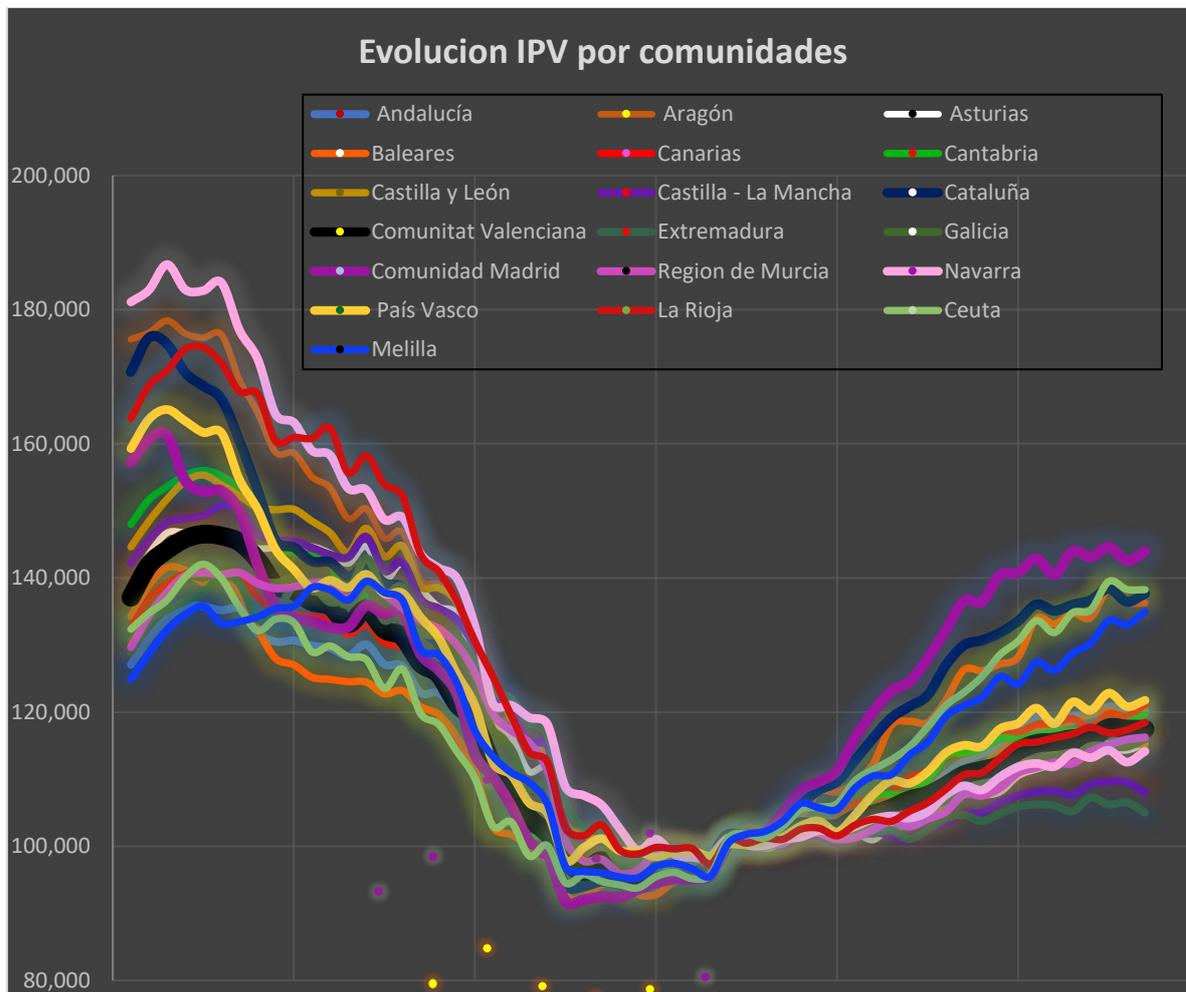


Gráfica 144 IPV Melilla

En la gráfica 144, se observa como el IPV de Melilla, se observa un incremento desde el primer trimestre del 2007 hasta el cuarto trimestre del 2010, a continuación, el IPV sufre un descenso desde el cuarto trimestre de 2010 hasta el cuarto trimestre del 2013 para finalmente desde el 1T de 2014 hasta 1T de 2021 se consiga un incremento similar a los datos del IPV de los años 2007 y 2008.

14.2.-Estudio global de las comunidades.

A continuación, se va a observar la siguiente gráfica y se va a estudiar la evolución del IPV de la C.V respecto a todas las otras comunidades.



Gráfica 145 Evolución IPV por Comunidades Autónomas

En la gráfica 145, se observa la evolución de los índices de precios de las viviendas de las diferentes Comunidades Autónomas desde el 2007 hasta el año 2021. En esta se puede observar representada la Comunidad Valenciana de color negro con puntos amarillos. La evolución del IPV de la C.V es una evolución muy similar al resto de comunidades.

Se podría afirmar que todas las comunidades autónomas de España han sufrido un descenso desde 1t de 2007 hasta 1T de 2013, a continuación, ha habido comunidades autónomas que han sufrido un proceso de estancamiento en su IPV y otras han empezado un incremento ligero en su IPV, finalmente a partir del 1t de 2018 todas las comunidades autónomas han empezado a tener un incremento de su IPV.

Se destaca la Comunidad Foral de Navarra, como aquella comunidad con menor incremento en el IPV frente al resto de comunidades autónomas.

Por el contrario, la Comunidad de Madrid ha conseguido que su IPV sea el más grande que el resto de las comunidades autónomas en el 1T de 2021, si tomamos el IPV en valores absolutos, seguida por Cataluña y Ceuta.

Se recuerda que Ceuta ha presentado el mayor aumento del IPV, partiendo de un IPV inferior al del resto de Comunidades Autónomas.

En relación con la Comunidad Valenciana su IPV se encuentra por debajo de la comunidad de Madrid y Cataluña, lo que sorprende en este análisis ya que la Comunidad Valenciana es una de las más importantes de España, gracias a su situación geográfica y climática, lo que lleva a la siguiente reflexión donde se afirma que en la Comunidad Valenciana no se ha desarrollado al 100% en el ámbito inmobiliario y por tanto tiene más recorrido de crecimiento frente a las principales Comunidades Autónomas líderes en el IPV.

Para finalizar, se observa que todas las comunidades autónomas en estos últimos años el IPV está creciendo progresivamente, lo que indica la gráfica que es positivo para todo el sector de la vivienda.

15.-Conclusiones finales

15.1. Conclusiones sobre las hipótesis planteadas

El presente trabajo parte de 3 hipótesis iniciales, las cuales se comentan a continuación.

Hipótesis 1: El IPV de la Comunidad Valenciana será inferior al de las comunidades de Cataluña y Madrid, pero cercano a estas, debido a que es una comunidad muy importante en España.

Como se ha visto en el apartado 14.2 la Comunidad Valenciana está bastante alejada en cuanto a los valores de las comunidades como lo son Cataluña y Madrid. Tal y como se ha comentado en el apartado correspondiente, este hecho puede ser debido a dos motivos. Se piensa que el IPV en la C.V puede mejorar y está en pendiente creciente, pero basándonos en el modelo ARIMAX, este crecimiento va a terminar de aquí 1 periodo, y los siguientes 3 volverá a declinarse, o puede darse el caso que siga subiendo y que llegue a los valores que tienen hoy en día las comunidades referentes en España como lo es Cataluña y Madrid.

Hipótesis 2: Se prevé que las variables que afectarán de forma más significativa en las variaciones del IPV, son los M² de los que dispone la vivienda, el nivel de salarios de las personas y el número de desempleados.

Como se ha ido observando en el Trabajo Final de Grado, las variables que se quedan en un modelo ARIMAX valido son M², Parados y en sustitución a la que se creía que a priori tenía una buena relación los Salarios, ha entrado en el modelo la variable PIB, esta variable es razonable debido a que según el estudio del arte muchos de los compañeros citaban el PIB como una variable imprescindible para el desarrollo del modelo, ya que está muy relacionada con las Viviendas.

Hipótesis 3: Se prevé la variable Euribor no tenga una relación directa con el IPV, debido a que en los últimos años se ha mantenido en valores cercanos a 0 o incluso negativos y se espera que no tenga una influencia significativa en el IPV.

La última hipótesis, que se planteó en el principio del trabajo, era contemplar si la variable Euribor al estar presentando valores negativos se ha pensado que no iba a ser válida y se esperaba que no tuviera una influencia significativa en el IPV.

Como se ha observado en el trabajo, la variable Euribor no tiene una influencia significativa sobre el IPV, debido a que en el modelo de regresión no se cuenta con ella. Por lo tanto, se puede decir que la hipótesis es válida.

15.2. Conclusiones generales

Referente al trabajo se puede afirmar la siguiente observación. Si se hubiera tenido más datos de cada una de las variables estudiadas, habría estado mejor para hacer la validación del modelo, debido a que en este trabajo se parte desde el año 2007 y se nota mucho que la escasez de datos de las variables hace que durante todo el trabajo se presenten todos los problemas que se han ido observando durante el trabajo.

El modelo que se ha obtenido, aun teniendo los problemas de la escasez de datos es un modelo ARIMAX válido con las variables M^2 , Parados y el PIB, aunque se puede apreciar que con más datos se podría haber sacado algún modelo más y se podría haber realizado un modelo ARIMAX con más variables.

El trabajo, a nivel personal ha sido muy bueno. He podido aplicar los conocimientos aprendidos durante la etapa universitaria y su vez, he profundizado en la herramienta de estadística STATGRAPHICS. También, he aprendido a utilizar otra herramienta de estadística como es R Studio y he adoptado conocimientos más amplios que los trabajados durante mi estancia en la universidad. Por lo que, a nivel personal, estoy muy contento por todo lo aprendido.

Por último, gracias a este trabajo me ha servido para darme cuenta de la importancia que tiene el índice de precios de la vivienda, debido a que este dato representa todo a nivel general, y gracias a las variaciones de este índice se puede ver si una comunidad está creciendo o no, o incluso si esta aumentado el sistema inmobiliario.

Bibliografía

- Rodríguez, O. (2021, 24 enero). *De la crisis del ladrillo a la del covid: el número de inmobiliarias crece un 58%*. El Independiente.
<https://www.elindependiente.com/economia/2021/01/24/de-la-crisis-del-ladrillo-a-la-del-covid-el-numero-de-inmobiliarias-crece-un-58/>
- Peña Cerezo, M. Á., Ruiz Herrán, V, V., & García Merino, J. D. (2002, mayo). *Análisis de los factores determinantes del precio del activo vivienda*. <https://addi.ehu.es/>.
[https://addi.ehu.es/bitstream/handle/10810/9100/Pe%
c3%b1a-Ruiz-Garc%
c3%ada_An%
c3%a1lisis%
20de%
20los%
20factores%
20determinantes.pdf?sequence=1&isAllowed=y](https://addi.ehu.es/bitstream/handle/10810/9100/Pe%c3%b1a-Ruiz-Garc%c3%ada_An%c3%a1lisis%20de%20los%20factores%20determinantes.pdf?sequence=1&isAllowed=y)
- Martínez Pagés, J., & Maza Lasierra, L. A. (2003). *Análisis del precio de la vivienda en España*. <https://repositorio.bde.es>.
<https://repositorio.bde.es/bitstream/123456789/6770/1/dt0307.pdf>
- Análisis regional del mercado laboral y de la inflación*. (2005). Dialnet.
<https://dialnet.unirioja.es/servlet/revista?codigo=8349>
- Abad Sánchez, J. (2013, junio). *Precios de la vivienda: Sobrevaloración y burbuja*. www.ucm.es. <https://www.ucm.es/data/cont/docs/518-2013-10-23-Abad13.pdf>
- Hernando, J. R. (2015, 17 julio). *Análisis y pronóstico del precio de la vivienda en España: modelo econométrico desde una perspectiva conductual | Revista de Estudios Empresariales. Segunda Época*. revistaselectronicas.ujaen.es.
<https://revistaselectronicas.ujaen.es/index.php/REE/article/view/2087>

- Vayá, E. (2018, 23 noviembre). *Dipòsit Digital de la Universitat de Barcelona: Factores determinantes de la diferencia de precios en el mercado inmobiliario. Caso aplicado a los municipios de Cataluña*. diposit.ub.edu.
<http://diposit.ub.edu/dspace/handle/2445/126381>
- LÓPEZ SERRANO, A. (2019, mayo). *Evolución de precios en el mercado inmobiliario (vivienda)*. <https://rua.ua.es/>.
https://rua.ua.es/dspace/bitstream/10045/93289/1/Evolucion_de_precios_en_el_mercado_inmobiliario_vivien_LOPEZ_SERRANO_ANTONIO.pdf
- Sagner T., A. (2011). *Determinantes del precio de viviendas en la región metropolitana de Chile*. www.scielo.org.mx.
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-718X2011000400813&lng=es&tlng=es.
- CONSUELO COLOM, M., & CRUZ MOLÉS, M. (2020). *Un análisis sobre el gasto en servicios de vivienda en España*. <https://www.ine.es>.
https://www.ine.es/ss/Satellite?blobcol=urldata&blobheader=application%2Fpdf&blobheadertype=Content-Disposition&blobheadervalue1=attachment%3B+filename%3D143_5.pdf&blobkey=urldata&blobtable=MungoBlobs&blobwhere=298%2F129%2F143_5%2C0.pdf&ssbinary=true
- INE - Instituto Nacional de Estadística. (s. f.). *Sección prensa / Índice de Precios de Vivienda (IPV)*. Recuperado 29 de julio de 2021, de
https://www.ine.es/prensa/ipv_prensa.htm

Anexos

- Anexo 1 Base de datos Excel de todas la variables utilizadas durante el trabajo
- Anexo 2 ANEXO II Instrucciones para la ejecución correcta en el programa R-Studio.