Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Estudios de Postgrado
Maestría en Estadística Aplicada

ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018

Lic. Carlos Enrique Santa Cruz Hernández
Asesorado por el Mtro. Luis Alejandro Samayoa Alvarado

Guatemala, noviembre de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

LIC. CARLOS ENRIQUE SANTA CRUZ HERNÁNDEZ
ASESORADO POR EL MTRO. LUIS ALEJADRO SAMAYOA ALVARADO

AL CONFERÍRSELE EL TÍTULO DE

MAESTRO EN ESTADÍSTICA APLICADA

GUATEMALA, NOVIEMBRE DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Mtro. Edwin Adalberto Bracamonte Orozco
EXAMINADOR	Mtro. William Eduardo Fagiani Cruz
EXAMINADOR	Dra. Aura Marina Rodríguez Pérez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería, con fecha 31 de enero de 2022.

Lic. Carlos Enrique Santa Cruz Hernández

Decanato
Facultad de Ingeniería
24189101- 24189102
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.798.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018,** presentado por: **Carlos Enrique Santa Cruz Hernández,** que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DECANA
FACULTAD DE INGENIERÍA

Inga. Aurelia Anabela Cordova Estrada
Decana

Guatemala, noviembre de 2022

AACE/gaoc

Guatemala, noviembre de 2022

LNG.EEP.OI.798.2022

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

"ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018"

presentado por **Carlos Enrique Santa Cruz Hernández** correspondiente al programa de **Maestría en artes en Estadística aplicada** ; apruebo y autorizo el mismo.

Atentamente,

"Id y Enseñad a Todos"

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA · FACULTAD DE INGENIERÍA · DIRECCIÓN · ESCUELA DE POSTGRADO

Mtro. Ing. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería

Guatemala, 29 de septiembre de 2022

M.A. Ing. Edgar Dario Alvarez Coti
Director
Escuela de Estudios de Postgrado
Presente

Estimado M.A. Ing. Alvarez Coti

Por este medio informo a usted, que he revisado y aprobado el **INFORME FINAL y ARTÍCULO CIENTÍFICO** titulado: **ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018** del estudiante **Carlos Enrique Santa Cruz Hernández** quien se identifica con número de carné **200812620** del programa de Maestria En Estadistica Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el ***Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014.*** Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

UNIVERSIDAD DE SAN CARLOS
ESCUELA DE POSTGRADO
FACULTAD DE INGENIERIA
DE GUATEMALA

Mtro. Ing. Edwin Adalberto Bracamonte Orozco
Coordinador
Maestria En Estadistica Aplicada
Escuela de Estudios de Postgrado

Guatemala, 29 de septiembre de 2022

M.A. Ing. Edgar Dario Alvarez Coti
Director
Escuela de Estudios de Postgrados
Presente

Estimado M.A. Ing. Alvarez Coti

Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: **"ANÁLISIS DE SERIES TEMPORALES DE LAS AFILIACIONES A LA SEGURIDAD SOCIAL Y SU RELACIÓN CON LOS SECTORES PRODUCTIVOS DE GUATEMALA, PERÍODO 2001-2018"** del estudiante **Carlos Enrique Santa Cruz Hernández** del programa de **Maestría En Estadística Aplicada** identificado con número de carné 200812620.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Msc. Lic. Luis Alejandro Samayoa Alvarado
Colegiado No. 24083
Asesor de Tesis

Luis Alejandro Samayoa Alvarado
ECONOMISTA
No. de Colegiado 24083

ACTO QUE DEDICO A:

Dios	Por su misericordia y bondad.
Mi Madre	Ana Gloria Hernández, por la vida, sus cuidados y amor incondicional.
Mi Esposa	Isabel de Santa Cruz, por su amor y apoyo en todo momento, te amo, eres el ser que Dios puso a mi lado para demostrarme su amor en cada instante.
Mis Hijos	Joao, Joann y Johan Santa Cruz, quiero ser un buen ejemplo para ustedes, agradezco todo su amor y comprensión por el tiempo que no tuvimos para jugar juntos, espero reponer esos fines de semana. Ustedes son mi motor y combustible para seguir esforzándome por una vida mejor.

AGRADECIMIENTOS A:

El Pueblo de Guatemala	Por el pago sus impuestos que cubren el precio de la inversión en la educación superior de sus habitantes.
La Universidad de San Carlos de Guatemala	Por volverme aceptar en sus aulas y seguir enseñándome tanto en la vida.
La Facultad de Ingeniería, Escuela de Estudios de Postgrado	Por la oportunidad de ser mi segunda facultad en donde he aprendido valiosos conocimientos para poner en práctica al servicio de nuestro país.
Mis amigos	Por su amistad sincera y desinteresada, los aprecio mucho.
Maestro y amigo	Luis Alejandro Samayoa Alvarado, por ser ejemplo de excelencia académica, compartir sus conocimientos, aportes y guía técnica para el logro de la presente tesis.
Mis docentes de la Maestría	Por la enseñanza en cada uno de los cursos del pensum de estudio; sin duda alguna, los conocimientos serán puestos en práctica para contribuir en la construcción de un mejor país.

Al Instituto Guatemalteco de Seguridad Social	Por ser una noble institución, producto de la lucha del pueblo que debe cuidarse y defenderse para que cada día nuestros hijos e hijas tengan las garantías mínimas de dignidad humana.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES V
LISTA DE SÍMBOLOS XI
GLOSARIO XIII
RESUMEN XIX
PLANTEAMIENTO DEL PROBLEMA XXI
OBJETIVOS XXVII
RESUMEN DEL MARCO METODOLÓGICO XXIX
INTRODUCCIÓN XXXV

1. MARCO REFERENCIAL 1
   1.1. Análisis de regresión múltiple 1
   1.2. Series de tiempo 5

2. MARCO TEÓRICO 7
   2.1. Series de tiempo 7
      2.1.1. Componente de una serie temporal 8
      2.1.2. Pruebas de estacionariedad 10
      2.1.3. Tipos de tendencia 11
      2.1.4. Prueba de Primera diferencia 11
      2.1.5. Pruebas de Estacionalidad 12
      2.1.6. Medias Móviles 13
      2.1.7. Método de Holt-Winters 14
      2.1.8. Pronósticos 15
      2.1.9. Modelos ARIMA 15
      2.1.10. Metodología de *Box-Jenkins* (BJ) 17

2.2. Modelos de regresión 18
2.2.1. Supuestos de un modelo de regresión 19
2.2.2. Pruebas de supuestos 21
2.2.3. Criterios de comparación 23
2.2.3.1. Cp de Mallows 24
2.2.3.2. Criterio de información de *Akaike* 24
2.2.3.3. Criterio *Bayesiano* (BIC) 24
2.2.4. Seguridad Social 25
2.2.4.1. Seguro social 26
2.2.4.2. Seguridad social 26
2.2.4.3. Asistencia Social 27
2.2.4.4. Previsión social 27
2.2.4.5. Seguridad Social en Guatemala 28
2.2.4.5.1. Campo de aplicación 30
2.2.5. Beneficios de la seguridad social 30
2.2.6. Financiamiento de la seguridad social 31
2.3. Actividades Económicas 32
2.4. Crecimiento económico, empleo y seguridad social 34

3. PRESENTACIÓN DE RESULTADOS 37
3.1. Grado de fuerza o asociación y dependencia estadística 37
3.1.1. Análisis exploratorio de la seguridad social 37
3.1.1.1. Pruebas de normalidad 37
3.1.2. Grado de fuerza o asociación lineal 44
3.2. Comprobación de supuestos 47
3.2.1. Supuestos del análisis de regresión múltiple 47
3.2.1.1. Relación lineal entre variables 47
3.2.1.2. Valores atípicos 49
3.2.1.3. Distribución normal 50

3.2.1.4. Independencia 51
3.2.1.5. Homocedasticidad 52
3.2.2. Supuestos del análisis de series de tiempo 53
3.2.2.1. Estacionariedad 53
3.2.2.2. Homogeneidad 55
3.2.2.3. Independencia en las observaciones... 56
3.3. Modelos de predicción 58
3.3.1. Modelo 1 de análisis de regresión múltiple 58
3.3.1.1. Diagnóstico modelo 1 59
3.3.1.1.1. Criterio Cp de Mallows . 60
3.3.1.1.2. Criterio VIF 61
3.3.1.1.3. Criterio *Akaike* (AIC)..... 61
3.3.1.1.4. Criterio *Bayesiano* (BIC) 61
3.3.1.2. Diagnóstico modelo 2 62
3.3.1.3. Diagnóstico modelo 3 64
3.3.1.4. Análisis residuales del modelo..... 66
3.3.1.4.1. Normalidad 66
3.3.1.4.2. Datos Atípicos 67
3.3.1.4.3. Independencia 68
3.3.1.4.4. Homoscedasticidad 68
3.3.1.4.5. Análisis de influencia 69
3.3.1.5. Interpretación general del modelo 3 71
3.4. Modelos de pronóstico 73
3.4.1. Pronóstico afiliaciones a la seguridad social 73
3.4.2. Pronósticos actividades económicas 83
3.4.2.1. Agricultura, Ganadería y caza 83
3.4.2.2. Industria manufacturera 88
3.4.2.3. Comercio al por mayor y al por menor. 93
3.4.2.4. Transporte 98

3.4.2.5. Administración pública y defensa 103
3.5. Objetivo general 107

4. DISCUSIÓN DE RESULTADOS 109
4.1. Análisis interno 109
4.2. Análisis externo 111

CONCLUSIONES 115
RECOMENDACIONES 117
REFERENCIAS 119
APÉNDICES 129

ÍNDICE DE ILUSTRACIONES

FIGURAS

1. Desestacionalización de una serie por medias móviles 14
2. Conjunto de dos datos con correlación de segundo grado 21
3. Gráfica de normalidad, seguridad social, agricultura e industria 40
4. Gráfica de normalidad de electricidad construcción, comercio privados y transporte 41
5. Gráfica de normalidad de intermediación financiera, alquiler de viviendas, servicios privados y administración pública..... 42
6. Variaciones inter trimestrales de la seguridad social 2001-2018 43
7. Variación inter trimestral de las actividades económicas 2001-2018……..... 43
8. Mapa de calor matriz de correlaciones, de las variaciones inter trimestrales de las afiliaciones a la seguridad social y las variaciones inter trimestrales de las actividades económicas 44
9. Grado de dependencia estadística de las variaciones inter trimestrales de las afiliaciones a la seguridad social, respecto de las variaciones inter trimestrales de las actividades económicas 45
10. Diagrama de cajas residuos no estandarizados primer modelo..... 49
11. Q-Q plot de los residuos..... 50
12. Dispersión en los residuos 52
13. Prueba SNHT alquiler de vivienda 56
14. Análisis de regresión múltiple del primer modelo 59
15. Análisis de regresión múltiple segundo modelo 62

16. Análisis de regresión múltiple tercer modelo.....64
17. Prueba de normalidad en los residuos del modelo 3 67
18. Análisis de independencia de los residuos modelo 3.....68
19. Resumen de casos influyentes y distancia de Cook 69
20. Apalancamiento Leverage 70
21. Distancia de Cook..... 70
22. Residuales ACF y PACF pronóstico de serie acumulada de las afiliaciones a la seguridad social.....76
23. Comparación cifras reales vs pronóstico 2019 afiliaciones a la seguridad social..... ……77
24. Pronóstico de las afiliaciones a la seguridad social en Guatemala año 2019 78
25. Residuales de ACF Y PACF pronóstico de serie con variaciones Inter trimestral de las afiliaciones a la seguridad social 81
26. Comparación de variación real vs pronóstico del año 2019 variaciones inter trimestrales de las afiliaciones a la seguridad social con un nivel de 95 % de confianza 82
27. ACF y PACF de residuales del modelo de pronóstico de la actividad económica de agricultura.....85
28. Pronóstico de la actividad económica de agricultura para el año 2019.. …… 86
29. Comparación de valores reales y valores pronosticados de las variaciones de la actividad económica agricultura 2019 con un nivel de 95 % de confianza 87
30. ACF y PACF residuales del modelo, industria manufacturera 90
31. Pronóstico 2019 industria manufacturera 91
32. Comparación de valores reales y valores pronosticados de las variaciones de la actividad económica industria manufacturera 2019 con un nivel del 95 % de confianza.....92

33. ACF y PACF residuales del modelo, comercio al por mayor y al por menor 95
34. Pronóstico 2019 de las variaciones de la actividad económica comercio al por mayor y por menor con un nivel del 95 % de confianza …….....……. 96
35. Comparación de valores reales vs valores pronosticados de las variaciones de la actividad económica comercio al por mayor y menor 2019, con un nivel de confianza del 95 % 97
36. ACF Y PACF de residuales del modelo de la actividad económica de transporte 100
37. Pronóstico 2019 de las variaciones de la actividad económica de transporte, almacenamiento y comunicaciones con un nivel del 95 % de confianza 101
38. Comparación de valores reales vs pronóstico 2019 de la actividad económica de transporte, con un nivel del 95 % de confianza..... 102
39. ACF Y PACF de residuales en modelo de la administración pública y defensa 105
40. Pronóstico 2019 de la actividad económica de administración pública y defensa..... 105
41. Pronóstico 2019 actividad económica de administración pública y defensa con un nivel del 95 % de confianza 106
42. Comparación de valores reales vs pronóstico de 2019 de la actividad económica administración pública y defensa 106

TABLAS

I. Variables del Estudio XXX
II. Pruebas de normalidad en variables de estudio 38
III. Estadísticos descriptivos variables de estudio 39
IV. Datos atípicos Prueba de Grubbs 46
V. Variables con datos atípicos 46
VI. Coeficientes de correlación y determinación.....47
VII. Pruebas de normalidad 51
VIII. Análisis de dependencia de los residuos 51
IX. Prueba de White 53
X. Estacionariedad 54
XI. Prueba de Homogeneidad Normal Estándar 55
XII. Pruebas de independencia de Ljung-Box 57
XIII. Ecuación de regresión primer modelo 60
XIV. Ecuación de regresión segundo modelo.....63
XV. Ecuación de regresión tercer modelo 64
XVI. Comparación de modelos de análisis de regresión lineal múltiple.....65
XVII. Datos atípicos en residuales en el modelo 3.....67
XVIII. Ecuación final de regresión múltiple, variaciones inter trimestrales de las afiliaciones a la seguridad social respecto a las variaciones inter trimestrales de las actividades económicas de Guatemala en el período 2001-2018..... 71
XIX. Modelo de pronóstico de mejor ajuste serie acumulada de afiliaciones a la seguridad social 73
XX. Estadísticos de ajuste del modelo serie acumulada de afiliaciones a la seguridad social 74
XXI. Diagnóstico del modelo afiliaciones a la Seguridad Social 74

XXII. Parámetros del modelo de suavizado exponencial, afiliaciones a la seguridad social 75
XXIII. Valores pronosticados de las afiliaciones de la seguridad social cuatrimestres del 2019, 95 % de confianza..... 76
XXIV. Modelo de pronóstico de mejor ajuste variaciones Inter trimestrales de la seguridad social..... 78
XXV. Estadísticos de ajuste del modelo con variaciones inter trimestrales de la seguridad social..... 79
XXVI. Diagnóstico del modelo de las variaciones Inter trimestrales de la seguridad social 80
XXVII. Pronóstico de la variación Inter trimestral de las afiliaciones a la seguridad social 2019 81
XXVIII. Modelo de pronóstico de la actividad de agricultura..... 83
XXIX. Estadísticos de ajuste del modelo ARIMA, actividad económica de agricultura 84
XXX. Diagnóstico del modelo de predicción, actividad económica de agricultura 84
XXXI. Pronóstico de la actividad económica de agricultura del año 2019 85
XXXII. Modelo de pronóstico de industria..... 88
XXXIII. Estadísticos de ajuste del modelo 89
XXXIV. Diagnóstico del modelo de industria manufacturera..... 89
XXXV. Pronóstico año 2019, industria manufacturera 90
XXXVI. Modelo de pronóstico de la actividad económica del comercio 93
XXXVII. Estadísticos de ajuste del modelo de pronóstico del comercio 94
XXXVIII. Diagnóstico del modelo de pronóstico del comercio 94
XXXIX. Pronóstico año 2019, comercio al por mayor y al por menor 96

XL. Modelo de pronóstico de la actividad económica trasporte 99
XLI. Estadísticos de ajuste modelo de transporte 99
XLII. Diagnóstico del modelo de la actividad económica de transporte..... 100
XLIII. Pronóstico 2019 de la actividad económica de transporte.....101
XLIV. Modelo de pronóstico de la actividad económica de administración pública 103
XLV. Estadísticos de ajuste de la actividad económica de administración pública 104
XLVI. Diagnóstico del modelo de la administración pública.....104

LISTA DE SÍMBOLOS

Símbolo	Significado
AE	Actividad Económica
$\boldsymbol{Y_i}$	Afiliaciones a la Seguridad Social
AR	Autorregresivos.
BJ	*Box-Jenkins.*
R	Coeficiente de Correlación de *Pearson*
$\mathbf{R^2}$	Coeficiente de determinación
$\mathbf{R^2A}$	Coeficiente de determinación ajustado
$\boldsymbol{\beta_1}$ **y** $\boldsymbol{\beta_2}$...	Coeficientes de regresión
AIC	Criterio de información *Akaike*
BIC	Criterio de información bayesiano
Δ	Delta
$\boldsymbol{D_i}$	Distancia de *Cook*
$\mathbf{X^2}$	Distribución Chi cuadrado
$\boldsymbol{d}$	*Durbin Watson*
ECMP	Error cuadrado medio poblacional
$\mathbf{u_t}$	Error de ruido
MAPE	Error porcentual absoluto medio
$\boldsymbol{h_{ii}}$	Estadístico de *Leverage*
$\boldsymbol{S_t}$	Factor estacional
ACF	Función de autocorrelación
PACF	Función de autocorrelación parcial
gl	Grados de libertad

H1	Hipótesis alterna
H0	Hipótesis nula
$\boldsymbol{VIF}$	Inflación de la varianza
$\boldsymbol{\beta_0}$	Intersección o intercepto
SK	*Kolmogórov-Smirnov*
$\boldsymbol{u}$	Media
MA	Medias móviles
Sig	Nivel de significancia
n	Número de observaciones
$\boldsymbol{\varphi p(L)}$	Operador de retardo polinomial
$\boldsymbol{t}$	Período
$\boldsymbol{t-1}$	Período menos 1
$\Delta \boldsymbol{Y_t}$	Primera diferencia
ARIMA	Proceso autorregresivo integrado por media móvil
PIB	Producto Interno Bruto
White	Prueba de *White*
SNHT	Prueba Normal Estándar de Homogeneidad
RES	Residuos
RESIDEST	Residuos estudentizados
ss	Seguridad Social
S-W	*Shapiro-Wilk*
P-value	Valor P
$\boldsymbol{Y}$	Variable dependiente
X1 y X2…	Variables independientes o explicativas
$\boldsymbol{\sigma^2}$	Varianza

GLOSARIO

Aleatoriedad	Palabra usada en sentidos no determinísticos (de un proceso aleatorio) a puramente por oportunidad, independientemente de otros eventos.
Análisis de Correlación	Es el estudio que se realiza para medir la intensidad o grado de la asociación que existe entre variables numéricas.
Análisis de Datos	Proceso de transformación de datos en información utilizable que a menudo se presenta en forma de un artículo analítico publicado, con el fin de agregar valor a la producción estadística.
Análisis de Regresión	Es el estudio que se realiza con el propósito de hacer predicciones. el objetivo es el desarrollo de un modelo estadístico que pueda ser utilizado para predecir valores de una variable dependiente, basado en los valores de la variable independiente.
Autocorrelación	Se denomina así a la correlación de una variable consigo misma cuando se desfasa uno o más periodos de tiempo. Se determina calculando el coeficiente de autocorrelación.
Caeterīs Pāribus	Mientras todo lo demás se mantenga constante.

Coeficiente de Correlación Lineal de Pearson	Es un número que mide la intensidad de la asociación lineal entre dos variables. El coeficiente de correlación se representa simbólicamente por r o R.
Coeficiente de Determinación	Es un valor que se obtiene elevando al cuadrado el coeficiente de correlación. Se representa por R^2.
Coeficientes de Regresión	Son los valores constantes de una ecuación de regresión lineal.
Correlograma	Es un gráfico que permite apreciar las autocorrelaciones r1, r2…, rk mediante el cual se identifican si los datos de una serie de tiempo tienen las siguientes características: estacionalidad, aleatoriedad, tendencia y estacionariedad.
Covarianza	Es una medida de la asociación lineal entre dos variables.
Cp Mallows	Prueba de evaluación del ajuste del modelo de regresión.
Desviación Estándar	Conocida también como desviación típica, es una medida de dispersión que se obtiene con la raíz cuadrada de la varianza.
Dickey Fuller	Prueba para detectar estacionariedad en una serie temporal.

Dispersión	Se refiere a la separación de los datos en una distribución, es decir, al grado en que las observaciones se separan.
Histograma	Es un tipo de gráfica de barras utilizado para representar la distribución de frecuencia de una variable.
Independencia Estadística	Se dice que dos variables X e Y son independientes, estadísticamente, cuando la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales en todos los casos.
Intervalo de confianza	El intervalo de confianza o límite de confianza, es un rango de valores entre los cuales se espera con cierto nivel de confianza, se encuentre el valor poblacional estimado según el estadístico resumen de interés.
Jamovi	Software estadístico gratuito y abierto para el análisis estadístico.
Minitab	Programa de computadora para ejecutar funciones estadísticas.
OIT	Organización Internacional del Trabajo.
Proporción	Cociente que representa una parte con referencia al todo. Se calcula utilizando, en el numerador y el

	denominador, información referida a una misma categoría de hechos u objetos.
Rstudio	Entorno de desarrollo integrado para el lenguaje de programación.
Series Temporales	Conocida también como serie cronológica. Es la sucesión de observaciones cuantitativas ordenadas en el tiempo de un fenómeno. Los datos tienen un orden que no es posible variar, la información puede ser mensual, trimestral, anual o de cualquier otro intervalo temporal.
SPSS	*Statistical Package for the Social Sciences.*
Validación	Actividad mediante la cual se corrobora que los procesos, herramientas y resultados de la producción estadística son correctos o bien que cumplan con los criterios de aceptabilidad definidos.
Variable Discreta	Una variable se considera variable discreta cuando esta puede tomar solo un número finito o contablemente infinito (dentro de los números enteros positivos) de valores distintos.
Variable	Son características o atributos de las unidades de estudio. Se les llama variables debido a que los valores pueden cambiar de una unidad de estudio a otra.

Varianza	Es un estimador de la dispersión de una variable aleatoria respecto de su media. Es la esperanza matemática del cuadrado de la diferencia de cada valor (poblacional o muestral) con respecto a su media, puede tomar valores entre 0 y 1.

RESUMEN

El propósito de la presente investigación consistió en la construcción de modelos de predicción y pronóstico aplicados en afiliaciones a la seguridad social y actividades económicas de Guatemala, de acuerdo con los datos de las afiliaciones a la seguridad social y las actividades económicas del período 2001 a 2018, para luego comparar los pronósticos con los registros reales de los cuatro trimestres del 2019.

El objetivo general del estudio fue estimar el grado de asociación y de dependencia estadística de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala, analizar las series temporales de las afiliaciones a la seguridad social y las actividades económicas de Guatemala del 2001 al 2018, mediante el análisis de regresión múltiple y de series de tiempo para determinar identificar los modelos de mejor ajuste de predicción y pronósticos en el corto plazo.

Para cumplir con los objetivos propuestos en la investigación, se utilizó el análisis de regresión múltiple y el análisis de series de tiempo, comprobando cada uno de los supuestos, el diagnóstico y validación para cada modelo, tanto en los datos puros, como en los residuos, siguiendo el principio de parsimonia donde se busca el modelo más sencillo con el mayor grado de explicación posible.

Como resultados, se identificaron las cinco principales actividades económicas que explicaron de mejor manera el comportamiento de las afiliaciones a la seguridad social el período antes mencionado; así como, la

comparación de los valores pronosticados en el corto plazo con los registros reales de los trimestres del 2019.

Con lo anterior, fue posible recomendar a la institución de seguridad social en Guatemala, que continúe proporcionando información para fortalecer los conocimientos y análisis en la realización de propuestas y con eso ampliar la cobertura de afiliados a un régimen tan importante para las familias guatemaltecas.

PLANTEAMIENTO DEL PROBLEMA

- Contexto general

En cuanto a la seguridad social, la ONU (2008) menciona que, la misma es un derecho humano reconocido por los Estados parte de esa Organización, en donde el Pacto Internacional de Derechos Económicos, Sociales y Culturales declaran que, El derecho a la seguridad social es de importancia fundamental para garantizar a todas las personas su dignidad humana cuando hacen frente a circunstancias que les privan de su capacidad para ejercer plenamente los derechos.

En este sentido, la ONU (2008) sigue explicando que, la seguridad social posee un carácter redistributivo, desempeña un papel importante para reducir y mitigar la pobreza, prevenir la exclusión social y promover la inclusión social.

Para el caso de Guatemala, la cobertura de la seguridad social no ha sido significativa; es decir, el país no ha alcanzado coberturas mayores al 29 % de la población económicamente activa (PEA). Según el informe de resultados del XII Censo Nacional de la Población y VII de Vivienda, en el año 2018, se registraron 9,928,561 personas mayores de 15 años (población en edad de trabajar); de acuerdo con estimaciones propias de la base de datos de dicho censo, las personas con edad igual o mayor a 15 años y menores y de igual edad a 65 años, fueron 9,167,226. Considerando que, según la definición del Diccionario Prehispánico, la PEA la constituye la población de 16-65 años (15-65 para Guatemala), se puede decir que la cobertura de la seguridad social fue de 14.46 %. La PEA tiene una clasificación de personas en edad de trabajar y que están

ocupadas y/o buscando empleo, en ese sentido, el Instituto Nacional de Estada dística (INE) reportó 5,020,548 personas; bajo este segundo escenario, para el año 2018, existió una cobertura de 26.40 %. (INE, 2019)

En cambio, de acuerdo con los datos del censo del 2002, la población entre 15 y 65 años fue de 6,038,155, es decir, 48.18 % menor al 2018; el registro de afiliados de ese año fue de 953,052, una cobertura de seguridad social de 15.78 %, de acuerdo con el criterio de la PEA ocupada y/o buscando empleo, el INE registró un total de 3,224,361, de esa cuenta, se puede mencionar una cobertura de 29.56 %. (INE,2002)

Los factores que intervienen en el aumento o disminución de la cobertura a la seguridad social son diversos; por una parte, para ser afiliado cotizantes al régimen de la seguridad social, el trabajador debe poseer una relación laboral de dependencia (empleo formal), las empresas que contratan a dichos trabajadores están obligadas por parte del Estado a cotizar al Seguro Social. No obstante, persiste una alta representación del empleo informal en la economía del país. Según el Banco de Guatemala en el año 2019, estimó que dicho sector representó el 22 % del Producto Interno Bruto. (BANGUAT, 2019)

Lo anterior supone que, las personas que trabajan en la informalidad no cuentan con garantías mínimas que brinda la seguridad social, tales como: servicios médicos en caso de enfermedad o accidentes, subsidios en casos de invalidez o sobrevivencia y retribuciones económicas (pensiones) cuando alcanzan la tercera edad o entre otras.

Los factores que intervienen en el aumento o disminución de la cobertura a la seguridad social son diversos, entre los que se pueden mencionar el empleo informal y las nuevas formas de contratación. Si bien, las empresas que contratan

trabajadores bajo una relación laboral de dependencia (empleo formal) están obligadas por parte del Estado a cotizar al Seguro Social, persiste una alta representación del empleo informal en la economía del país. Según el Banco de Guatemala en el año 2019, estimó que dicho sector representó el 22 % del Producto Interno Bruto. (BANGUAT, 2019)

En ese sentido; se considera que, las personas que trabajan en la informalidad no cuentan con garantías mínimas que brinda la seguridad social, tales como: servicios médicos en caso de enfermedad o accidentes, subsidios en casos de invalidez o sobrevivencia y retribuciones económicas (pensiones) cuando alcanzan la tercera edad o entre otras.

Por otro lado, cada vez crece el número de empresas que emplean modalidades distintas a la de contratación formal (relación de dependencia) promoviendo que las personas trabajen bajo subcontratos o por modalidad de facturación por servicios prestados, Para estos, según la ley nacional vigente no existe obligatoriedad de una relación formal de empleo, provocando que las personas no sean incluidas en planillas y por tal razón, no coticen al régimen de seguridad social, entre otros factores.

Además, el crecimiento económico sostenido del país no se ha traducido en un crecimiento de la cobertura de la seguridad social. El nivel de crecimiento económico de Guatemala expresado por la variación interanual del Producto Interno Bruto (PIB) a precios constantes, ha mantenido un ritmo sostenido promedio de 3.5 % del 2001 al 2018, con una desviación estándar de 1.2 %. (BANGUAT, 2019)

Sin embargo, para el caso de Guatemala, (en diversas variables macroeconómicas) que haya crecimiento económico no se traduce en beneficio

de las mayorías sociales, en lo que respecta a la seguridad social, el hecho que se registre crecimiento económico no significa (aunque debería) que exista crecimiento de personas afiliadas al régimen de seguridad social, considerando que, son las personas como fuerza de trabajo las que contribuyen al crecimiento económico en cada actividad productiva.

- Descripción del problema

No se conocen los modelos de mejor ajuste de predicción y pronóstico, realizado con el análisis de regresión múltiple y de series de tiempo entre las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018.

- Formulación del problema

Para la formulación se hace necesario plantarse una pregunta central y cuatro auxiliares en donde se pueda dar respuesta a la descripción del problema.

  - Pregunta central

¿Cuáles son los modelos de mejor ajuste de predicción y pronóstico realizado con el análisis de regresión múltiple y de series de tiempo, para las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018?

  - Preguntas auxiliares

    - ¿Cuál es el grado de asociación y de dependencia estadística entre las afiliaciones a la seguridad social

respecto de las actividades económicas de Guatemala en el período 2001-2018?

- ¿Se cumplen los supuestos del análisis de regresión múltiple y de series de tiempo en el registro de las afiliaciones a la seguridad social y el registro de las actividades económicas de Guatemala en el período 2001-2018?

- ¿Cuál es el modelo de predicción de mejor ajuste de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018?

- ¿Cuál es el modelo de pronóstico de mejor ajuste de corto plazo de las afiliaciones a la seguridad social y de las actividades económicas en Guatemala en el período 2001-2018?

- Delimitación del problema

El problema fue analizado con base en el registro trimestral de las afiliaciones a la seguridad social y de las actividades económicas del país, estos últimos, expresados en millones de quetzales a precios constantes del año de referencia 2001, con una duración de 18 años, que corresponde al período 2001-2018.

OBJETIVOS

General

Estimar el grado de asociación y dependencia estadística de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala, del 2001 al 2018, mediante el análisis de regresión múltiple y de series de tiempo para construir los modelos de mejor ajuste de predicción y pronósticos en el corto plazo (períodos menores a un año).

Específicos

1. Estimar el grado de fuerza o asociación y dependencia estadística entre las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala entre los años 2001 y 2018 a través del análisis de correlación y de regresión múltiple.

2. Comprobar los supuestos del análisis de regresión múltiple y de series de tiempo de las afiliaciones a la seguridad social y de las actividades económicas de Guatemala del período de 2001 a 2018 a través de las pruebas estadísticas para cada tipo de análisis.

3. Estimar el modelo de predicción de mejor ajuste de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala entre los años 2001 y 2018 a través del diagnóstico y validación de cada modelo.

4. Estimar el modelo de pronóstico de mejor ajuste a corto plazo de las afiliaciones a la seguridad social y de las actividades económicas en Guatemala según los datos de los años 2001 a 2018 a través del diagnóstico y validación del modelo.

RESUMEN DEL MARCO METODOLÓGICO

A continuación, se presentan los criterios establecidos para el proceso de investigación, se resumen las características del estudio, se describen las variables de interés (unidades de análisis) y cada una de las fases que se ejecutaron para el logro de los objetivos propuestos.

- Características del estudio

El Estudio propuesto tuvo un enfoque cuantitativo, derivado que sus variables principales fueron: el registro monetario trimestral en moneda nacional a precios reales del año 2001 al 2018 de las actividades económicas del país, variable tipo numérica de naturaleza cuantitativa, continua, con escala de razón; y, el registro trimestral de las afiliaciones a la seguridad social, variable tipo numérica, de naturaleza cuantitativa, discreta, con escala de razón.

El diseño fue no experimental (observacional), dado que no es posible incidir en los resultados registrados en cada período; por tal razón se trabajó con las series tal cual fue su comportamiento absoluto y relativo en los años seleccionados.

El tipo del estudio fue correlacional, pues se relacionaron las variables de las actividades económicas y las afiliaciones a la seguridad social en Guatemala en el período de 2001 a 2018, el alcance fue descriptivo, correlacional y predictivo, dado que se utilizó la estadística descriptiva e inferencial para conocer primeramente el comportamiento de las variables, luego, la determinación de la relación existente entre las afiliaciones a la seguridad social y las actividades

económicas del país, por último, la estimación de pronósticos de corto plazo (períodos menores a un año).

- Unidades de Análisis

Las unidades de análisis correspondieron al total de la población afiliada a la seguridad social en Guatemala y el registro de los valores en unidades monetarias nacionales de las actividades económicas que conformaron el Producto Interno Bruto por el lado de la producción, a precios de referencia del año 2001, desde el 2001 al 2018.

- Variables

Tabla I. **Variables del Estudio**

Variable	Definición Teórica	Definición Operativa	Tipo, Naturaleza y Escala
Variación inter trimestral de los afiliados a la Seguridad Social (ss)	Constituye el número de cotizantes a la seguridad social en Guatemala, muestra la cobertura de la población, respecto al número de población económicamente activa (PEA) (Cotizantes a la seguridad social /PEA) período x *100) =cobertura	El registro de la población afiliada o cotizante de la seguridad social, la realiza el Instituto Guatemalteco de Seguridad Social como ente rector de dicha materia, se cuenta con el registro por trimestre y por actividad económica del cual se tiene acceso a través de la ley de libre acceso a la información pública Decreto 57-2008.	Tipo numérica, de naturaleza cuantitativa, discreta, con escala de razón.

Continuación tabla I.

Variación inter trimestral de las actividades económicas del País año de referencia 2001 (ae)	Corresponde a los sectores productivos que registra el Banco de Guatemala, también denominadas actividades económicas las cuales en su conjunto conforman el Producto Interno Bruto (PIB) por el lado de la producción	Tomando como referencia el año 2001, el Banco de Guatemala, registró 11 actividades económicas, siendo estas: 1) Agricultura, ganadería, caza, silvicultura y pesca; 2) Explotación de minas y canteras; 3) Industrias manufactureras 4) Suministro de electricidad y captación de agua; 5) Construcción 6) Comercio al por mayor y menor; 7) Trasporte, almacenamiento y comunicaciones; 8) Intermediación financiera, seguros y actividades auxiliares; 9) Alquiler de vivienda 10) Servicios privados y 11) Administración pública y defensa	Tipo numérica, de naturaleza cuantitativa, continua, con escala de razón.

Fuente: elaboración propia.

- Fases del estudio

En este apartado se describen cada una de las 7 fases que se desarrollaron durante el presente estudio, las cuales fueron realizadas de forma sistemática y cronológica.

  - Fase 1: revisión de literatura

La primera fase consistió en la revisión de la información documental y literatura sobre las actividades económicas, el registro de valores en unidades monetarias nacionales a precios reales del año de referencia 2001, el conjunto de las actividades económicas que conformaron el Producto Interno Bruto (PIB) por el lado de la producción, las cuales reflejaron el comportamiento económico del país del 2001 al 2018; asimismo, se consultó la literatura respecto a la seguridad social en Guatemala desde sus inicios, sus avances y su comportamiento en cuanto a la cobertura nacional.

  - Fase 2: recolección de información

La segunda fase correspondió a la obtención de los datos, primero, del registro de los valores monetarios a precios constantes del año de referencia 2001 de las actividades económicas que conformaron el PIB por el lado de la producción, para ello se consultó la página oficial del Banco de Guatemala en el apartado de estadísticas macroeconómicas que se encuentran disponibles de forma oficial al público en general. Asimismo, se recurrió a los informes y estadísticas que publica el Instituto Guatemalteco de Seguridad Social en su sitio de internet, en el cual están contenidos los registros históricos de las afiliaciones a la seguridad social.

- Fase 3: análisis estadístico

En la tercera fase se desarrolló el análisis estadístico de la información obtenida, inicialmente, con el análisis exploratorio de los datos y de estadística descriptiva para aumentar la comprensión de la información y el comportamiento de las variables, tanto para las afiliaciones a la seguridad social, como para el registro de las actividades económicas del país.

Así mismo, se realizó el análisis de regresión múltiple y el análisis de series de tiempo, comprobando los supuestos para ambos tipos de análisis, lo que dio paso para la realización de pronósticos de corto plazo (períodos menores a un año).

Las pruebas y cálculos realizados se ejecutaron en *Rstudio, Minitab, Jamovi, Infostat* y en *Statistical Package for the Social Sciences* (*SPSS)*.

- Fase 4: interpretación de resultados

Sobre los diversos cálculos y análisis estadísticos realizados, la cuarta fase consistió en la interpretación de los resultados de acuerdos a las pruebas estadísticas de descripción, correlación y análisis de series de tiempo.

Mediante la estimación de modelos de mejor ajuste, se conoció el grado de asociación y de dependencia estadística de las afiliaciones a la seguridad social respecto a las actividades económicas del país, se conocieron los resultados de los pronósticos estimados y se comparó la información pronosticada con los resultados reales registrados en los trimestres inmediatos posteriores al estudio.

- Fase 5: análisis de los resultados

La quinta fase constituyó el análisis de los resultados, los cuales versaron desde la información que otorgaron las variables en cada modelo de regresión múltiple, los análisis de los coeficientes de *Cp de Mallows*, la inflación de la varianza al extraer las variables que presentaron cierto nivel de multicolinealidad; así como, la interpretación de los coeficientes de correlación y determinación. Posteriormente, se realizó el análisis de los modelos de pronósticos realizados, comparando efectivamente, las estimaciones sobre las cifras reales de cada variable; se pudo observar que, diversos modelos de pronósticos otorgaban valores cercanos a la realidad registrada durante los trimestres del año 2019; en ese sentido, se atendieron los objetivos propuestos de la investigación.

- Fase 6: redacción de informe final

En la a sexta fase se elaboró el informe final, juntamente con el apoyo del profesional asesor del trabajo de investigación, se discutieron a detalle los resultados obtenidos conforme a los modelos de mejor ajuste que explicaron el grado de dependencia y asociación estadística de las afiliaciones a le seguridad social respecto a los sectores productivos de Guatemala conforme a los pronósticos estimados.

- Fase 7: redacción de conclusiones y recomendaciones

En la séptima y última fase, se elaboraron las conclusiones y recomendaciones del estudio con base en los objetivos descritos en el trabajo de investigación.

INTRODUCCIÓN

La investigación propuesta trata de una innovación en el ámbito de la seguridad social y su relación estadística con las actividades económicas que conforman el Producto Interno Bruto (PIB) por el origen de la producción.

Se desarrolla el análisis de regresión múltiple y de series de tiempo estimando predicciones y pronósticos de corto plazo (períodos menores a un año) con las observaciones de las variaciones inter trimestrales de las afiliaciones a la seguridad social y las variaciones inter trimestrales de las actividades económicas de Guatemala de 2001 a 2018.

Según la ONU (2008): la seguridad social garantiza:

> La seguridad social garantiza a todas las personas su dignidad humana cuando hacen frente a circunstancias que les privan de su capacidad para ejercer plenamente sus derechos; así mismo, debido a su carácter redistributivo, desempeña un papel importante para reducir y mitigar la pobreza, prevenir la exclusión social y promover la inclusión social. (p. 2)

Sin embargo, Guatemala no ha registrado una cobertura significativa de seguridad social para las mayorías sociales, principalmente a la población económicamente activa (PEA); según los resultados de los censos 2002 y 2018, la cobertura de seguridad social respecto de la PEA (personas en edad entre 15 y 65 años) fue de 15.78 % y 14.46 % respectivamente.

La baja cobertura se debe a diversos factores, desde la alta presencia de empleo informal en donde las empresas según sus características no cotizan al seguro social, hasta la evasión de pago y registro real de los trabajadores de las empresas que si lo hacen.

No obstante, Guatemala ha registrado un crecimiento económico expresado por el Producto Interno Bruto (PIB) a precios constantes, significativamente bueno, el ritmo sostenido promedio ha sido de 3.5 % del 2001 al 2018. (BANGUAT, 2019)

Para el caso de Guatemala, el crecimiento económico, no se ha traducido en beneficio de la mayoría de la población en lo relativo a la seguridad social, pues el hecho que se registre crecimiento económico no ha significado crecimiento de la cobertura al régimen de seguridad social, considerando que, son las personas como fuerza de trabajo las que contribuyen al crecimiento económico en cada actividad productiva.

Es de mucha importancia comprender la dinámica de la cobertura a la seguridad social de la población guatemalteca, el grado de asociación y dependencia estadística respecto de las actividades económicas que conforman el PIB por el lado de la producción, dado que, las políticas públicas en función de mejorar las condiciones de vida de la población guatemalteca deben sustentarse con información basada en evidencia científica.

En ese orden de ideas, el conocer las fluctuaciones de las afiliaciones a la seguridad social ante variaciones de las actividades económicas del país, sumado a la estimación de pronósticos de corto plazo, mediante el análisis de series de tiempo, se cuenta con información adicional y oportuna que sirva de referencia Guatemala autoridades del Instituto Guatemalteco de Seguridad

Social, del Gobierno Central y analistas en temas de seguridad social y crecimiento económico.

La presente investigación se realizó en 7 fases organizadas sistemáticamente mediante el uso de la estadística descriptiva e inferencial, se comprobaron los supuestos según los requerimientos teóricos del análisis de regresión múltiple y de series de tiempo; por último, se estimaron pronósticos de corto plazo, los cuales fueron verificados con los resultados reales de los trimestres inmediatos posteriores al período de estudio.

La investigación fue posible de acuerdo con la estimación de factibilidad del autor, la cual contó con: un estudiante investigador, un asesor profesional, recursos tecnológicos y financieros que fueron cubiertos en su totalidad por el estudiante investigador.

El presente estudio se encuentra estructurado en cuatro capítulos generales los cuales se describen a continuación.

En el primer capítulo, se da a conocer el marco referencial, que fue de utilidad para conocer las experiencias de estudios previos sobre seguridad social, crecimiento económico, análisis de regresión múltiple y series de tiempo.

En el segundo capítulo, se presenta el marco teórico que describe la teoría sobre el análisis de regresión múltiple, las series de tiempo y sobre la seguridad social en Guatemala.

El tercer capítulo, incluye la presentación de resultados de los cuatro objetivos específicos y del objetivo general, donde se estima el grado de fuerza y asociación lineal entre las afiliaciones a la seguridad social y las actividades

económicas de Guatemala, la comprobación de supuestos, el modelo de mejor ajuste de predicción y los diferentes pronósticos estimados y comparados con los trimestres inmediatos posteriores al estudio.

Como cuarto capítulo, se presenta la discusión de resultados, el análisis interno y externo, las conclusiones y recomendaciones con base en los diferentes diagnósticos y validaciones internas a cada modelo propuesto.

1. MARCO REFERENCIAL

A continuación, se describen estudios y consultas realizadas de trabajos previos en el marco del análisis de regresión lineal como técnica multivariada ampliamente utilizada para realizar predicciones; así como, el análisis de series de tiempo para realizar pronósticos.

1.1. Análisis de regresión múltiple

Para poder establecer la relación de dependencia y correlación de dos o más variables existen diversas técnicas enmarcadas dentro de la familia de análisis multivariado, para el caso que ocupa el presente estudio, se eligió, el análisis de regresión múltiple.

En cuanto a los modelos de regresión, Madrigal (2014) opina que, pueden ser preferidos para establecer pronósticos o predicciones entre unas variables respecto de otras, debido a su sencillez o por costumbre, principalmente en el área económica; asimismo, indica que la regresión pueda darse entre una variable respecto de otra (regresión simple) o teniendo en cuenta más de una variable explicativa o independiente (regresión múltiple).

Respecto al cumplimiento de los supuestos, Meneses (2019) menciona que, el cumplimiento de los supuestos es un requerimiento importante para la aplicación de las técnicas del análisis multivariante. Destaca de que el análisis está basado en la interpretación de las asociaciones observadas entre las variables; por ello, el presente estudio, esta información ayudó a establecer teóricamente la estimación del grado de asociación entre las variables

seleccionadas, para cumplir con uno de los requerimientos de este tipo de análisis fue contar con una variable dependiente que sea numérica o cuantitativa.

Posterior a establecer los modelos de predicción, Rosas et, al (2006), en su estudio de Modelos de regresión múltiple en presencia de variables cuantitativas y cualitativas, menciona la importancia de comparar los criterios de selección; como algunos más usados, menciona, el cuadrado medio del error (CME), el coeficiente de determinación (R^2), la suma de cuadrados de predicción (PRESS) y los estadísticos de *Cp de Mallows.*

Rosas continua mencionando que, puede establecerse la significancia del modelo a través de la prueba F, una prueba que muestra la bondad de ajuste, se describen las hipótesis en donde los betas son iguales a cero o al menos un beta es distinto de cero; para el primer caso, un modelo que compruebe que $\mathrm{H0}: \beta = 0$, constituiría un modelo no apropiado para realizar predicciones; sin embargo, si se comprueba que, al menos un beta es distinto de cero ($\mathrm{H0}: \beta \neq 0$), el modelo constituiría un buen modelo para predecir.

Así mismo, una alternativa a la prueba F sería una prueba t de dos colas, es decir $t_1 - \alpha/2$ a un nivel de significancia con n grados de libertad; para este caso, el rechazo de H0, se da cuando la t calculada es igual o mayor a $t_1 - \alpha/2$.

Sobre los criterios de selección, Guerra, et al (2003) en su artículo de criterios para la selección de modelos estadísticos en la investigación científica, menciona la importancia de los aspectos que deben tomarse en cuenta para la selección de modelos estadísticos explicativos y dinámicos considerando 14 criterios para decidir el que mejor se ajuste en los datos.

En esa selección de criterios, menciona la importancia de los métodos de ajuste del modelo, el error estándar de los estimadores, los coeficientes de determinación R^2 y R^2 ajustado, el cuadrado medio residual, el diagnóstico y tratamiento de la multicolinealidad, la validación de las validaciones del modelo, los estadísticos de *Cp de Mallows,* el análisis de los residuales, las pruebas de normalidad, de autocorrelación, de homoscedasticidad, entre otros.

De esa cuenta, menciona que, el método de mínimos cuadrados ordinarios tiene amplia aplicación práctica y, en la actualidad, este tipo de modelos está incorporado en distintos paquetes y software estadísticos, los cuales pueden utilizarse siempre y cuando se comprueben los supuestos estadísticos para su aplicación.

Sobre el problema de multicolinealidad, Del Valle et, al (2012) en su artículo La Multicolinealidad en modelos de Regresión Lineal Múltiple menciona que, al aplicar la regresión mediante el método de mínimos cuadrados ordinarios, se obtienen los estimadores del parámetro, los errores estándar de los estimadores, la prueba t de los parámetros, la prueba F sobre el modelo (ANOVA), el coeficiente de determinación el error cuadrado medio y el error de estimación.

Estos estimadores son comparables entre sí, al aplicar distintos modelos y la selección del mejor modelo, dependerá de los criterios de selección tanto de los parámetros como de los estadígrafos obtenidos, algo interesante de este artículo es la comparación de los coeficientes de variación entre diferentes modelos, que ayudan a elegir aquel que muestre la menor variación posible.

Es importante el cumplimiento de los criterios, según Abuín, J. R. (2007) en su estudio de Regresión lineal múltiple, menciona la importancia que deben

cumplirse al utilizar este tipo de modelos, entre ellos indica: que las variables tengan sentido numérico, que no exista variables redundantes o repetidas, las variables elegidas deben de tener justificación teórica, la relación entre variables del modelo y el número de casos, debe ser 1 a 10 y la variable dependiente debe tener relación lineal con las variables explicativas.

De esa cuenta, al seleccionar el modelo, los coeficientes son elegidos de tal forma que la suma de cuadrados entre las observaciones y las predicciones sea mínima; es decir, que la varianza residual se minimiza.

El modelo de regresión múltiple propuesto por Abuín (2007) es el siguiente: $y = b_0 + b_1 * x_1 + b_2 * x_2 + + b_k * x_k + u$.

Donde los coeficientes de b indican el incremento de la variable dependiente (y) por el incremento de la unidad de cada variable explicativa (x), manteniéndose el resto de las variables constantes (*Caeterīs pāribus*).

Así mismo, el autor muestra los supuestos a lo que él llama consideraciones que deben seguirse; tales como: linealidad, homoscedasticidad (las perturbaciones tienen la misma varianza), independencia (las perturbaciones aleatorias son independientes entre sí), normalidad (las distribuciones de las perturbaciones siguen una distribución teórica) y, las variables x_k se obtienen sin errores en la media.

Por último, el autor establece los contrastes de la regresión, o también llamada prueba de bondad de ajuste (prueba F) del modelo, en donde se indica el planteamiento de hipótesis en relación con las betas, la H0; es que todas las betas son nulas = y la H1, alternativa o complementaria: es que existe al menos una beta distinta de cero (0).

1.2. Series de tiempo

Sobre las series temporales, Bracamontes (2020) menciona la importancia de dichas series, en lo que respecta al registro de datos de la seguridad social los cuales son de importante ayuda para los tomadores de decisiones, el autor desarrolló una serie de pasos para la elaboración de las series temporales, pasos que van desde acceder a los datos e identificar las variables, construir y aplicar filtros necesarios, fusionar tablas de datos hasta realizar los cálculos necesarios para el proceso de preparación de análisis de los datos recabados.

Así mismo, Madrigal (2011) expone la necesidad de realizar pronósticos, realizado con las series temporales para tomar decisiones tácticas, estratégicas con ayuda de la planificación eficiente y eficaz, de esa cuenta, los pronósticos son de mucha importancia para los tomadores de decisiones según la disciplina en la que se desenvuelvan.

Respecto a los problemas que puedan presentarse en series de tiempo, Antunez (2011) menciona que, es importante corregir los problemas como la no estacionariedad en la serie, estacionalidad, la presencia de raíz unitaria, la autocorrelación, entre otras; de esa cuenta, muestra la importancia de descomponer una serie temporal en sus elementos: estacionalidad, tendencia, componente irregular y ciclo.

Para los estudio de aplicación, Páez et, al (2016) utiliza las series de tiempo en la realización de pronósticos de producción, hace uso del análisis de series de tiempo para pronosticar la producción de pozos petroleros, describiendo las componentes cíclicas, estacionales e irregulares, haciendo estimaciones de modelos de medias móviles, ARMA y ARIMA, realizando a su vez, comparaciones entro los coeficientes de cada modelo, tales como: el R^2

ajustado, el error cuadrado medio (MSE), los criterios de información de *Akaike* y *Bayesiano* y el error absoluto porcentual.

Concluyendo en su estudio menciona que, la estimación de pronósticos mediante series de tiempo reduce su efectividad en el largo plazo; no obstante, el análisis constituye una herramienta con mayores precisiones que otras técnicas tradicionales.

Dada la información de los estudios anteriores, se consideró importante y oportuna la aplicación del análisis de regresión múltiple y de series de tiempo a las actividades económicas que conforman el PIB por el origen de la producción y las afiliaciones a la seguridad social en Guatemala, durante el período 2001-2018; la primera, para establecer modelos que indiquen el comportamiento de la seguridad social respecto a las variaciones de las actividades económicas y, la segunda, para estimar pronósticos de corto plazo teniendo como observancia principal la comprobación de los supuestos teóricos, el diagnóstico y validación de cada modelo estimado.

2. MARCO TEÓRICO

2.1. Series de tiempo

Gujarati (2001) indica que una serie de tiempo es un conjunto de observaciones sobre los valores que toma una variable en diferentes momentos del tiempo. Las observaciones pueden darse en momentos específicos guardando distancias iguales (equidistantes), las mediciones pueden ser: por segundo, minuto, hora, día, semana, mes, trimestre, semestre, entre otros. (p. 22)

Anderson et al. (2008) establece que, el objetivo del análisis de series de tiempo es realizar pronósticos de valores futuros basados en series históricas (p. 798).

Una de las características fundamentales de las series de tiempo es que, las observaciones sucesivas no son independientes entre sí, es decir, las pruebas estadísticas de independencia no son aplicables a las series de tiempo, dado que, los datos observados dependen de la misma serie en tiempos pasados.

Gujarati (2001) establece que, toda serie de tiempo puede ser generada por un proceso estocástico o aleatorio (p.740). En una serie temporal se busca que un proceso estocástico sea estacionario (p.737), lo anterior se cumple (proceso estocástico estacionario) cuando la media de las observaciones analizadas y su varianza, se mantienen en el tiempo (constantes), a su vez, el valor de las covarianzas entre dos períodos depende solamente de la distancia

entre esos dos períodos y no del tiempo en el cual se ha calculado, el hecho de que la serie sea estacionaria es una de las bases fundamentales que permite realizar pronósticos.

Para explicar mejor el concepto anterior, Gujarati (2001) propone las siguientes propiedades:

- Media: $E(Y_t) = u$ (Ec.1)
- Varianza: $var(Y_t) = E(Y_t - u)2 = \sigma^2$ (Ec.2)
- Covarianza: $yk = E[Y_t - u)(Y_{t+k} - u)]$ (Ec.3)

Toda serie temporal, tiene diversos componentes que son necesarios analizar y determinar previo a trabajar con ellas.

2.1.1. Componente de una serie temporal

En cuanto a los componentes de la serie temporal, Anderson et, al (2008) establece los siguientes:

- Componente tendencial
- Componente cíclico
- Componente estacional
- Componente irregular

Componente tendencial: se le denomina componente de tendencia al cambio gradual o desplazamiento de observaciones a lo largo del tiempo, este puede ser en sentido ascendente (tendencia positiva) o descendente (tendencia negativa).

Componente cíclico: consiste a desplazamientos superiores o inferiores a la tendencia, se considera un ciclo cuyas duraciones sean mayores a un año. (Anderson et, al. 2008)

Componente estacional: corresponde al proceder de los datos en un tiempo corto, regularmente menor a un año, los datos son influenciados por las estaciones del año, es decir, comportamientos con un patrón observable en tiempos específicos.

Componente irregular: son aquellos factores impredecibles, constituyen los factores residuales, dan cuenta de las desviaciones de los valores reales de la serie de tiempo de los valores que se esperan al considerar los efectos de los componentes de tendencia, cíclicos y estacionales. Anderson et, al (2008)

Las series de tiempo pueden ser utilizadas y aplicadas en distintos campos de estudio como la economía, comercio, demografía, medio ambiente, experimentos científicos, medicina, entre otros.

Según Mauricio (2007), “el principal objetivo de las series temporales consiste en elaborar un modelo estadístico que describa adecuadamente la procedencia de dicha serie” (p.8), posterior a elaborar el modelo, la serie puede utilizarse para:

- Describir la evolución de la serie

- Prever la evolución futura

- Contrastar alguna teoría existente sobre las características de las variables

2.1.2. Pruebas de estacionariedad

Como se mencionó anteriormente, previo a trabajar con una serie temporal, es necesario comprobar si los datos tienen comportamiento estacionario; para ello, se pueden realizar dos pruebas estadísticas que ayudan a conocer ese extremo.

Según Gujarati (2001), “la prueba de estacionariedad basada en un correlograma, se realiza en función de autocorrelación” (p.749).

(ACF) denotada bajo la siguiente expresión:

$$pk = \frac{yk}{y0} = \frac{covorianza\ al\ rezago\ k}{varianza} \qquad \text{(Ec.4)}$$

Donde: pk es un número adimensional que se encuentra entre -1 y 1.

Al graficar pk frente a k, se obtiene una gráfica de correlograma. (Gujarati, 2001)

La segunda prueba para conocer si una serie es estacionaria, es la prueba de raíz unitaria para ello es necesario considerar el siguiente modelo:

$$Y_t = Y_{t-1} + u_t \qquad \text{(Ec.5)}$$

Donde: u_t = representa el error estocástico, denominado error de ruido blanco (media cero varianza constante $\sigma 2$).

Al determinar el coeficiente Y_{t-1}= 1, se menciona que los datos tienen problemas de raíz unitaria; razón por la cual, la serie es no estacionaria también conocida como, caminata aleatoria. (Gujarati, 2001)

"Los términos de no estacionariedad, caminata aleatoria, raíz unitaria y tendencia estocástica constituyen sinónimos" (Gujarati, 2010, p. 702)

Mauricio (2007) señala que, "los modelos de series temporales que tiene raíces unitarias pueden mostrar resultados erróneos al realizar algún tipo de inferencia con ellos" (p.211).

No obstante, si los valores encontrados del coeficiente Y_{t-1}<1, se dice que, la serie es estacionaria en sentido puro.

2.1.3. Tipos de tendencia

- Tendencia determinística: si la serie es completamente predecible (lo que se espera para realizar pronósticos)

- Tendencia estocástica: no predecible

2.1.4. Prueba de Primera diferencia

Según Gujarati (2001), "sí, a través de la prueba de estacionariedad se ha establecido que la serie es no estacionaria, es posible convertir la serie a través de la prueba de diferencias d" (p. 420).

- $Y_t - Y_{t-1} = \beta_2(x_t - x_{t-1}) + (u_t - u_{t-1})$ (Ec.6)

- $Yt = \beta_2(x_t - x_{t-1}) + \varepsilon_t$ (Ec.7)

- $\Delta Y_t = \beta_2 \Delta\, x_t + \varepsilon_t$ (Ec.8)

Donde: Δ, delta es el operador de la primera diferencia al cual le corresponde el símbolo u para las diferencias consecutivas de dos valores dados.

2.1.5. Pruebas de Estacionalidad

Sobre la estacionalidad, González (2009) enseña que, la estacionalidad de una serie es un movimiento cíclico en un año que se repite sistemáticamente, la estacionalidad está estrechamente relacionada con la tendencia de las siguientes maneras:

- Si la amplitud del componente estacional se mantiene estable a lo largo del tiempo, la variación observada se sumaría a la línea de tendencia de forma aditiva, en esta forma la estacionalidad se representa por media igual a cero a lo largo del año.

- Si la amplitud crece según crece la tendencia, entonces se dice que lo hace de forma multiplicativa, en ese caso, el promedio anual de los factores estacionales ha de ser igual a uno (1).

2.1.6. Medias Móviles

Para desestacionalizar una serie, existen diferentes métodos, uno de los más reconocidos, las medias móviles, el cual consiste en extraer la tendencia de la serie. (González, 2009)

- Estimar una tendencia local, es necesario tener presente la temporalidad de las observaciones, es decir, si los datos a evaluar son mensuales, se debe considerar una media móvil de orden 12, ejemplo:

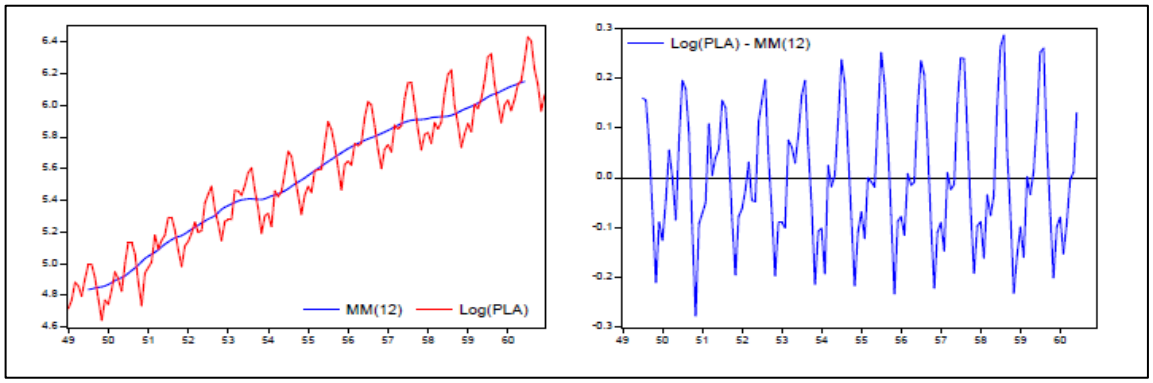
$$Yt - 6 + \cdots + Yt - 1 + Yt + Yt + 1 + \cdots + Yt + 5 \,/12 \text{ (Ec. 9)}$$

Calcular los índices de los datos originales de acuerdo con las medias móviles centradas:

$$\frac{Y_t}{M_t^{12}} \; t = 7,8, \ldots, T - 6 \qquad \text{(Ec. 10)}$$

La Figura 1, muestra el ejemplo de una serie estimando la tendencia y por el otro lado eliminando la misma.

Figura 1. **Desestacionalización de una serie por medias móviles**



Fuente: González (2009). *Técnicas de predicción económica.*

2.1.7. Método de Holt-Winters

Extraer la estacionalidad mediante este método, consiste en aplicar tres ecuaciones, "las cuales suavizan un factor asociado a cada uno de los componentes de la serie: aleatoriedad, tendencia y estacionalidad" (González, 2009, p. 65). Para este método es posible considerar el efecto aditivo y multiplicativo de la serie.

En el modelo aditivo se considera la siguiente expresión:

$$m_t = \alpha\,(Y_t - S_{t-S}) + (1-\alpha)\,[m_{t-1} + b_{t-1}] \qquad \text{(Ec.11)}$$

$$b_t = \beta\,(m_t - m_{t-1}) + (1-\beta)\,b_{t-1} \qquad \text{(Ec. 12)}$$

$$S_t = \gamma\,(Y_t - m_t) + (1-\gamma)\,S_{t-S} \qquad \text{(Ec. 13)}$$

En lo que respecta al método *Holt-Winters*, González (2009) establece que, es un método intuitivamente muy atractivo dado que la ecuación de la

actualización del factor estacional S_t se puede comparar con el índice estacional, en el cual la interpretación puede definirse primeramente, como la diferencia entre el valor actual de la serie y el valor suavizado, si dicha diferencia es mayor, lo demás tendrá positividad y si fuera un valor menor, el resto será negativo (inferior a cero); sin embargo, es importante considerar que el método de alisamiento posee tendencia más no estacionalidad.

2.1.8. Pronósticos

Sobre pronósticos, Andrés Carrión (2001) expone al inicio de su obra que, el hombre de alguna manera siempre ha querido conocer el futuro, para ello se han realizado diversas acciones dados los elementos disponibles en cada época, desde métodos poco racionales próximos a la magia y a la religión hasta métodos más racionalistas, realizado con la matemática y estadística.

Cualesquiera que sean los casos, es importante mencionar que para conocer el futuro es necesario conocer el pasado, Nicolás Maquiavelo en el siglo XIV expuso acertadamente que para predecir lo que ha de suceder, hay que observar antes lo que ha ocurrido anteriormente. (Carrión, 2001)

2.1.9. Modelos ARIMA

ARIMA corresponde a la definición de proceso autorregresivo integrado por media móvil, en el cual se basan los supuestos de las series de tiempo donde se consideran la estacionariedad de la serie en sentido débil, en la práctica, principalmente en las series económicas, las series son no estacionarias, es por ello que diversos autores, recomiendan aplicar el método de las diferencias, para convertir las series en estacionarias y posteriormente aplicar el modelo *ARMA* en

donde AR corresponde a los modelos autorregresivos que se basan en un orden los cuales dependen de la periodicidad en la series observadas. (Gujarati, 2001) y (Aceves, 2003)

Un modelo AR de orden uno (1) puede ser representado con la siguiente expresión:

$$\mathrm{Y\,t} = \varphi 0 + \varphi 1 \mathrm{Yt} - 1 + \mathrm{a} \qquad \text{(Ec. 14)}$$

Mientras que un modelo AR de orden (p) se expresa con la siguiente ecuación:

$$\mathrm{Y\,t} = \varphi 0 + \varphi 1 \mathrm{Y\,t} - 1 + \varphi 2 \mathrm{Yt} - 2 + \ldots\ldots + \varphi\ \mathrm{pYt} - \ \mathrm{p} + \mathrm{at} \qquad \text{(Ec. 15)}$$

Donde: $\varphi \mathrm{p}(\mathrm{L})$ es conocido como el operador de retardo polinomial. (Aceves, 2003)

Sobre las medias móviles, Gujarati (2001) menciona que MA(q) es aquel cuyo valor que explica una variable en período t en función de un término independiente y que corresponde a una sucesión de errores de períodos previos, los cuales pueden ser ponderados según la temporalidad de las observaciones, suponga la siguiente ecuación:

$$Y_t = \mu + \beta 0 \mu_t + \ \beta 1 \mu_{t-1} \qquad \text{(Ec.16)}$$

Dónde: μ es una constante $Y\mu$ corresponde al error estocástico de ruido blanco, según esta expresión Y en el período t es igual a la constante más un promedio móvil en términos de error presente y pasado, para el caso del presente, entonces se dice que, Y sigue una condición de media móvil de primer

orden MA (1), en resumen, se puede indicar que un proceso de medio móvil es una combinación lineal en termino de error de ruido blanco. (Gujarati, 2001)

2.1.10. Metodología de *Box-Jenkins* (BJ)

Para esta metodología, Gujarati (2001) enseña que, para utilizar el método BJ es necesario hacer uso de cuatro pasos que ayudarán a comprender los valores de $p, q\ y\ d$ que representan al modelo *ARIMA*.

Paso 1. Identificación, en este paso se determinan los valores de $p, q\ y\ d$ para encontrar estos valores es muy útil utilizar los gráficos de correlogramas y correlogramas parciales.

Paso 2. Estimación, posterior a determinar los valores de $p, q\ y\ d$ apropiadamente, corresponde terminar los parámetros autorregresivos y de media móvil dentro del modelo.

Paso 3. Verificación de diagnóstico, luego de estimar el modelo ARIMA y sus parámetros, es necesario verificar si los valores se ajustan a los datos de forma razonable, esto se debe a que puede existir otro modelo ARIMA que también lo haga, el autor opina que, la estimación de modelos ARIMA constituye un arte, para ello deben de realizarse diversas pruebas seleccionando los residuales que mejor ruido blanco presenten.

Paso 4. Predicción, la popularidad de utilizar los modelos ARIMA se debe al éxito de las predicciones obtenidas, por ellos es necesario comprobar los resultados obtenidos con información de las variables evaluadas que tengan disponibilidad.

2.2. Modelos de regresión

Gujarati (2001) menciona:

> El análisis de regresión trata del estudio de la dependencia de la variable dependiente, en una o más variables; las variables explicativas, con el objeto de estimar y/o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las últimas (p.16).

A pesar de que, las variables dependientes están en función de las variables predictoras o independientes, sin importar que tan fuerte sea el grado de dependencia o asociación entre ellas, en ningún momento se puede establecer causalidad entre ambas. (Gujarati, 2003)

Es importante mencionar que, la regresión y correlación son términos que suelen confundirse; sin embargo, conceptualmente corresponde a análisis distintos, el primero se basa en determinar el grado de dependencia estadística de las variables independientes sobre la variable dependiente, para ello se utiliza el "coeficiente de determinación (R^2)", con valores que pueden oscilar entre -1 a 1, mientras que el segundo, "coeficiente de correlación (R)" cuyos valores de la misma manera que R^2 pueden oscilar entre $-1 \leq r \geq 1$, dicho valor muestra el grado de asociación o fuerza entre las variables analizadas. (Gujarati, 2003)

En una regresión múltiple, Walpole et al. (2008) menciona que puede expresarse de la siguiente forma:

$$Y = \beta_0 + \beta_{1*x1} + \beta_{2*x2} \qquad \text{(Ec.17)}$$

Donde: Y representa a la variable dependiente, $X1\ y\ X2$ ... las variables independientes, β_0 la intersección o intercepto, β_1 y $\beta_{2...}$ son los coeficientes de regresión parcial.

2.2.1. Supuestos de un modelo de regresión

Para construir un modelo de predicción a mediante el análisis de regresión múltiple, es necesario comprobar los supuestos estadísticos que permitan evidenciar la correcta utilización de los aspectos técnicos y no caer en errores de predicción.

A continuación, se enumeran los supuestos más importantes sobre este tipo de análisis:

- Supuesto de Multicolinealidad

Se dice que, dos variables explicativas tienen presencia de colinealidad cuando se explican entre sí; es decir, existe una relación lineal perfecta o aproximadamente exacta entre dichas variables. (Gujarati, 2001)

- Supuesto de Heteroscedasticidad

La heteroscedasticidad se da en un modelo cuando todas las perturbaciones μ_i no tienen varianzas σ^2 iguales (homoscedasticidad), representa uno de los supuestos más críticos en los modelos de regresión. (Gujarati, 2001)

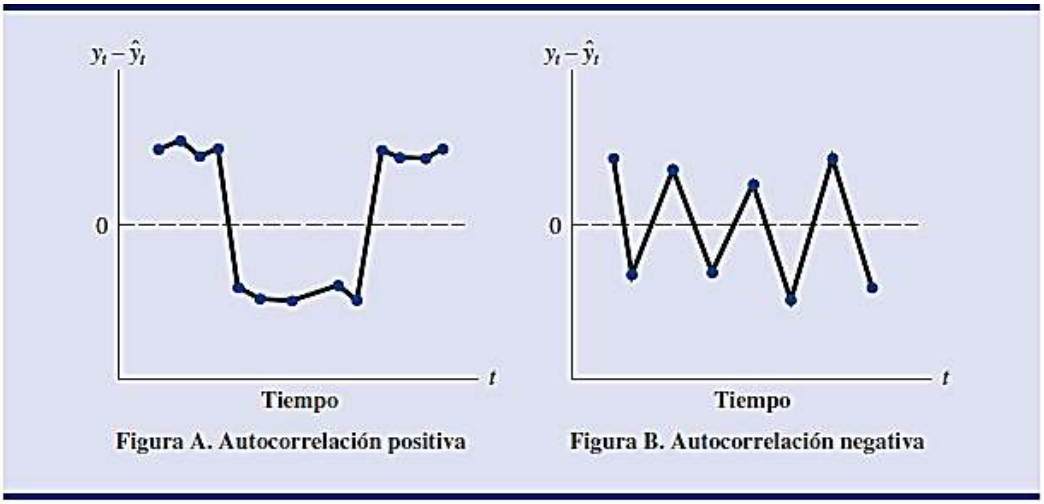
- Supuesto de Autocorrelación

Tintner citado por Gujarati (2001), define la autocorrelación como una serie dada consigo misma, rezagada por un número de unidades de tiempo, cuando se analiza una correlación entre la misma serie; sin embargo, la definición de autocorrelación entre dos series de tiempo distintas, los rezagos de la primera observación son iguales a los rezagos de la última observación entre series, a ello Tintner le denomina autocorrelación serial.

Si existe autocorrelación, Anderson et al. (2008) enseña que se transgrede uno de los supuestos del modelo de regresión, indicando que los términos del error no son independientes entre sí. Las Autocorrelaciones pueden darse con relación a períodos anteriores; es decir, el valor de y en un período t, puede estar relacionado con el valor de y de un período anterior $(t-1)$, en ese caso, se denota una autocorrelación de grado uno (1), si la relación del valor y es con relación a dos períodos $(t-2)$ existe autocorrelación de segundo grado (2), así sucesivamente.

Para detectar autocorrelación, Gujarati (2001) menciona una forma a través del método gráfico, observando el comportamiento de los residuales contrastados frente al tiempo.

Figura 2. **Conjunto de dos datos con correlación de segundo grado**



Fuente: Anderson et al. (2008). *Estadística para administración y economía.*

2.2.2. Pruebas de supuestos

Al describir los supuestos de los modelos de regresión múltiple, es importante mencionar las pruebas para detectar la presencia de multicolinealidad, autocorrelación y heteroscedasticidad.

Para el caso de la multicolinealidad, pueden aplicarse las pruebas de inflación de la varianza (VIF) por sus siglas en inglés, este tipo de prueba permite la determinación de regresar cada columna X (Variable explicativa) sobre las restantes, un R^2 muy alto en una o más regresiones, evidencia una relación lineal aproximada entre la variable dependiente la cual toma como independiente, la ecuación puede expresarse. (Tusell, 2011)

$$VIF(i) = \frac{1}{1-\mathrm{R}^2(\mathrm{i})} \qquad \text{(Ec. 18)}$$

Es importante mencionar el criterio de esta prueba que permite indicar si existe multicolinealidad o no:

Sí los valores VIF (i) encontrados son superiores a 10, se considera que existe presencia de multicolinealidad. (Tusell, 2011)

Para Anderson et al (2008), el caso de la Autocorrelación, se puede utilizar la prueba *Durbin-Watson,* si los valores residuales sucesivos se encuentran muy cercanos unos entre y otros, es decir existe una autocorrelación positiva el valor será pequeño, en caso contrario se encontrarán alejados entre ellos, se dice que, existirá una autocorrelación negativa, para este caso, el estadístico estimado será grande, la prueba *Durbin-Watson* puede expresarse con la siguiente ecuación:

$$d = \frac{\sum_{t=2}^{n}(e_2 - e_{t-1})2}{\sum_{t=1}^{n} e_t^2} \quad \text{(Ec. 19)}$$

De la misma manera que en la prueba de autocorrelación, existen criterios para determinar si una variable tiene presencia de autocorrelación con la prueba *Durbin-Watson.*

Anderson et al. (2008) menciona que, los valores que se estimen en el estadístico pueden oscilar entre 0 a 4, si el valor calculado fuera dos (2), entonces se dice que no existe autocorrelación, si es menor a dos, existe autocorrelación positiva y fuera mayor a dos, existe presencia de autocorrelación negativa.

Para detectar si existe heteroscedasticidad en las variables analizadas, hay distintas pruebas que pueden ayudar a detectar el problema, desde pruebas formales como la prueba *Park,* el método *Goldfeld y Quandt,* la prueba *Glejser y*

las pruebas de correlación por rango, cada una de las pruebas mencionadas tiene sus alcances y desventajas.

Sin embargo, una de las pruebas mayormente utilizadas es la prueba general de heteroscedasticidad de *White,* la cual, a diferencia de las pruebas mencionadas anteriormente, esta prueba necesita ordenar las observaciones respecto a la variable X, un supuesto importante de mencionar se debe a que, la prueba *White* no se apoya en el supuesto de normalidad lo cual también supone cierta facilidad de aplicabilidad. (Gujarati, 2001)

La prueba general de heteroscedasticidad puede expresarse con el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \text{(Ec. 20)}$$

Los criterios según los resultados obtenidos en las pruebas de *White* establecen que, si los valores obtenidos en ji cuadrado no excedente al ji cuadrado crítico a un nivel de significancia dado, se dice que no existe presencia de heteroscedasticidad.

2.2.3. Criterios de comparación

Para poder comparar los estadígrafos entre distintos modelos, que permitan seleccionar variables regresoras que mejor expliquen el comportamiento de la variable dependiente, es posible auxiliarse entre los valores *Cp de Mallows,* criterios de Información de *Akaike* y *Bayesiano.*

2.2.3.1. Cp de Mallows

Rodríguez, A. A. (2005) menciona que, al determinar los estadísticos de Cp, corresponde a los menores valores enteros calculados.

$$C_p = \frac{SCEp}{CME} - (n - 2 * p) \qquad \text{(Ec.21)}$$

2.2.3.2. Criterio de información de *Akaike*

Márquez Cebrián, M. D. (2002) cita a Akaike (1974) indicando que el problema de la identificación de la teoría de la decisión estadística, según la función de perdida de información más adecuada es elegir el mínimo del criterio de información AIC (criterio de especificación). Este criterio es basado en la diferencia entre dos tipos de distribuciones (la distribución observada y la distribución teórica), a partir del logaritmo de máxima verosimilitud (log-verosimilitud).

$$AIC = 2(k - \ln L) \qquad \text{(Ec. 22)}$$

Donde; $2ln$ = máxima verosimilitud + 2 (n = parámetros independientes ajustados).

2.2.3.3. Criterio *Bayesiano* (BIC)

Márquez Cebrián, M. D. (2002) enseña que, el criterio de AIC presenta ciertos problemas de sobre estimación; de esa cuenta, *Akaike* (1978) elabora un nuevo criterio de especificación, en el cual se introduce al AIC una modificación de tipo *bayesiano* de acuerdo con la siguiente expresión:

$$BIC = k * \ln(n) - 2 * \ln L)$$ (Ec. 23)

Donde: $2ln =$ (máxima verosimilitud) + $ln\ (N)\ k\ y\ N$= el número total de observaciones muestrales.

Márquez Cebrián, M. D. (2002) menciona que, la diferencia entre los criterios *AIC* y *BIC* radica en que este último penaliza más los modelos con un número mayor de parámetros estimados en donde se corrige el problema de sobre estimación.

2.2.4. Seguridad Social

La Organización Internacional del Trabajo (OIT) en su portal de internet menciona que, el concepto de Seguridad Social nació en Alemania en el siglo XIX, el cual fue propuesto por el Canciller Otto Von Bismarck, el Canciller propuso un seguro social para la vejez con el fin de promover el bienestar de los trabajadores, el programa que benefició a la población en edad adulta, junto a programas previos de indemnización junto al seguro de enfermedad, brindó al país alemán un sistema completo de seguridad social.

Narro et al. (2010) hace hincapié que, posterior al ejemplo alemán de establecer esquemas de seguridad social basados en aportes de parte de patronos y trabajadores, varios países de la región europea imitaran dicho esquema, entre ellos Inglaterra, en 1946 estableció su Servicio Nacional de Salud. Bajo esta premisa varios países de esa región tuvieron en cuenta la posibilidad de que los Estados diseñaran mecanismos de seguridad social para la población más necesitada.

Respecto a las necesidades fundamentales de los ciudadanos, Mesa-Lago et al. (1997) opina que, la seguridad social puede está presente en múltiples formas en los procesos de desarrollo de las sociedades.

En ese sentido, se establecen ciertos principios que deben de ser considerados al definir la seguridad social, por ejemplo: los principios de universalidad y solidaridad. Los principios de solidaridad y universalidad, adaptados y configurados a la luz de la situación local, siguen siendo elementos fundamentales de un sistema de seguridad social, que pretenda solidez y durabilidad social y política. (Mesa-Lago et al.,1997)

Es importante considerar las siguientes definiciones, regularmente suelen confundirse y utilizarse como sinónimos: seguridad social, seguro social, asistencia social y previsión social:

2.2.4.1. Seguro social

Valenzuela (2006) menciona que, “el seguro social es el principal mecanismo del cual la previsión social se vale para cumplir su función aseguradora y previsora, el brazo ejecutor, un instrumento de la seguridad social” (p. 4). Corresponde a la institución que ejerce la función del cumplimiento de la seguridad social.

2.2.4.2. Seguridad social

Ruiz (1997) citado por Valenzuela (2006) menciona que, la seguridad social es:

> La protección que la sociedad proporciona a sus miembros, mediante una serie de medidas públicas, contra las privaciones económicas y sociales que de no ser así ocasionarían la desaparición o una fuerte reducción de los ingresos por causa de enfermedad, maternidad, accidente de trabajo o enfermedad laboral, desempleo, invalidez, vejez y muerte; y también la protección en forma de asistencia médica y de ayuda a las familias con hijos, (p.4).

2.2.4.3. Asistencia Social

La diferencia entre asistencia social y seguridad social radica primordialmente a que, la primera es una forma de beneficencia pública, una ayuda temporal.

Ruiz (1997) citado por Valenzuela (2006) menciona a la asistencia social como:

> El conjunto de normas de todo tipo, que integran una actividad del Estado y en su caso de los particulares, destinadas a procurar una condición lo más digna, decorosa y humana, para aquellas personas que, imposibilitadas para satisfacer por sí mismas sus necesidades elementales y de bienestar social, requieren del socorro y la ayuda altruista, no obligatoria, de los demás (p.2).

2.2.4.4. Previsión social

Ruiz (1997) citado por Valenzuela (2006) expresa que, “la previsión social es útil para temperar o disminuir la inseguridad, así como los males que padecen

los trabajadores vistos como clase social económicamente débil, dentro o fuera del trabajo” (p.3).

Es decir, la previsión social constituye un eslabón importante de la seguridad social. A la previsión social, le corresponde procurar el aprovisionamiento de los recursos necesarios que serán utilizados en el futuro, la cual puede ser traducida como pensiones y subsidios.

2.2.4.5. Seguridad Social en Guatemala

Según la CPRG (1993), en el Artículo 100 establece que, “el Estado reconoce y garantiza el derecho a la seguridad social para beneficio de los habitantes de la Nación. Su régimen se instituye como función pública, en forma nacional, unitaria y obligatoria” (p.19).

El Congreso de la República de Guatemala según el Decreto 295 publicado en el Diario Oficial el 31 de octubre y 4 de noviembre de 1946, decretó la Ley Orgánica del Instituto Guatemalteco de Seguridad Social, institución encargada de aplicar el régimen de seguridad obligatoria.

La autoridad máxima del Instituto Guatemalteco de Seguridad Social (IGSS), se conforma a través de una Junta Directiva, la cual tiene representación directa de 6 representantes titulares y 6 suplentes.

Los órganos que representan la Junta Directiva son: 1) Dos del Poder Ejecutivo del país (nombrado por el presidente de la República); 2) dos representantes de la Junta Monetaria del Banco Central; 3) dos representantes del Consejo Superior Universitario de la Universidad autónoma de San Carlos de Guatemala; 4) dos representes del Colegio

de Médicos y Cirujanos de Guatemala; 5) dos representantes de asociaciones o sindicatos patronales que estén registrados conforme la Ley y, 6) dos representes de los sindicatos de trabajadores que estén registrados conforme. (Decreto 295, 1946, p. 5)

Los órganos superiores del Instituto Guatemalteco de Seguridad Social, luego de la Junta Directiva son:

- La Gerencia del Instituto
- El Consejo Técnico
- La Contraloría General del Instituto
- Las Subgerencias del Instituto (8 en total al año 2022)
    - Subgerencia de Planificación Y Desarrollo
    - Subgerencia Financiera
    - Subgerencia Administrativa
    - Subgerencia de Recursos Humanos
    - Subgerencia de Prestaciones en Salud
    - Subgerencia de Prestaciones Pecuniarias
    - Subgerencia de Integridad y Transparencia Administrativa

- Subgerencia de Tecnología

Sobre la obligatoriedad, los beneficios y las fuentes de financiamiento de la seguridad social en Guatemala se pueden mencionar lo siguiente:

2.2.4.5.1. Campo de aplicación

Todos los habitantes de Guatemala que sean parte activa del proceso de producción de artículos o servicios, están obligados a contribuir al sostenimiento del régimen de Seguridad Social en proporción a sus ingresos y tienen el derecho de recibir beneficios para sí mismos o para sus familiares que dependan económicamente de ellos, en la extensión y calidad de dichos beneficios que sean compatibles con el mínimum de protección que el interés y la estabilidad sociales requieran que se les otorgue. (Decreto 295, 1946, p. 16)

2.2.5. Beneficios de la seguridad social

Según el Artículo 28 del Decreto 295 (1946):

El régimen de Seguridad Social comprende protección y beneficios en caso de que ocurran los siguientes riesgos de carácter social:

a) Accidentes de trabajo y enfermedades profesionales;

b) Maternidad;

c) Enfermedades generales;

d) Invalidez;

e) Orfandad;

f) Viudedad;

g) Vejez;

h) Muerte (gastos de entierro); e

i) Los demás que los reglamentos determinen. (p. 17)

2.2.6. Financiamiento de la seguridad social

La forma de financiamiento de la seguridad social en Guatemala se realiza de forma tripartita donde las aportaciones son cubiertas por los trabajadores, patronos y el Estado de Guatemala, con base a los siguientes porcentajes:

En cuanto a los reglamentos de cuotas y montos, el Artículo 39 del Decreto 295 (1946) establecen que, estos, deben determinar en cada caso, el monto de las cuotas o contribuciones, de acuerdo con el costo total que para los respectivos beneficios establezcan las estimaciones actuariales, así como la manera y momento de cobrar o de percibir aquéllas y el procedimiento o normas que se deben seguir para calcularlas.

Dentro del costo total, quedan comprendidos el pago de prestaciones, los gastos administrativos y la capitalización de obligaciones. En la etapa prevista por el inciso a) del artículo anterior, se deben observar las siguientes reglas.

“Las tres partes deben contribuir a sufragar el costo total de los beneficios que en determinado momento se den, en la siguiente proporción: Trabajadores 25 %; Patronos 50 % y Estado 25 %” (Decreto 295,1946, p. 20).

Para el caso de la contribución que le corresponde como Estado y como patrono, la misma debe ser cubierta con impuestos específicos para el efecto, tal como lo consigna el artículo 40 de la referida Ley:

“La cuota del Estado como tal y como patrono se debe financiar con los impuestos que al efecto se creen o determinen, los cuales han de ser disponibilidades privativas del Instituto” (Decreto 295,1946, p. 21).

Las contribuciones por parte de los trabajadores corresponden el 4.83 % del salario afecto, mientras que a los patronos les corresponde el pago del 10.67 % sobre el mismo salario, los porcentajes financian directamente los programas de Enfermedad, Maternidad y Accidentes -EMA y los de Invalidez, Vejez y Sobrevivencia -IVS-.

2.3. Actividades Económicas

Respecto de las actividades económicas, Black et al. (1997) menciona que se define como toda producción y consumo de bienes y servicios. Ésta puede ser mediante mercados organizados o a través de actividades no comerciales como la producción doméstica. La extensión de la actividad económica en un país se mide por el ingreso nacional y el aumento de la actividad económica se refleja en el crecimiento económico; ambas medidas son incompletas si no incluyen todas las formas de producción doméstica.

Según Sepúlveda (1995):

> La actividad económica es aquella que tiende a incrementar la actividad productiva de bienes y servicios de una economía, para satisfacer las necesidades humanas en un período, involucra los aspectos sociales de tales actividades y también proporciona un cierto nivel de vida a la

población, Por lo tanto, la actividad económica abarca los aspectos de producción, consumo, intercambio y distribución. (p. 19)

Para la formación del Producto Interno Bruto (PIB) en Guatemala, del 2001 al 2018, el Banco Central, encargado del registro de las actividades económicas del país y del cálculo de este, clasificó 11 actividades económicas, las cuales se enumeran a continuación:

- Agricultura, Ganadería, Caza, Silvicultura Y Pesca
- Explotación de Minas y Canteras
- Industrias Manufactureras
- Suministro de Electricidad y Captación De Agua
- Construcción
- Comercio al Por Mayor y Menor
- Trasporte, Almacenamiento y Comunicaciones
- Intermediación Financiera, Seguros y Actividades Auxiliares
- Alquiler de Vivienda
- Servicios Privados

- Administración Pública Y Defensa

2.4. Crecimiento económico, empleo y seguridad social

El crecimiento económico de un país es expresado por el Producto Interno Bruto (PIB). Una de las formas ampliamente utilizada y aceptada para la formación del PIB es por el origen de la producción, la cual está dada por el registro monetario de cada actividad económica. Las actividades económicas de acuerdo con su naturaleza requieren en mayores o menores medidas de los factores de la producción, de acuerdo con la teoría clásica estos han sido la tierra, el trabajo y el capital.

La organización Internacional del Trabajo (OIT) menciona que el crecimiento económico necesita de la fuerza de trabajo intrínsecamente.

El crecimiento económico es un requisito previo para aumentar el empleo productivo; es el resultado combinado de incrementos en el empleo e incrementos en la productividad laboral. Por lo tanto, la tasa de crecimiento económico establece los límites absolutos dentro de los cuales se puede presentar el crecimiento en el empleo y el crecimiento en la productividad laboral.

CEPAL (2000) en su libro reformas económicas, crecimiento y empleo mencionan que, la demanda laboral de las empresas es una variable que a nivel agregado depende en gran parte del crecimiento económico y de las características de las tecnologías aplicadas.

Dado que, las personas afiladas a la seguridad en social en Guatemala dependen totalmente de las empresas legalmente constituidas y que cotizan al

seguro social, la relación teórica de las variables del presente estudio es válida de acuerdo con la teoría existente.

3. PRESENTACIÓN DE RESULTADOS

3.1. Grado de fuerza o asociación y dependencia estadística

Objetivo 1: estimar el grado de fuerza o asociación y de dependencia estadística entre las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala entre los años 2001 y 2018 a través del análisis de correlación y de regresión múltiple.

A continuación, se describen los resultados que corresponden a la fase del análisis exploratorio y descriptivo del comportamiento de las afiliaciones de la seguridad y de las actividades económicas de Guatemala en el período 2001-2018.

3.1.1. Análisis exploratorio de la seguridad social

En lo que respecta al número afiliados a la seguridad social en Guatemala, durante el período de estudio, se obtuvo la serie de datos acumulados trimestralmente del registro que realiza el Instituto Guatemalteco de Seguridad social; sobre las actividades económicas, los valores están expresados en millones de quetzales a precios de referencia del 2001; en ese orden de ideas, se muestran los siguientes resultados:

3.1.1.1. Pruebas de normalidad

A continuación, las pruebas de normalidad *Kolmogorov Smrinov* y *Shapiro Wilk* para cada variable de estudio, las cuales representan las variaciones inter

trimestrales; para el caso de las afiliaciones de la seguridad social, corresponden las personas afiliadas al régimen, mientras que; las actividades económicas, corresponden las variaciones inter trimestrales de los valores monetarios expresados en millones de quetzales.

Tabla II. **Pruebas de normalidad en variables de estudio**

Variable	Kolmogórov-Smirnov		Shapiro-Wilk	
	Estadístico	**Sig.**	**Estadístico**	**Sig.**
Seguridad Social	.153	.000	.947	.005
1. Agricultura, ganadería, caza, silvicultura y pesca	.128	.006	.908	.000
2. Explotación de minas y canteras	.062	.200	.950	.007
3. Industrias manufactureras	.163	.000	.933	.001
4. Suministro de electricidad y captación de agua	.067	.200	.991	.894
5. Construcción	.174	.000	.868	.000
6. Comercio al por mayor y al por menor	.172	.000	.868	.000
7. Transporte, almacenamiento y comunicaciones	.086	.200	.977	.223
8. Intermediación financiera, seguros y actividades auxiliares	.106	.045	.974	.148
9. Alquiler de vivienda	.123	.009	.955	.012
10. Servicios privados	.161	.000	.914	.000
11. Administración pública y defensa	.099	.079	.941	.002

*. Esto es un límite inferior de la significación verdadera.
a. Corrección de significación de Lilliefors

Fuente: elaboración propia, realizado con SPSS 25.

De acuerdo con la tabla II, de las pruebas de normalidad de *Kolmogórov-Smirnov* de las 12 variables de estudio, la variable de minas y canteras, suministro de electricidad, transporte y administración pública, distribuyen de forma normal y de acuerdo con la prueba de *Shapiro-Wilk*, las actividades económicas de suministro de electricidad, transporte e intermediación financiera se ajustan significativamente a una distribución normal.

Tabla III. **Estadísticos descriptivos variables de estudio**

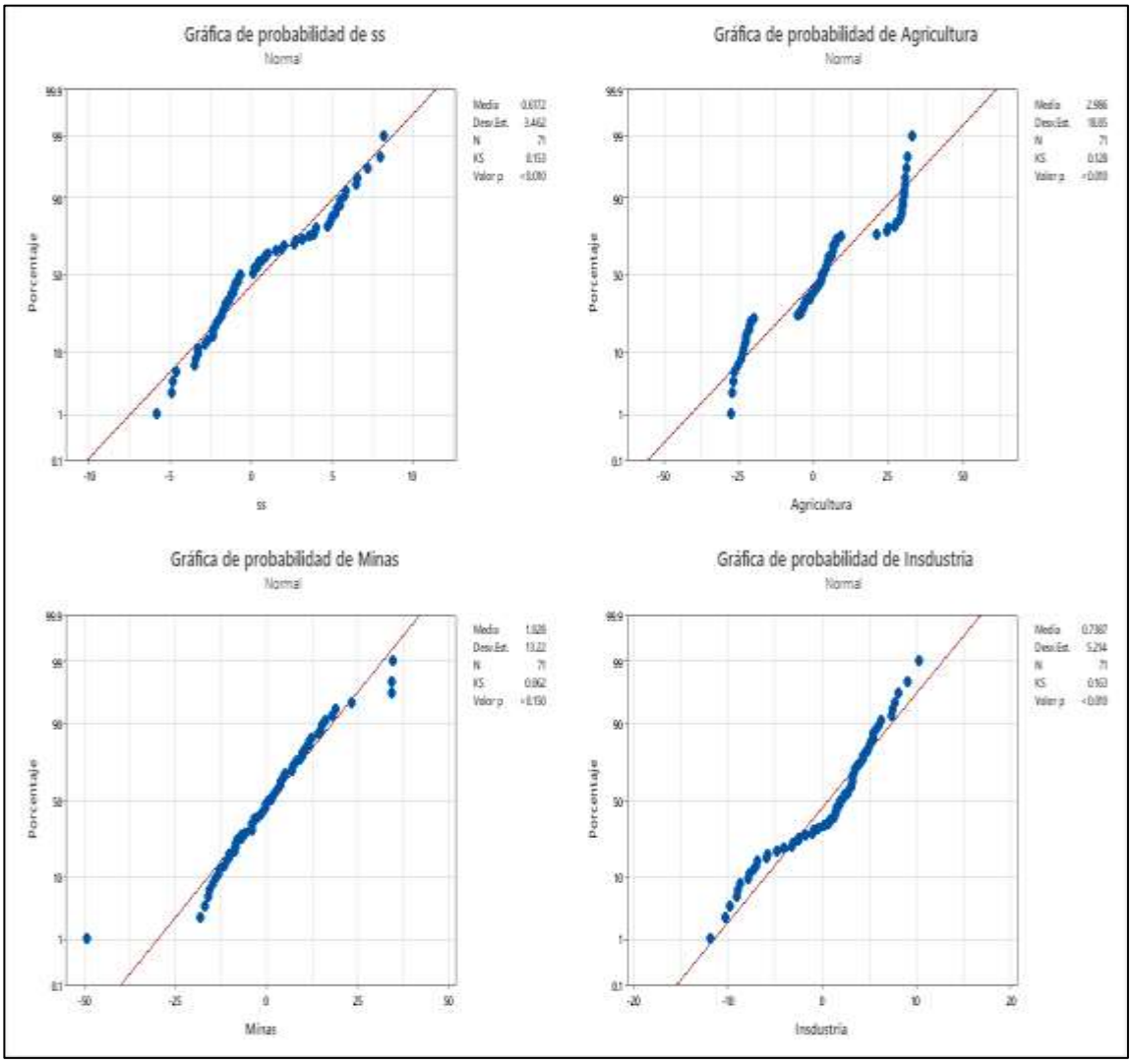
Variable	Media	Mediana	Desv. Desviación	Rango
Seguridad Social	0.6172	-0.6800	3.46215	14.01
1. Agricultura,	2.9861	2.7700	18.85040	60.49
2. minas y canteras	1.0283	1.0400	13.22500	83.71
3. Industrias	0.7387	2.0700	5.21387	22.17
4. electricidad	1.0614	1.5000	3.44069	17.72
5. Construcción	3.3003	8.7100	20.88127	67.43
6. Comercio	1.6568	4.0400	11.63105	40.74
7. Transporte,	2.0121	2.6900	5.24644	24.92
8. Intermediación financiera,	2.3937	1.6400	8.19179	35.58
9. Alquiler de vivienda	0.8032	0.8000	0.27679	2,837
10. Servicios privados	0.9442	1.5900	1,509	5,120
11. Administración pública	1.2479	3,649	883	2,764

Fuente: elaboración propia, realizado con SPSS 25.

En lo que respecta a los estadísticos descriptivos de las variables, la tabla III muestra la media, la mediana, la desviación estándar y el rango intercuartílico para cada una de las variables, las cuales muestran una distribución teórica gaussiana y, por tanto, es correcto expresar el promedio junto a su desviación estándar. Mientras que, las variables que no presentan una distribución teórica, lo correcto es mencionar la mediana junto a su rango intercuartílico para describir su comportamiento.

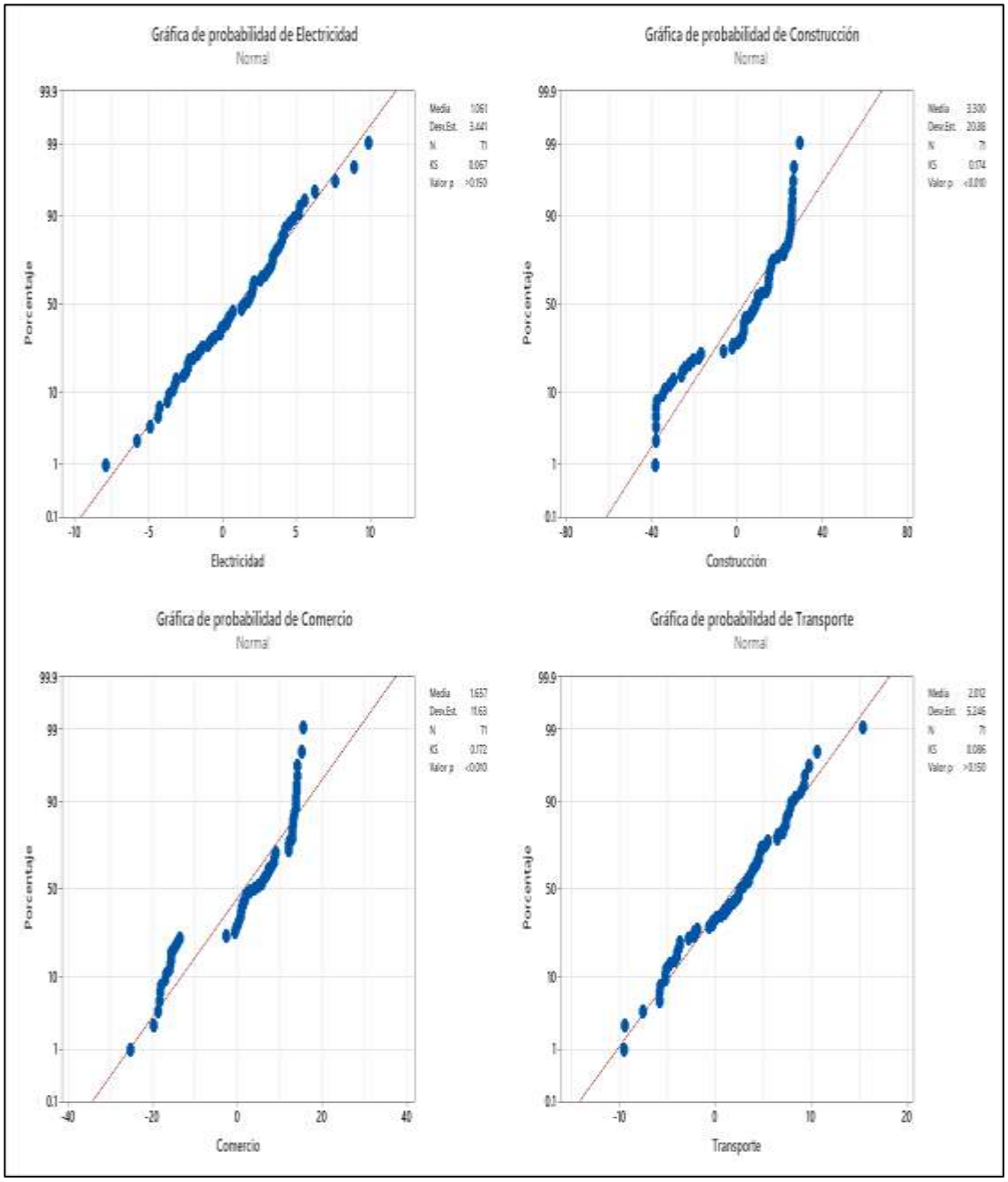
A continuación, se presentan las gráficas de normalidad para cada una de las 12 variables de estudio, mostrando efectivamente, aquellas en las que la distribución de los datos no se da alrededor de una tendencia teórica, por facilidad de visualización gráfica, se muestran 3 grupos de 4 variables cada uno.

Figura 3. **Gráfica de normalidad, seguridad social, agricultura e industria**



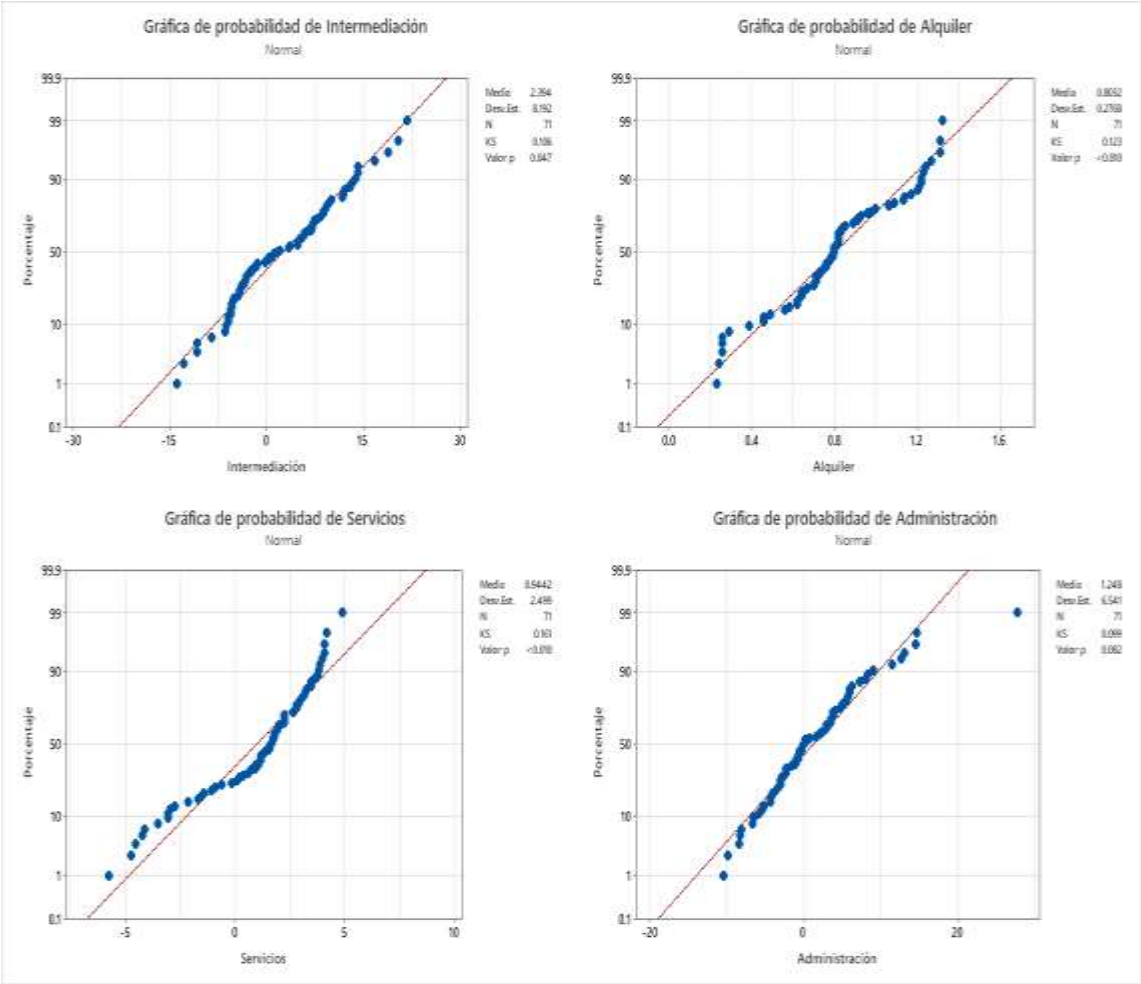
Fuente: elaboración propia, realizado con Minitab 20.

Figura 4. **Gráfica de normalidad de electricidad, construcción, comercio privados y transporte**



Fuente: elaboración propia, realizado con Minitab 20.

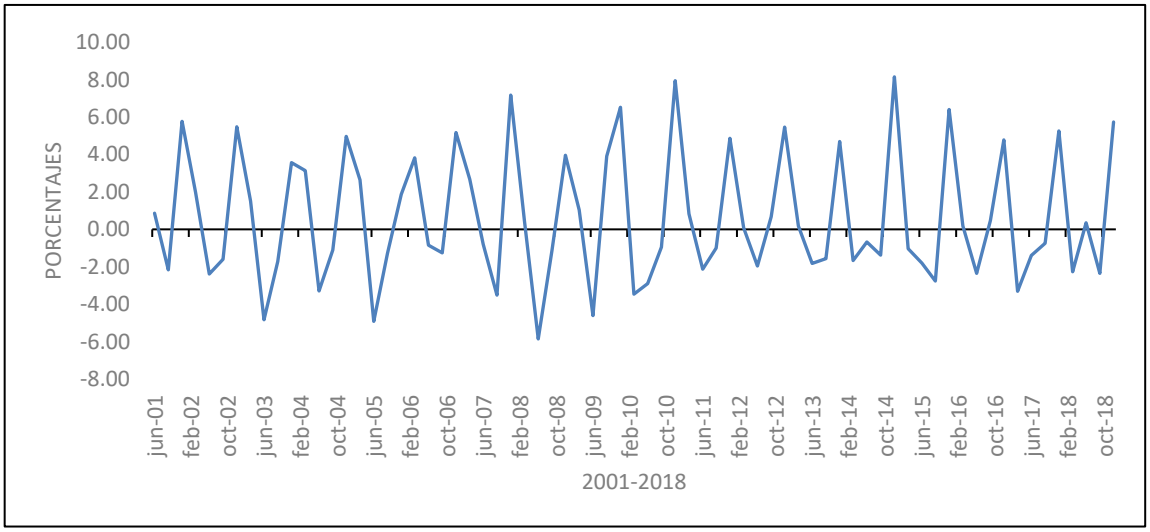
Figura 5. **Gráfica de normalidad de intermediación financiera, alquiler de viviendas, servicios privados y administración pública**



Fuente: elaboración propia, realizado con Minitab 20.

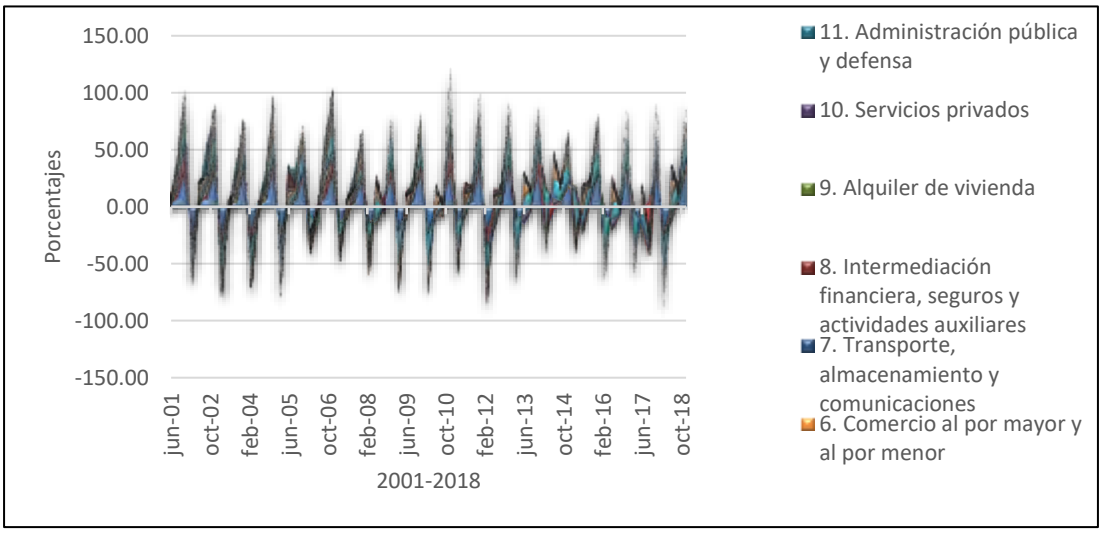
En las figuras 6 y 7 se puede observar el comportamiento a lo largo de los 19 años, tanto de las variaciones inter trimestrales de las personas afiliadas a la seguridad social como el comportamiento de las variaciones inter trimestrales de las actividades económicas de Guatemala que; en su conjunto, conforman el PIB.

Figura 6. **Variaciones inter trimestrales de la seguridad social 2001-2018**



Fuente: elaboración propia, realizado con Microsoft Excel.

Figura 7. **Variación inter trimestral de las actividades económicas 2001-2018**

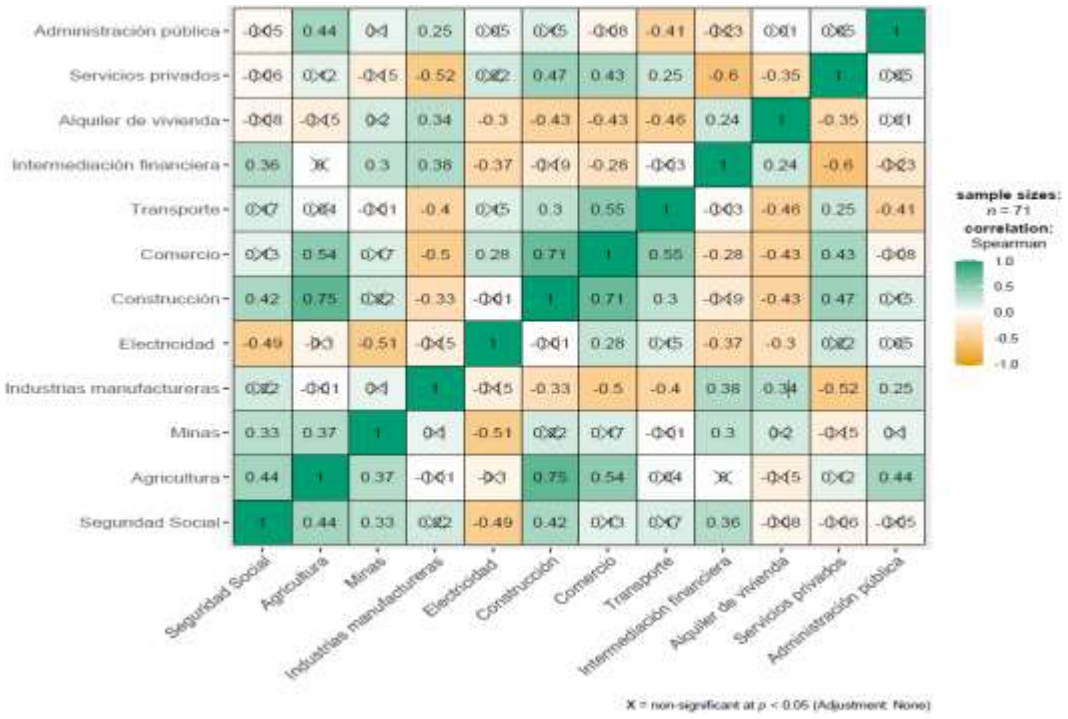


Fuente: elaboración propia, realizado con Microsoft Excel.

3.1.2. Grado de fuerza o asociación lineal

Para poder realizar el análisis de correlación lineal entre las variables de afiliaciones a la seguridad social y las actividades económicas de Guatemala, se procedió a calcular las variaciones inter trimestrales, eliminando el factor de tendencia; de lo contrario, las correlaciones saldrían positivas; no obstante, no es la intención del presente estudio, establecer correlaciones espurias entre las variables, sino conocer objetivamente el grado de asociación y de dependencia estadística real.

Figura 8. **Mapa de calor matriz de correlaciones, de las variaciones inter trimestrales de las afiliaciones a la seguridad social y las variaciones inter trimestrales de las actividades económicas**

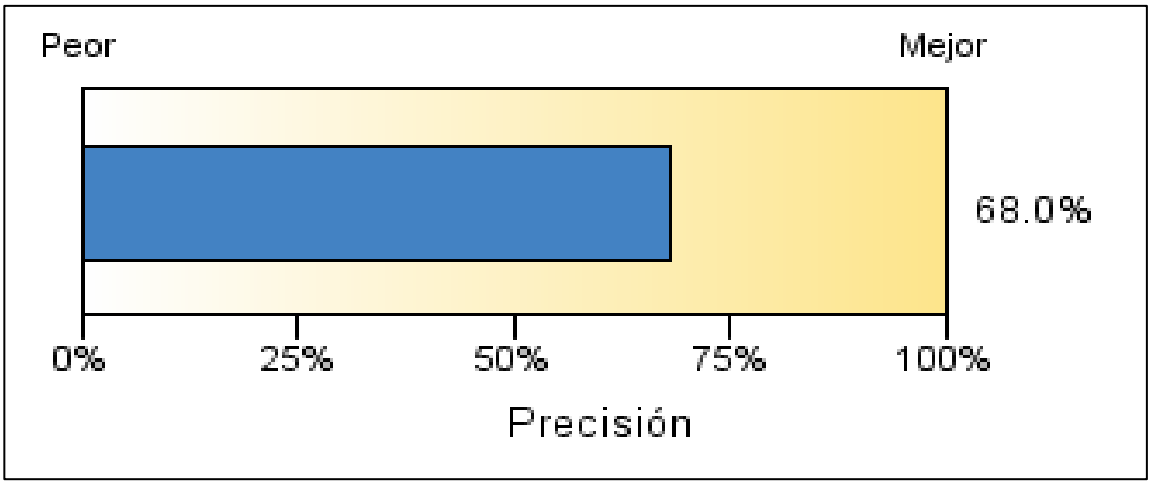


Fuente: elaboración propia, realizado con *Jamovi*.

Tal como lo muestra la figura 8, las correlaciones significativas de las variaciones de las afiliaciones a la seguridad social corresponden a las variaciones de las actividades económicas de agricultura, explotación de minas y canteras, construcción e intermediación financiera. La electricidad muestra una correlación negativa; las demás actividades económicas tienen correlación; sin embargo, estas no son significativas.

Sobre el grado de dependencia estadística, obtenido mediante la estimación del coeficiente de determinación ajustado (R^2A), el mismo es de 0.68; es decir, las variaciones inter trimestrales de las afiliaciones a la seguridad social son explicadas en un 68 % de las variaciones inter trimestrales de las 11 actividades económicas de Guatemala en el período 2001-2018.

Figura 9. **Grado de dependencia estadística de las variaciones inter trimestrales de las afiliaciones a la seguridad social, respecto de las variaciones inter trimestrales de las actividades económicas**



Fuente: elaboración propia, realizado con SPSS 25.

Tabla IV. **Datos atípicos Prueba de Grubbs**

Variable	N	Media	Desv.Est.	Mín.	Máx.	G	P
Seguridad Social	71	0.617	3.462	-5.856	8.149	2.18	1.000
1. Agricultura	71	2.99	18.85	-27.38	33.11	1.61	1.000
2. Minas	71	1.03	13.22	-49.31	34.40	3.81	0.004
3. Industrias manufactureras	71	0.739	5.213	-11.860	10.309	2.42	0.983
4. Electricidad	71	1.062	3.440	-7.866	9.849	2.60	0.568
5. Construcción	71	3.30	20.88	-38.14	29.29	1.98	1.000
6. Comercio	71	1.66	11.63	-25.15	15.59	2.31	1.000
7. Transporte	71	2.012	5.246	-9.526	15.394	2.55	0.653
8. Intermediación financiera	71	2.394	8.192	-13.918	21.660	2.35	1.000
9. Alquiler de vivienda	71	0.8026	0.2767	0.2313	1.3234	2.06	1.000
10. Servicios privados	71	0.944	2.498	-5.781	4.819	2.69	0.415
11. Administración pública	71	1.248	6.542	-10.419	27.834	4.06	0.001

Fuente: elaboración propia, realizado con SPSS 25.

Tabla V. **Variables con datos atípicos**

Variable	Fila	Valor atípico
2. Explotación de minas	67	-49.3090
11. Administración pública	7	27.8339

Fuente: elaboración propia, realizado con SPSS 25.

De acuerdo con la prueba de datos atípicos de *Grubss* en la tabla IV y V, se registra presencia de valores extremos en la actividad económica de minas y canteras y de la administración pública con variaciones Inter trimestrales de -49.31 % y 27.83 % respectivamente.

Tabla VI. **Coeficientes de correlación y determinación**

Regression Analysis			
R^2	0.761		
Adjusted R^2	0.68	n	71
R	**0.872**	k	12
Std. Error	1.861	Dep. Var.	**Y**

Fuente: elaboración propia, realizado con *Megastat.*

3.2. Comprobación de supuestos

Objetivo 2: comprobar los supuestos del análisis de regresión múltiple y de series de tiempo de las afiliaciones a la seguridad social y de las actividades económicas de Guatemala del período de 2001 a 2018 a través de las pruebas estadísticas para cada tipo de análisis.

3.2.1. Supuestos del análisis de regresión múltiple

Existencia de relación lineal entre las variables de las afiliaciones de la seguridad social y las actividades económicas de Guatemala.

3.2.1.1. Relación lineal entre variables

El coeficiente de correlación de Pearson (0.87), indica que existe fuerza de asociación estadística entre las variaciones inter trimestrales de las afiliaciones de la seguridad social y las variaciones inter trimestrales de las actividades económicas; así también, existe un fuerte grado de dependencia estadística, según el coeficiente de determinación (R^2 = 0.76 y R^2 Ajustado = 0.68), las variaciones de las afiliaciones de la seguridad social dependen de las variaciones de las actividades económicas en un 68 %.

El primer modelo de análisis de regresión múltiple, expuesto a continuación, parte del supuesto de que todas las variables independientes (variables inter trimestrales de las actividades económicas) explican el comportamiento de las variaciones inter trimestrales de las afiliaciones a la seguridad social (variable dependiente).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \ldots \beta_{11} X_{11} + \mu_i \text{ (Ec. 24)}$$

Donde:

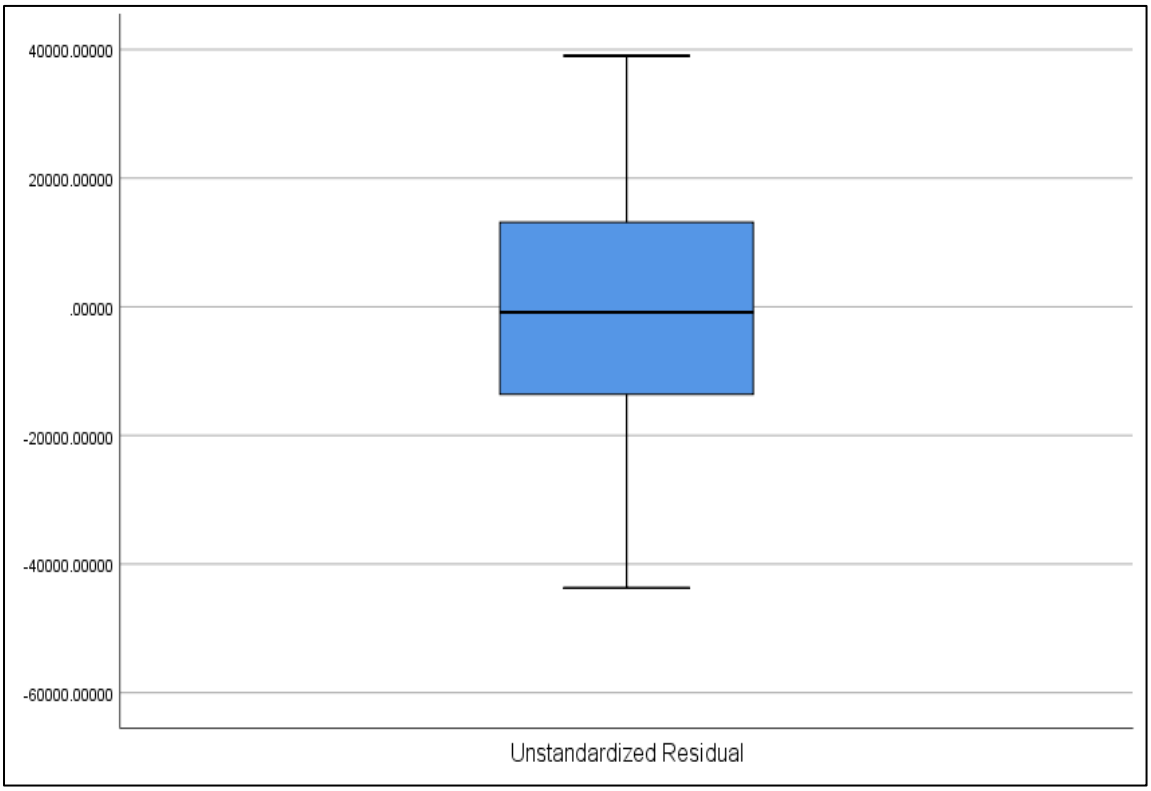
Y_i: variaciones inter trimestrales de las afiliaciones a la seguridad social
β_0: intercepto
β_1 ...: Coeficientes angulares para cada variable dependiente
X_1: Agricultura, Ganadería, Caza, Silvicultura y Pesca
X_2: Explotación de Minas y Canteras
X_3: Industrias Manufactureras
X_4: Suministro de Electricidad y Captación de Agua
X_5: Construcción
X_6: Comercio al por Mayor y Menor
X_7: Trasporte, Almacenamiento y Comunicaciones
X_8: Intermediación Financiera, Seguros y Actividades Auxiliares
X_9: Alquiler de Vivienda
X_{10}: Servicios Privados
X_{11}: Administración Pública y Defensa
μ_i: error estocástico (variable aleatoria no observable)
Sin embargo, dado que existe diferencia entre el coeficiente R^2 y R^2

Ajustado, es posible que ciertas variables independientes estén correlacionadas entre sí, es decir, estén brindado información redundante al modelo.

3.2.1.2. Valores atípicos

De acuerdo con la Figura 10, el diagrama de *box-plot* evidencia la no existencia de valores atípicos en los residuales del primer modelo, por tal razón, se cumple el segundo supuesto.

Figura 10. **Diagrama de cajas residuos no estandarizados primer modelo**

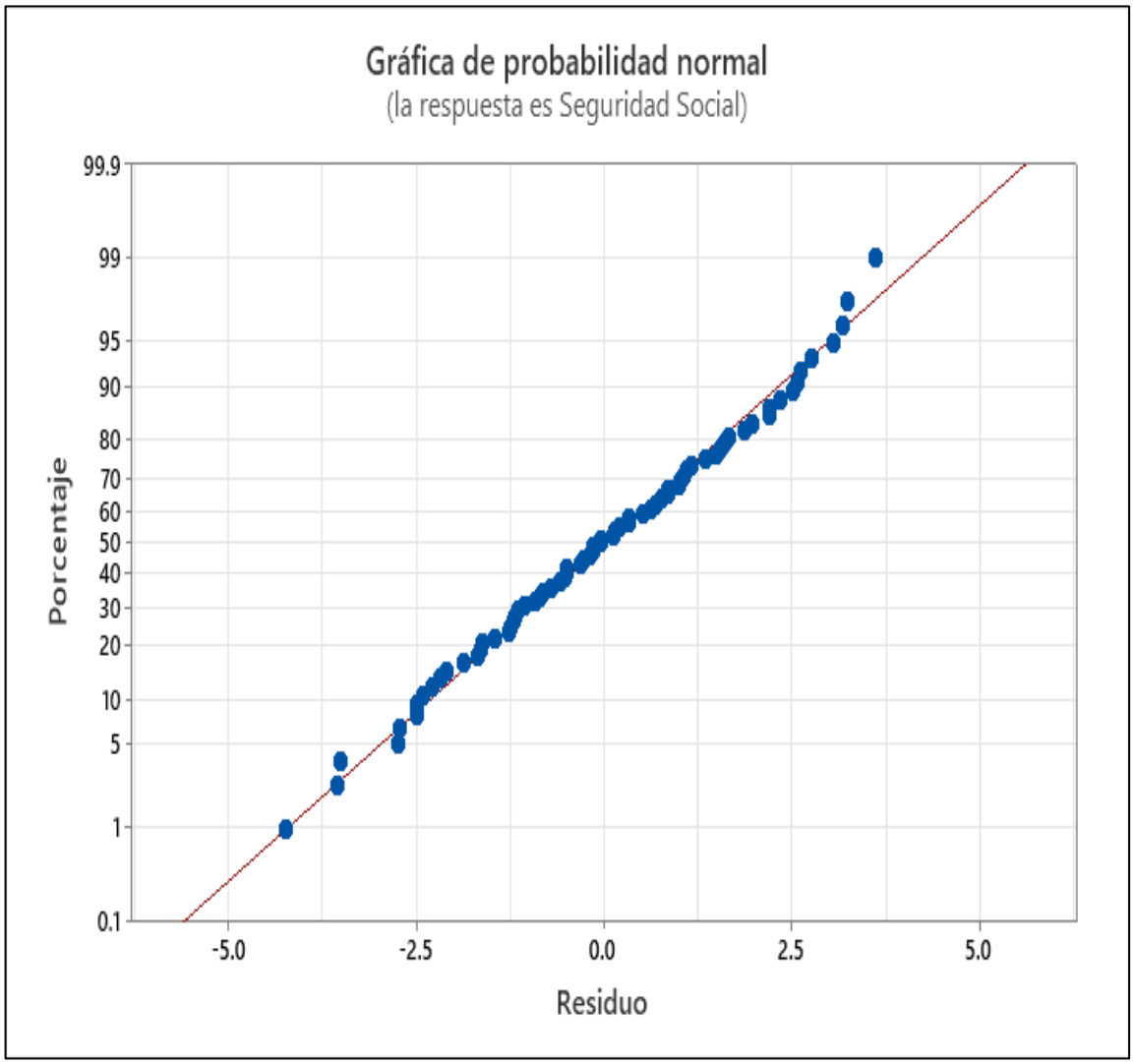


Fuente: elaboración propia, realizado con SPSS 25.

3.2.1.3. Distribución normal

Este supuesto indica que los residuos del modelo deben distribuirse de forma normal, de acuerdo con la figura 11, los residuos del primer modelo cumplen con este supuesto.

Figura 11. **Q-Q plot de los residuos**



Fuente: elaboración propia, realizado con Minitab 20.

Tabla VII. **Pruebas de normalidad**

Residuos no estandarizados	**Kolmogórov-Smirnov**			**Shapiro-Wilk**		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
	.070	71	.200*	.986	71	.630

Fuente: elaboración propia, realizado con SPSS 25.

De acuerdo con las pruebas de normalidad de *Kolmogórov-Smirnov* y *Shapiro-Wilk*, los residuales del primer modelo de regresión múltiple, distribuyen de forma normal, se cumple el tercer supuesto de análisis de regresión múltiple.

3.2.1.4. Independencia

De acuerdo con la tabla VIII, el valor de la prueba *Durbin-Watson* = 2.567, valor que se encuentra dentro del rango entre 1 y 3; según la teoría, si el valor de la prueba *D-W* es cercano a 2, los residuos son independientes; en ese sentido, se cumple el cuarto supuesto del análisis de regresión múltiple.

Tabla VIII. **Análisis de dependencia de los residuos**

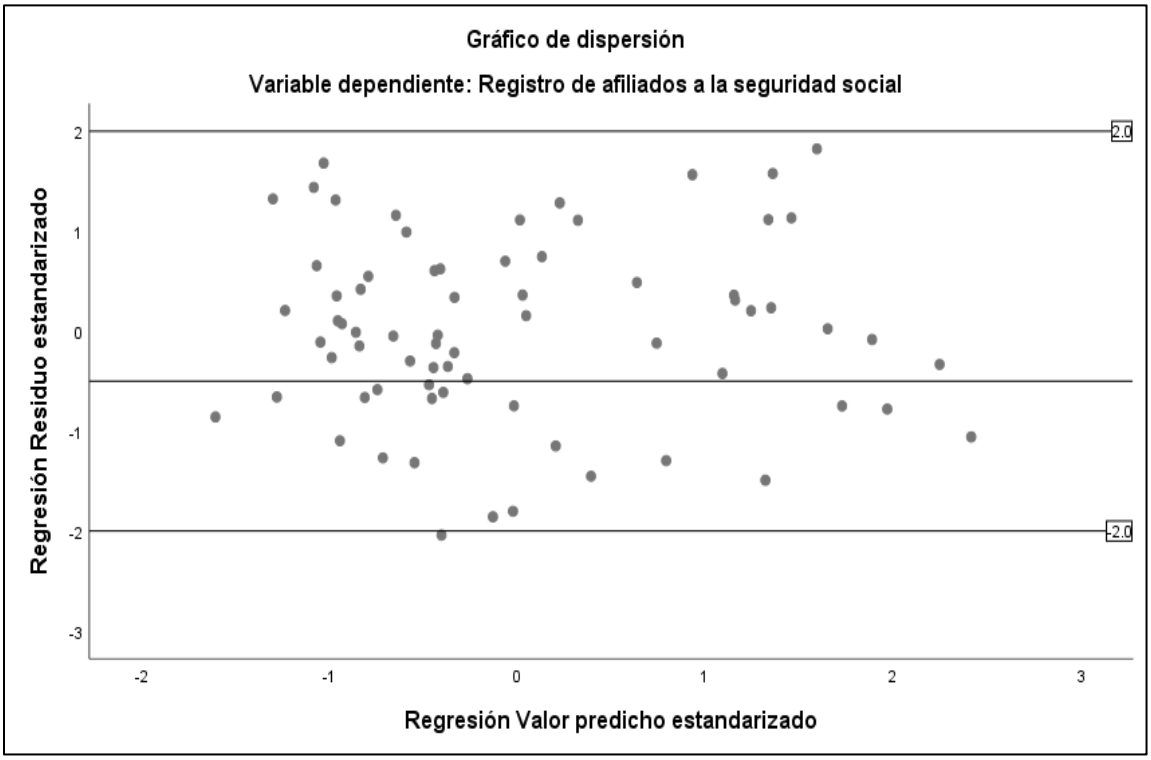
Modelo	**R**	**R cuadrado**	**R cuadrado ajustado**	**Error estándar de la estimación**	**Durbin-Watson**
1	.861[a]	.741	.693	21416.021	**2.567**

Fuente: elaboración propia, realizado con SPSS 25.

En la figura 12, se muestra en forma gráfica las distribuciones de los residuos, los cuales no muestran un patrón establecido (independencia), lo que

permitirá una predicción objetiva de la seguridad social ante cambios que se den en las actividades económicas.

Figura 12. **Dispersión en los residuos**



Fuente: elaboración propia, realizado con SPSS 25.

3.2.1.5. Homocedasticidad

De acuerdo con la estimación realizada en la prueba de White, se establece un planteamiento de H0: existe homoscedasticidad y H1: no existe homoscedasticidad. Dado que, el valor P (0.534) es > 0.05, no existe suficiente evidencia estadística para rechazar H0; por tal razón, se cumple el quinto supuesto del análisis de regresión múltiple.

Para el análisis de la prueba de White, fue necesario estimar una regresión lineal múltiple de 23 variables, se tomaron las variables independientes originales elevadas al cuadrado (22), así también, los residuos de la regresión múltiple del primer modelo elevados al cuadrado (1), se multiplicó el valor de las observaciones (70) por la nueva estimación de R^2 (0.2969) que dio como resultado el valor de la prueba de White (20.783); por último, se estimó el valor de P (0.534 > 0.05), el cual obedece a una distribución de Chi cuadrado con cola derecha.

Tabla IX. **Prueba de White**

R cuadrado	**0.2969**
observaciones	71
Prueba White	20.783
Valor de P	0.534

Fuente: elaboración propia.

3.2.2. Supuestos del análisis de series de tiempo

En este apartado se describen los supuestos de series de tiempo para las doce variables de estudio.

3.2.2.1. Estacionariedad

Para el análisis de serie de tiempo, es necesario que las series temporales tengan estacionariedad; es decir, que su esperanza matemática, varianza y autocorrelación no varíen en el tiempo (sean constantes).

De acuerdo con la tabla X, se pueden observar las series temporales de las afiliaciones a la seguridad social, agricultura, industria manufacturera, electricidad, comercio, y alquiler de vivienda, cumplen con el supuesto de estacionariedad en sentido débil de media constante igual a cero.

La prueba *Dickey Fuller* establece el siguiente planteamiento de hipótesis: H0: raíz unitaria (serie no estacionaria), H1: raíz no unitaria (serie estacionaria).

Tabla X. **Estacionariedad**

Variable	**Estacionariedad / Dickey-Fuller aumentado**		
	Valor P	**Decisión**	**Interpretación**
Afiliaciones a la seguridad social	0.01	Se rechaza H0	Serie estacionaria
1. Agricultura, ganadería, caza, silvicultura y pesca	0.01	Se rechaza H0	Serie estacionaria
2. Explotación de minas y canteras	0.15	No se rechaza H0	**Serie no estacionaria**
3. Industrias manufactureras	0.01	Se rechaza H0	Serie estacionaria
4. Suministro de electricidad y captación de agua	0.01	No se rechaza H0	Serie estacionaria
5. Construcción	0.11	No se rechaza H0	**Serie no estacionaria**
6. Comercio al por mayor y al por menor	0.01	Se rechaza H0	Serie estacionaria
7. Transporte, almacenamiento y comunicaciones	0.45	No se rechaza H0	**Serie no estacionaria**
8. Intermediación financiera, seguros y actividades auxiliares	0.18	No se rechaza H0	**Serie no estacionaria**
9. Alquiler de vivienda	0.01	Se rechaza H0	Serie estacionaria
10. Servicios privados	0.18	No se rechaza H0	**Serie no estacionaria**
11. Administración pública y defensa	0.25	No se rechaza H0	**Serie no estacionaria**

Fuente: elaboración propia, realizado con *Rstudio.*

3.2.2.2. Homogeneidad

El supuesto de homogeneidad describe que las observaciones de los datos en el tiempo no presentan cambios estructurales significativos, es decir, que la serie no deba trabajarse por segmentos dado el cambio estructural encontrado.

Tabla XI. **Prueba de Homogeneidad Normal Estándar**

Variable	Homogeneidad / SNHT		
	Valor P	**Decisión**	**Interpretación**
Afiliaciones a la seguridad social	0.69	No se rechaza H0	Serie homogénea
1. Agricultura	0.88	No se rechaza H0	Serie homogénea
2. minas y canteras	0.70	No se rechaza H0	Serie homogénea
3. Industrias	0.22	No se rechaza H0	Serie homogénea
4. Suministro de electricidad y captación de agua	0.15	No se rechaza H0	Serie homogénea
5. Construcción	0.94	No se rechaza H0	Serie homogénea
6. Comercio	1.00	No se rechaza H0	Serie homogénea
7. Transporte,	0.59	No se rechaza H0	Serie homogénea
8. Intermediación financiera,	0.78	No se rechaza H0	Serie homogénea
9. Alquiler de vivienda	0.04	Se rechaza H0	Serie no homogénea
10. Servicios privados	1.00	No se rechaza H0	Serie homogénea
11. Administración pública	0.95	No se rechaza H0	Serie homogénea

Fuente: elaboración propia con base en resultados de Rstudio.

De acuerdo con las pruebas de Homogeneidad realizadas a cada una de las series, la única actividad económica que no presenta homogeneidad es la de alquiler de viviendas; es decir, existe un cambio estructural significativo en la observación 13, que corresponde al cuarto trimestre de 2003.

Figura 13. **Prueba SNHT alquiler de vivienda**

```
        Standard Normal Homogeneity Test (SNHT)

data:  alquiler.ts
T = 9.4263, n = 71, p-value = 0.03835
alternative hypothesis: true delta is not equal to 0
sample estimates:
probable change point at time K
                              13
```

Fuente: elaboración propia, realizado con *Rstudio.*

3.2.2.3. Independencia en las observaciones

En el análisis de series temporales, se busca que la serie sea independiente; es decir, que los valores no estén auto correlacionados en el tiempo. De no considerarse lo anterior, se podría llegar a cometer errores en los modelos de pronósticos, consecuentemente, las interpretaciones de los resultados serían erróneas. Para ello, la prueba de *Ljung-Box* ayuda a establecer si la serie es independiente o no.

Tabla XII. **Pruebas de independencia de *Ljung-Box***

Variable	Independencia /Ljung-Box		
	Valor de P	**Decisión**	**Interpretación**
Afiliaciones a la seguridad social	0.20	No se rechaza H0	serie independiente
1. Agricultura, ganadería, caza, silvicultura y pesca	0.00	Se rechaza H0	serie no independiente
2. Explotación de minas y canteras	0.20	No se rechaza H0	serie independiente
3. Industrias manufactureras	0.00	Se rechaza H0	serie no independiente
4. Suministro de electricidad y captación de agua	0.04	Se rechaza H0	serie no independiente
5. Construcción	0.00	Se rechaza H0	serie no independiente
6. Comercio al por mayor y al por menor	0.00	Se rechaza H0	serie no independiente
7. Transporte, almacenamiento y comunicaciones	0.00	Se rechaza H0	serie no independiente
8. Intermediación financiera, seguros y actividades auxiliares	0.01	Se rechaza H0	serie no independiente
9. Alquiler de vivienda	0.36	No se rechaza H0	serie independiente
10. Servicios privados	0.02	Se rechaza H0	serie no independiente
11. Administración pública y defensa	0.04	Se rechaza H0	serie no independiente

Fuente: elaboración propia, realizado con Rstudio.

Según los resultados en la prueba de independencia de *Ljung-Box* para cada una de las variables de estudio se establece que, las afiliaciones a la seguridad social, agricultura y alquiler de vivienda son variables independientes, es decir no están auto correlacionadas.

Es importante mencionar que, de acuerdo con el modelo de pronóstico de mejor ajuste estimado, es posible corregir los problemas cuando no se cumple alguno de los supuestos de las series de tiempo, más adelante se mencionan las variables seleccionadas que mejor se ajustan a los datos del modelo de predicción y de pronóstico.

3.3. Modelos de predicción

Objetivo 3: estimar el modelo de predicción de mejor ajuste de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala entre los años 2001 y 2018 a través del diagnóstico y validación de cada modelo.

En la primera parte del objetivo anterior, se comprobaron los supuestos de análisis de regresión múltiple del modelo que incluyó las variaciones inter trimestrales de las actividades económicas de Guatemala del período 2001-2018 como variables que explican ($X1, X2 \ldots X11$) las variaciones inter trimestrales de las afiliaciones a la seguridad social en el país; sin embargo, en este apartado se incluye el diagnóstico de ese primer modelo, verificando si efectivamente es el modelo de mejor ajuste que permita realizar predicciones objetivas de las variaciones inter trimestrales a las afiliaciones a la seguridad social ante variaciones de las actividades económicas.

3.3.1. Modelo 1 de análisis de regresión múltiple

Tal como se mencionó antes, el primer modelo corresponde a la interacción de todas las variaciones inter trimestrales de las actividades económicas como variables explicativas a las variaciones inter trimestrales de las afiliaciones a la seguridad social.

3.3.1.1. Diagnóstico modelo 1

Para el diagnóstico se comparan los criterios de R^2, R^2 Ajustado, *Cp de Mallows,* análisis de inflación de la varianza (VIF), criterios de información de *Akaike (*AIC) y criterios *bayesianos* (BIC) entre modelos, seleccionando el que mejor se ajusta a los datos.

Figura 14. **Análisis de regresión múltiple del primer modelo**

Análisis de regresión lineal

Variable	N	R^2	R^2 Aj	ECMP	AIC	BIC
Seguridad Social	71	0.72	0.67	5.82	311.31	340.73

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor	CpMallows	VIF
const	1.43	1.16	-0.90	3.75	1.23	0.2248		
1. Agricultura, ganadería,..	0.22	0.05	0.12	0.32	4.45	<0.0001	29.77	15.34
2. Explotación de minas y ..	0.01	0.02	-0.04	0.06	0.36	0.7172	10.13	1.73
3. Industrias manufacturer..	0.22	0.07	0.09	0.35	3.34	0.0015	21.15	2.09
4. Suministro de electrici..	-0.16	0.11	-0.38	0.06	-1.45	0.1531	12.10	2.62
5. Construcción	-0.07	0.05	-0.17	0.03	-1.33	0.1872	11.78	20.62
6. Comercio al por mayor y..	-0.16	0.06	-0.29	-0.03	-2.50	0.0152	16.25	10.14
7. Transporte, almacenamie..	0.18	0.06	0.05	0.30	2.84	0.0061	18.09	1.93
8. Intermediación financie..	0.01	0.04	-0.08	0.10	0.15	0.8781	10.02	2.43
9. Alquiler de vivienda	-1.90	1.29	-4.49	0.68	-1.47	0.1459	12.17	2.28
10. Servicios privados	0.39	0.19	0.02	0.76	2.09	0.0408	14.37	3.86
11. Administración pública..	-0.14	0.05	-0.24	-0.05	-3.04	0.0035	19.25	1.72

Fuente: elaboración propia, realizado con *Infostat.*

Tabla XIII. **Ecuación de regresión primer modelo**

ss	=	1.45 + 0.2188 (agr) + 0.0086 (Expmc) + 0.2194 (im) - 0.162 4. (se) – 0.0691 (Cons) - 0.1617 (com) + 0.1776 (trans) + 0.0068 (intermf) – 1.94 (av) + 0.389 (sp) – 0.1446 (ap)

Fuente: elaboración propia, realizado con Minitab 20.

3.3.1.1.1. Criterio Cp de Mallows

Para el primer criterio de diagnóstico del modelo, se realiza un análisis de *Cp de Mallows* entre los predictores. De acuerdo con la teoría, se busca que los coeficientes sean los mayores posibles, esto quiere decir que, a mayor *Cp de Mallows* mayor información está brindando esa variable al modelo, al eliminar los variables con menor coeficiente, se evita el sesgo en la estimación de los verdaderos coeficientes de regresión otorgando resultados más precisos en las predicciones (Ec.21).

La figura 14 muestra que, el coeficiente del *Cp de Mallows* de las actividades de agricultura, industria manufacturera, comercio al por mayor y menor, transporte y almacenaje y la administración pública, brindan más información al modelo.

3.3.1.1.2. Criterio VIF

Adicional al criterio de *Cp de Mallows*, es importante agregar conjuntamente el criterio de inflación de la varianza (VIF por sus siglas en inglés, *variance inflation factor*), (Ec.18).

De acuerdo con la teoría, el VIF(i) encontrado no debe superar en 10, dado que, las variables pueden tener presencia de multicolinealidad.

En ese sentido, las variables de agricultura, construcción y comercio al por mayor y menor (esta última por muy poco) no cumplen con este supuesto.

3.3.1.1.3. Criterio *Akaike* (AIC)

La utilización de este criterio se basa en identificar la menor distancia existente entre las distribuciones a partir del logaritmo de máxima verosimilitud de cada modelo estimado; dado que, al realizar distintos modelos (considerando los criterios anteriores), el modelo de menor AIC calculado, es el que mejor se ajustará a las observaciones y por ende dará mejores predicciones de las variaciones de la seguridad social ante cambios de las variaciones de las actividades económicas seleccionadas en el modelo (Ec. 22).

3.3.1.1.4. Criterio *Bayesiano* (BIC)

Similar al criterio de AIC, el BIC introduce una modificación bayesiana al AIC; es decir, se sustituye el logaritmo de N por orden 2, con este elemento en la fórmula, se corrige el error de sobre estimación en las variables observadas, este criterio castiga de alguna forma los modelos con mayores parámetros de cada modelo, se selecciona el modelo con menor BIC calculado (Ec. 23).

Tomando en consideración los elementos anteriores, se presentan los resultados de un segundo modelo de análisis de regresión múltiple.

3.3.1.2. Diagnóstico modelo 2

Para construir el modelo 2, se dejaron las variables que mayor información (*Cp de Mallows*) proporcionaron al modelo y las de mayor coeficiente de inflación de la varianza (VIF), evitando la multicolinealidad de las variables.

Figura 15. **Análisis de regresión múltiple segundo modelo**

Análisis de regresión lineal

Variable	N	R^2	R^2 Aj	ECMP	AIC	BIC
Seguridad Social	71	0.71	0.68	4.85	306.96	327.33

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor	CpMallows	VIF
const	-0.06	0.30	-0.66	0.53	-0.21	0.8366		
1. Agricultura, ganadería,..	0.17	0.03	0.11	0.23	5.59	<0.0001	37.24	5.95
3. Industrias manufacturer..	0.24	0.06	0.11	0.36	3.77	0.0004	20.25	1.95
4. Suministro de electrici..	-0.18	0.10	-0.37	0.01	-1.86	0.0672	9.47	1.96
6. Comercio al por mayor y..	-0.17	0.06	-0.29	-0.05	-2.88	0.0054	14.30	8.61
7. Transporte, almacenamie..	0.21	0.06	0.10	0.33	3.82	0.0003	20.62	1.56
10. Servicios privados	0.23	0.12	-0.01	0.48	1.92	0.0592	9.69	1.68
11. Administración pública..	-0.15	0.04	-0.23	-0.06	-3.45	0.0010	17.89	1.41

Fuente: elaboración propia, realizado con *Infostat.*

Tabla XIV. **Ecuación de regresión segundo modelo**

$$ss = \begin{array}{c} -0.061 + 0.1699\ (agr) \\ + 0.2373\ (im) \\ - 0.1780\ (se) \\ - 0.1706\ (Com) \\ + 0.2137\ (Trans) \\ + 0.234\ (sp) \\ - 0.1472\ (ap) \end{array}$$

Fuente: elaboración propia, realizado con Minitab 20.

Dado que, la actividad económica de agricultura, ganadería y caza brindó la mayor información (*Cp de Mallows*) en el primer modelo, no se consideró oportuno eliminarla a pesar del VIF >10; sin embargo, al estimar el segundo modelo, es evidente que no existe inflación de la varianza (5.95<10); es decir, en esta variable no se evidencia multicolinealidad respecto a las demás (no sobre explica el modelo).

El segundo modelo de análisis de regresión múltiple, seleccionando 7 variables independientes, agricultura, industria manufacturera, suministro de electricidad, comercio al por mayor menor, transporte, servicios privados y administración pública, constituye un modelo que mejor su ajusta a los datos en comparación con el primero; sin embargo, en esta nueva estimación, se puede observar a través del *Cp de Mallows*, que las actividades económicas de suministros de electricidad y captación de agua con servicios privados, son los que otorgan menor información al modelo, con coeficientes 9.47 y 9.69 respectivamente.

3.3.1.3. Diagnóstico modelo 3

Para establecer el tercer modelo, se compararon los criterios de información R^2 Ajustado, AIC y BIC principalmente y los criterios de *Cp de Mallows* y VIF seguidamente, llegando a estimar el modelo con la menor cantidad de variables que explican de mejor las afiliaciones a la seguridad social.

Figura 16. **Análisis de regresión múltiple tercer modelo**

Análisis de regresión lineal

Variable	N	R^2	R^2 Aj	ECMP	AIC	BIC
Seguridad Social	71	0.68	0.65	4.89	310.35	326.19

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor	CpMallows	VIF
const	-0.01	0.28	-0.56	0.55	-0.03	0.9789		
1. Agricultura, ganadería,..	0.20	0.02	0.16	0.25	8.46	<0.0001	75.51	3.49
3. Industrias manufacturer..	0.19	0.06	0.06	0.31	3.03	0.0035	13.20	1.77
6. Comercio al por mayor y..	-0.21	0.05	-0.30	-0.12	-4.58	<0.0001	25.01	4.77
7. Transporte, almacenamie..	0.21	0.06	0.10	0.33	3.70	0.0005	17.67	1.56
11. Administración pública..	-0.17	0.04	-0.25	-0.08	-3.85	0.0003	18.84	1.36

Fuente: elaboración propia, realizado con *Infostat.*

Tabla XV. **Ecuación de regresión tercer modelo**

ss	=	-0.01+ 0.2043* (agr) + 0.1886* (im) - 0.2097* (comer) + 0.2143* (transp) - 0.1670* (ap) * significancia del 0.01, **significancia al 0.05 *** significancia al 0.10

Fuente: elaboración propia, realizado con Minitab 20.

Respetando el principio de parsimonia el cual menciona que, un modelo entre más sencillo sea y brinde la mayor explicación con la menor cantidad de variables posible, será un modelo de mejor ajuste y mayor precisión en sus predicciones.

En este tercer modelo mostrado en la figura 16, se evidencia mayor precisión en los criterios de *Cp de Mallows* y en los factores de la inflación de la varianza, el R^2 ajustado, disminuye en 2 puntos porcentuales (0.65). La siguiente tabla muestra la comparación de cada modelo de acuerdo con los criterios anteriores.

Tabla XVI. **Comparación de modelos de análisis de regresión lineal múltiple**

Criterio	**Modelo 1**	**Modelo 2**	**Modelo 3**	**Mejor Modelo**
R^2 ajustado	0.67	0.68	0.65	2
AIC	311.31	306.96	310.35	2
BIC	340.73	327.33	326.19	3
Rango valores *Cp de Mallows*	(10.02 – 29.77)	(9.47 – 37.24)	(13.20-75.51)29	3
Rango valores VIF	(1.72 - 20.62)	(1.56 - 8.61)	(1.36 - 4.77)	3
Variables independientes (parsimonia)	11	7	5	3

Fuente: elaboración propia.

De acuerdo con la tabla XVI, realizado con el principio de parsimonia y dado que, los valores de R^2 ajustado y los criterios de AIC y BIC no cambian significativamente entre el modelo 2 y 3, este último, constituye el de mejor ajuste.

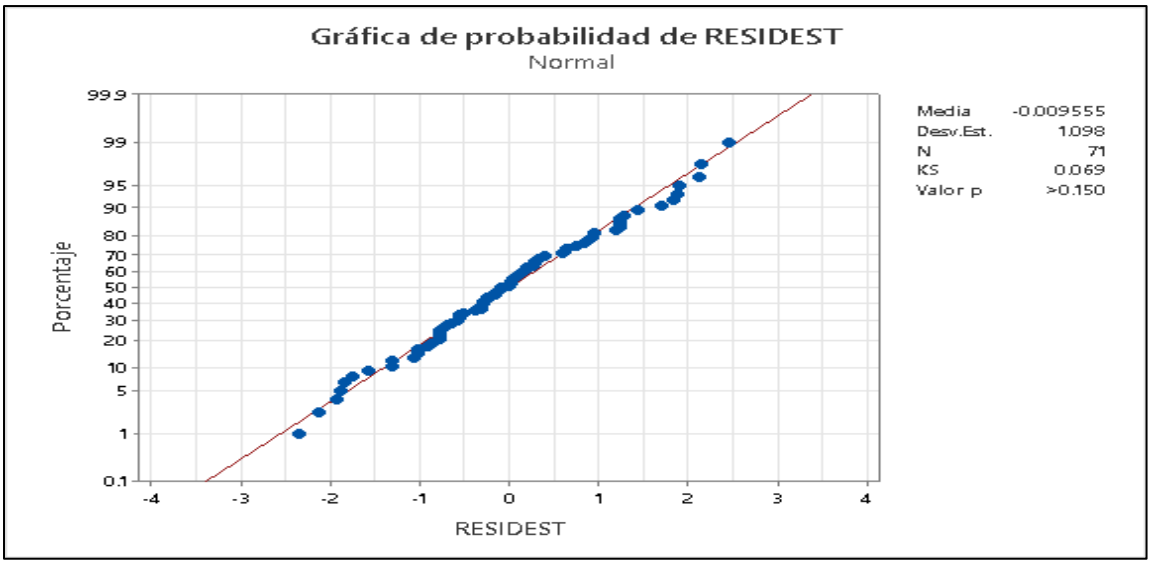
3.3.1.4. Análisis residuales del modelo

A continuación, se presenta un diagnóstico más profundo aplicado a los residuales del modelo 3 para confirmar que, efectivamente sea el modelo de mejor ajuste que permita predecir las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala según los datos del período 2001-2018.

3.3.1.4.1. Normalidad

De acuerdo con la figura 17, los residuos estandarizados distribuyen de forma normal.

Figura 17. **Prueba de normalidad en los residuos del modelo 3**



Fuente: elaboración propia, realizado con Minitab 20.

3.3.1.4.2. Datos Atípicos

Datos atípicos en los residuos: según los resultados de la prueba de en los residuos estandarizados del modelo 3, no hay presencia de datos atípicos.

Tabla XVII. **Datos atípicos en residuales en el modelo 3**

Variable	N	Media	Desv.Est.	Mín.	Máx.	G	P
RESIDEST	71	-0.010	1.098	-2.356	2.456	2.25	1.000

Fuente: elaboración propia, realizado con Minitab 20.

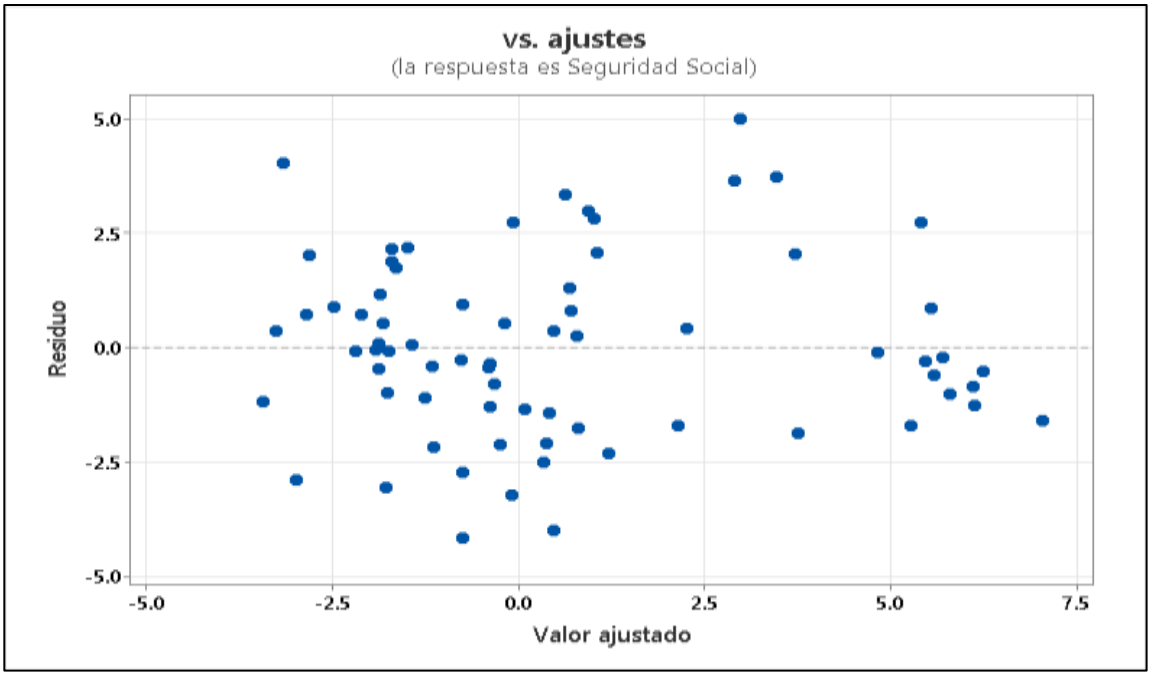
3.3.1.4.3. Independencia

Análisis de independencia entre los residuos estandarizados: de acuerdo con la prueba de *Durbin-Watson* = 2.75, la cual, se encuentra dentro del rango entre 1 y 3, se dice que, los residuos del modelo 3 son independientes.

3.3.1.4.4. Homoscedasticidad

La figura 18, muestra gráficamente que los residuos no muestran un patrón, es decir son homoscedásticos, la varianza en cada una de las observaciones es homogénea.

Figura 18. **Análisis de independencia de los residuos modelo 3**



Fuente: elaboración propia, realizado con *Infostat.*

3.3.1.4.5. Análisis de influencia

De acuerdo con el análisis de influencia del modelo 3, por medio del estadístico de *Leverage* (h_{ii}), el cual indica si existe un valor x_i y su contribución al valor ajustado (intercepto) que marque una distancia significativa que provoque un apalancamiento (influencia) en dicha constante, la particularidad de este análisis está en función de las variables independientes y no de la variable respuesta como se acostumbra.

Distancia de *Cook* (D_{cook}), en combinación con el estadístico de *Leverage*, es posible determinar la distancia de los puntos influyentes que provocan la distorsión del modelo.

$$D_i = e_i^2 * \frac{h_{ii}}{p*s^2*(1*h_{ii})^2} \qquad \text{(Ec. 25)}$$

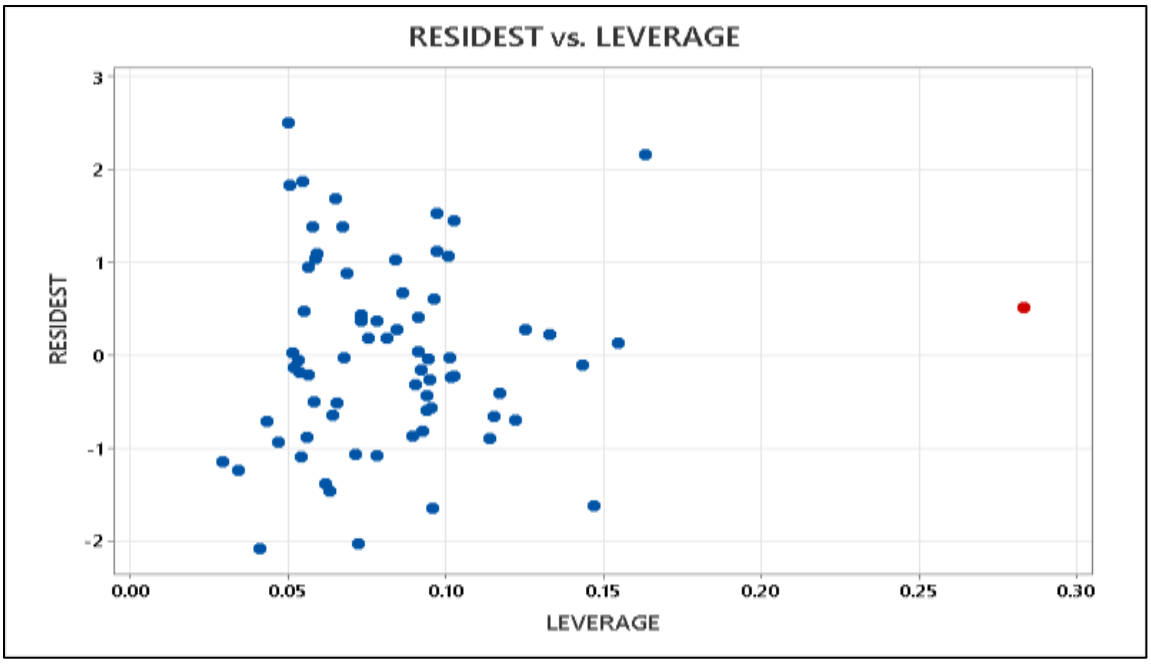
Figura 18. **Resumen de casos influyentes y distancia de Cook**

Tabla resumen criterios de diagnóstico

Caso	rs(min)	Caso	rs(max)	Caso	res(min)	Caso	res(max)	Caso	lev(max)	Caso	Cook(max)
17	-2.08	39	2.50	17	-2.14	39	2.61	6	0.28	1	0.15
26	-2.03	1	2.16	26	-2.08	1	2.23	1	0.16	9	0.08
13	-1.65	27	1.87	13	-1.68	27	1.91	32	0.15	39	0.06
9	-1.62	35	1.82	9	-1.64	35	1.86	9	0.15	26	0.05
29	-1.46	31	1.69	29	-1.47	31	1.71	7	0.14	13	0.05
36	-1.38	34	1.53	36	-1.39	34	1.55	24	0.13	34	0.04
2	-1.25	20	1.46	2	-1.25	20	1.47	69	0.13	20	0.04

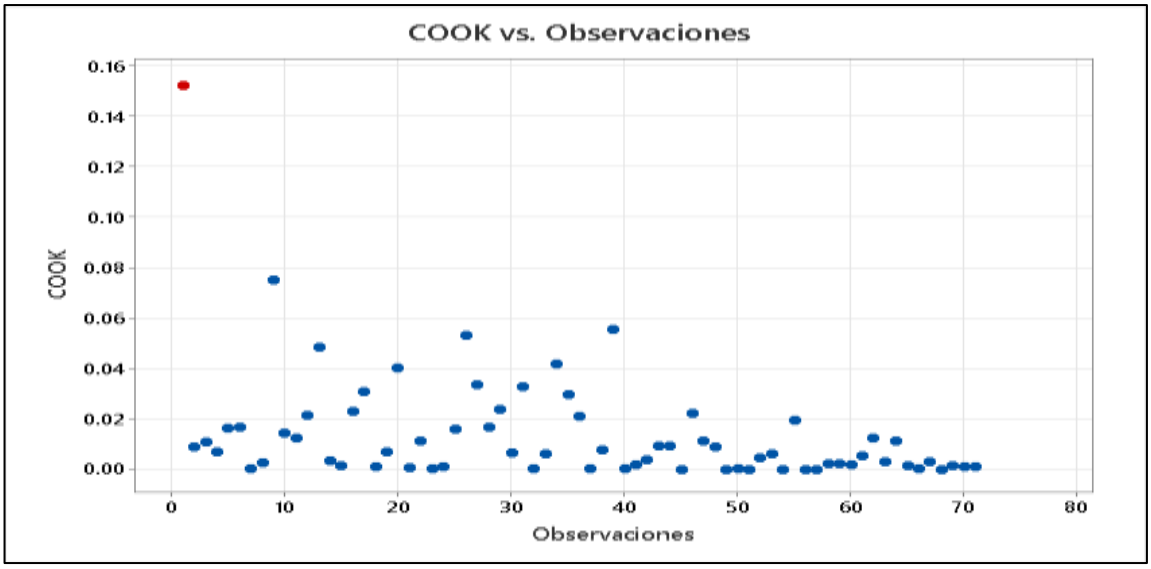
Fuente: elaboración propia, realizado con *Infostat.*

Figura 19. **Apalancamiento Leverage**



Fuente: elaboración propia, realizado con Minitab 20.

Figura 20. **Distancia de Cook**



Fuente: elaboración propia, realizado con Minitab 20.

De acuerdo al análisis de influencia, el caso 6 y el 39, representan las observaciones de mayor apalancamiento; sin embargo, no son los más influyentes, tal como lo muestra la Distancia de *Cook*, en los casos 1, 9, 26. Es decir, dichas observaciones afectan el modelo de predicción; es necesario mencionar que, las observaciones no se pueden extraer del modelo, dado que son registros reales (variaciones inter trimestrales) de las afiliaciones a la seguridad social; no obstante, es importante mencionar su influencia para la estimación del modelo de mejor ajuste.

3.3.1.5. Interpretación general del modelo 3

Las variaciones inter trimestrales de las actividades económicas de agricultura, ganadería y caza, industrias manufactureras, comercio al por mayor y menor, transporte, almacenamiento y comunicaciones y la administración pública y defensa explicaron en 65 % el promedio de las variaciones inter trimestrales de las afiliaciones a la seguridad social en Guatemala en el período 2001 y 2018, existe un 32 % no explicado que se deben a factores o perturbaciones estocásticas no incluidas en el modelo.

Tabla XVIII. **Ecuación final de regresión múltiple, variaciones inter trimestrales de las afiliaciones a la seguridad social respecto a las variaciones inter trimestrales de las actividades económicas de Guatemala en el período 2001-2018**

$$ss = -0.01 + 0.2043\,(agr) + 0.1886\,(im) - 0.2097\,(com) + 0.2143\,(transp) - 0.1670\,(ap)$$

Fuente: elaboración propia, realizado con Minitab 20.

- Si la actividad de agricultura aumenta su variación en un punto porcentual (1 %), las afiliaciones a la seguridad social aumentarán en veinte puntos básicos (0.20 %) en promedio, *Caeterīs pāribus.*

- Si la actividad económica de la industria manufacturera aumenta su variación en un punto porcentual (1 %), las afiliaciones a la seguridad social aumentarán en diecinueve puntos básicos (0.19 %) en promedio, *Caeterīs pāribus.*

- Si la actividad económica de comercio al por mayor y al por menor aumenta su variación en un punto porcentual (1 %), las afiliaciones a la seguridad social disminuirán en veintiún puntos básicos (-0.21 %) en promedio, *Caeterīs pāribus*.

- Si la actividad económica de transporte, almacenamiento y comunicación aumenta su variación en un punto porcentual (1 %), las afiliaciones a la seguridad social aumentarán en veintiún (0.21 %) en promedio, *Caeterīs pāribus*.

- Si la actividad económica de administración pública y defensa aumenta su variación porcentual en un punto básico (1 %), las afiliaciones a la seguridad social disminuirán en diecisiete puntos básicos (-0.17 %) en promedio, *Caeterīs pāribus*. Tanto este resultado, como el del comercio al por mayor y menor que muestran relaciones inversamente proporcional respecto a las variaciones de la seguridad social; podría deberse, en el caso de la administración pública, que el incremento en su aporte al PIB se dé por medio de la contratación de mayores trabajadores bajo contratos sin prestaciones y sin afiliación al IGSS y para el caso del comercio al por mayor y menor, el desarrollo de nuevos mecanismos de comercialización

que requieren menos uso incentivo en mano de obra; por ejemplo, el incremento de ventas por medios digitales, entre otros; no obstante, estos extremos debe ser confirmados por estudios donde se planten y comprueben ese tipo de hipótesis.

3.4. Modelos de pronóstico

Objetivo 4: estimar el modelo de pronóstico de mejor ajuste de corto plazo de las variaciones inter trimestrales de las afiliaciones a la seguridad social y de las actividades económicas en Guatemala según los datos del 2001 a 2018 a través del diagnóstico y validación del modelo.

3.4.1. Pronóstico afiliaciones a la seguridad social

Para estimar el primer pronóstico, se tomaron las afiliaciones acumuladas de cada trimestre del 2001 al 2018.

Tabla XIX. **Modelo de pronóstico de mejor ajuste serie acumulada de afiliaciones a la seguridad social**

	Descripción del modelo		
			Tipo de modelo
ID de modelo	Afiliaciones a la seguridad social, serie trimestral acumulada	Modelo	Multiplicativo de Winters

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XX. **Estadísticos de ajuste del modelo serie acumulada de afiliaciones a la seguridad social**

Estadístico de ajuste	Media
R cuadrado-estacionaria	.412
R cuadrado	.981
RMSE	18296.948
MAPE	1.212
MaxAPE	4.497
MAE	13480.65
MaxAE	49279.48
BIC normalizado	19.807

Fuente: elaboración propia, realizado con SPSS 25.

De acuerdo con el resumen estadístico de ajuste del modelo Multiplicativo de *Winters*, el modelo evidencia un nivel de error de 1.21 %, es decir, el modelo se ajusta bien a los datos de las afiliaciones trimestrales acumuladas de la seguridad social en el período 2001-2018.

Tabla XXI. **Diagnóstico del modelo afiliaciones a la Seguridad Social**

Modelo	Ljung-Box Q (18)			Número de valores atípicos
	Estadísticos	DF	Sig.	
Afiliaciones a la seguridad social, serie trimestral acumulada	18.351	15	.245	0

Fuente: elaboración propia, realizado con SPSS 25.

La tabla XXI, muestra el estadístico de *Ljung-Box*, el cual establece una H0: existencia de ruido blanco *(*White *noise),* media constante igual a cero $E(et) = 0$; varianza constante $E\ (et^2)\ =\ s^2$ y no autocorrelación $E\ (et, et\ *)\ =\ 0$, dado que, el valor P = 0.245 > 0.05, se puede mencionar que, el modelo se ajusta a los datos trimestrales acumulados de las afiliaciones a la seguridad social.

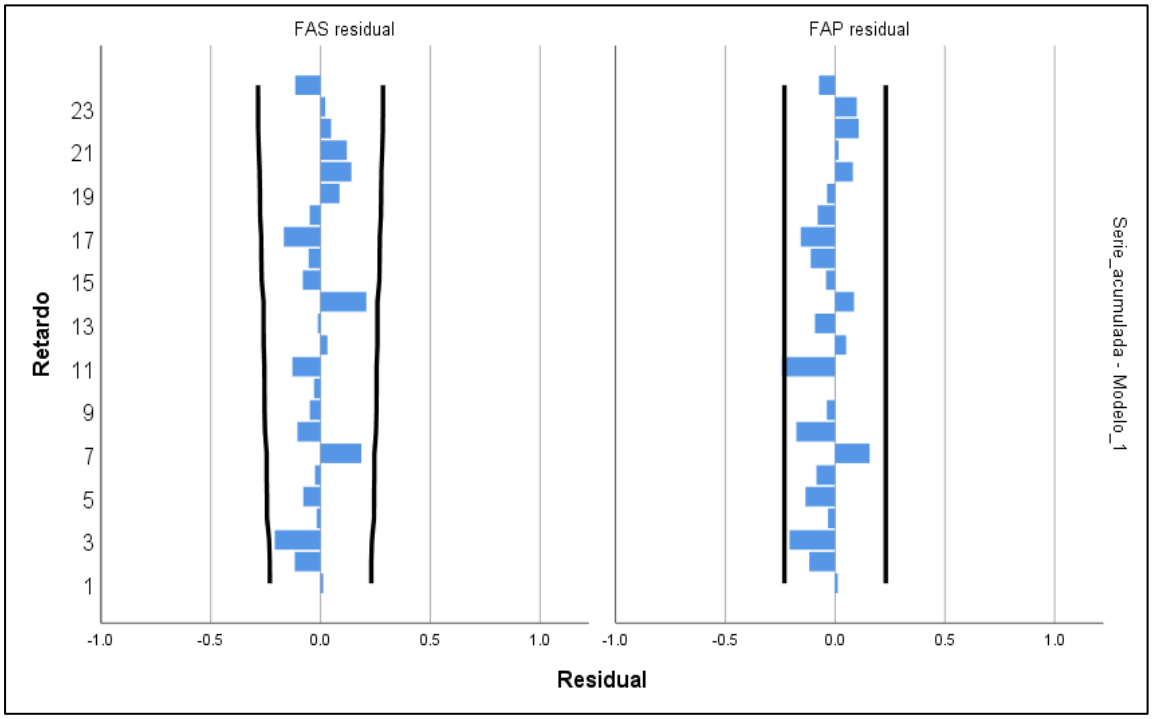
Modelo Multiplicativo de *Winters*. Según IBM (2022), este modelo es apropiado para series con una tendencia lineal y un efecto estacional que depende del nivel de la serie. Sus parámetros de suavizado son el nivel, la tendencia y la estación. El suavizado exponencial multiplicativo de Winters no es similar a ningún modelo ARIMA.

Tabla XXII. **Parámetros del modelo de suavizado exponencial, afiliaciones a la seguridad social**

Modelo			Estimación	SE	t	Sig.
Afiliaciones a la seguridad social, serie trimestral acumulada	Ninguna transformación	Alfa (nivel)	.663	.111	5.982	.000
		Gamma (tendencia)	.001	.071	.014	.989
		Delta (estacionalidad)	.495	.228	2.177	.033

Fuente: elaboración propia, realizado con SPSS 25.

Figura 21. **Residuales ACF y PACF pronóstico de serie acumulada de las afiliaciones a la seguridad social**



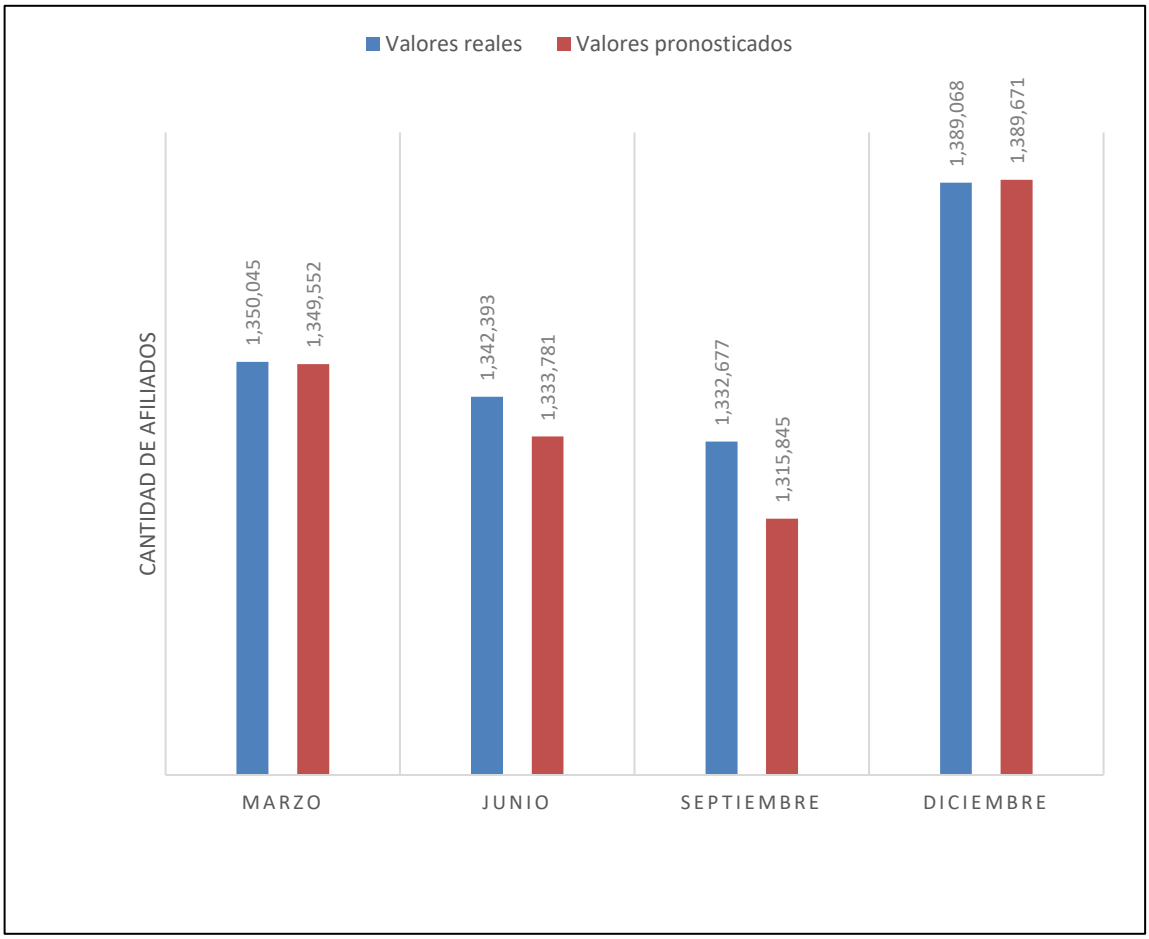
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXIII. **Valores pronosticados de las afiliaciones de la seguridad social cuatrimestres del 2019, 95 % de confianza**

Modelo		Predicción			
		T1 2019	**T2 2019**	**T3 2019**	**T4 2019**
Afiliaciones a la seguridad social, serie trimestral acumulada	Predicción	1,349,552	1,333,781	1,315,845	1,389,671
	Intervalo superior	1,386,053	1,377,378	1,365,320	1,446,386
	Intervalo inferior	1,313,050	1,290,185	1,266,370	1,332,955

Fuente: elaboración propia, realizado con SPSS 25.

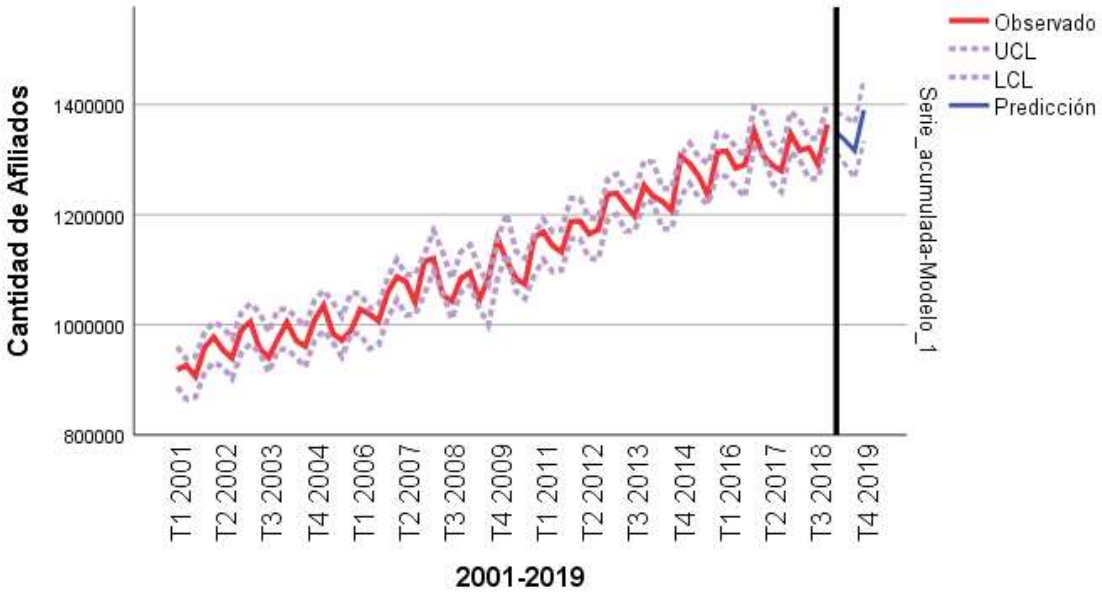
Figura 22. **Comparación cifras reales vs pronóstico 2019 afiliaciones a la seguridad social**



Fuente: elaboración propia.

La figura 23, muestra la comparación de cifras reales registradas durante los cuatrimestres del 2019 y los pronósticos realizados por el modelo Multiplicativo de *Winters*, la mayor diferencia se registra en el tercer cuatrimestre de ese año (-16,832); sin embargo, los valores pronosticados en los demás cuatrimestres son muy cercanos a la realidad de ese año.

Figura 23. **Pronóstico de las afiliaciones a la seguridad social en Guatemala año 2019**



Fuente: elaboración propia, realizado con SPSS 25.

Dado que, en el primer pronóstico de las afiliaciones a la seguridad social, se utilizó la variable acumulada de afiliaciones para cada trimestre en evaluación, se considera importante mostrar el ajuste del modelo con las variaciones inter trimestrales, extrayendo el componente tendencial de la serie.

Tabla XXIV. **Modelo de pronóstico de mejor ajuste variaciones Inter trimestrales de la seguridad social**

Descripción del modelo			
ID de modelo	Seguridad Social	Modelo_1	Tipo de modelo Estacional simple

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXV. **Estadísticos de ajuste del modelo con variaciones inter trimestrales de la seguridad social**

Estadístico de ajuste	Media
R cuadrado-estacionaria	.685
R cuadrado	.742
RMSE	1.773
MAPE	95.701
MaxAPE	935.183
MAE	1.300
MaxAE	5.577
BIC normalizado	1.265

Fuente: elaboración propia, realizado con SPSS 25.

La tabla XXV, muestra el resumen de los estadísticos de ajuste, donde se visualiza un error porcentual absoluto medio (MAPE) de 95.701 %; de acuerdo con este criterio, los datos no tienen un buen ajuste; sin embargo, la comparación siguiente de las cifras reales del año 2019, el pronóstico es muy cercano a la realidad.

De acuerdo con la teoría, el MAPE, es altamente influenciado “debido a que divide el error absoluto entre los datos reales, los valores que se aproximan a 0 pueden aumentar significativamente el MAPE” (https://support.minitab.com, recuperado 11 de julio 2022), por ello, se deben considerar otros criterios de selección del modelo.

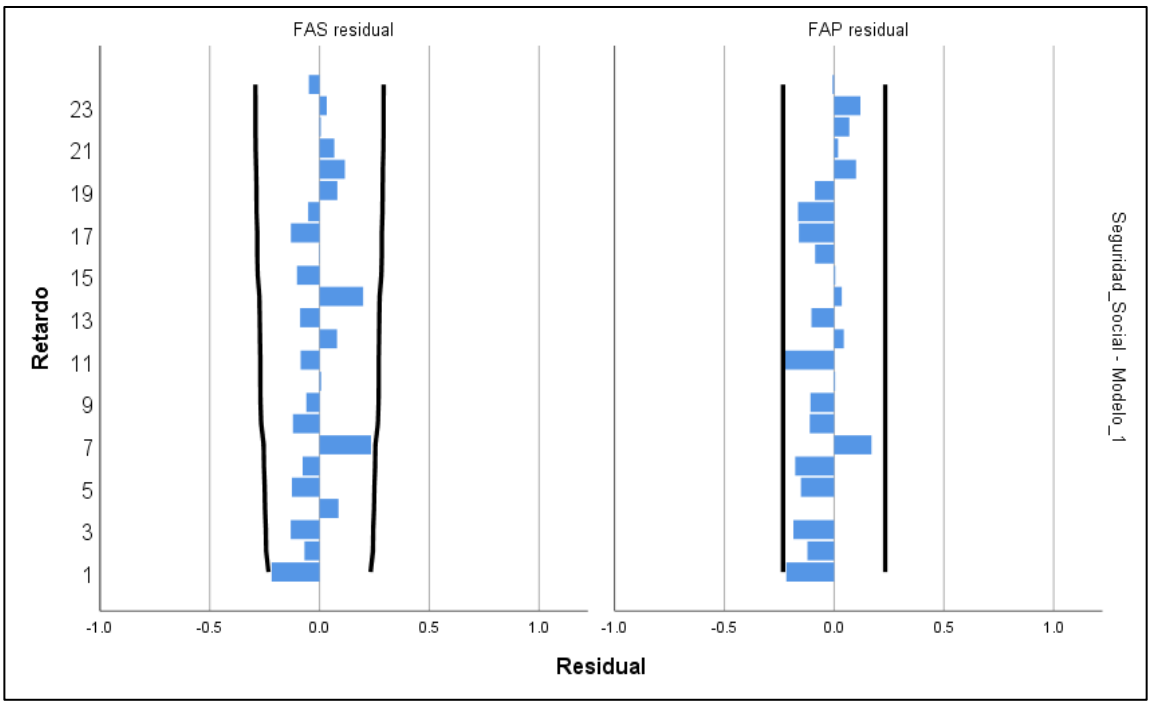
Tabla XXVI. **Diagnóstico del modelo de las variaciones Inter trimestrales de la seguridad social**

Modelo	Ljung-Box Q (18)			Número de valores atípicos
	Estadísticos	DF	Sig.	
Variaciones Inter trimestrales de la seguridad Social-Model	21.813	16	.149	0

Fuente: elaboración propia, realizado con SPSS 25.

La tabla XXVI, muestra el estadístico de *Ljung-Box* = 0.149 >0.05, H0: existencia de ruido blanco (White *noise*), media constante igual a cero E(et)=0; varianza constante E (et^2) = s^2 y no autocorrelación E (et, et*) = 0, el modelo se ajusta a los datos de las variaciones trimestrales de las afiliaciones a la seguridad social.

Figura 24. **Residuales de ACF Y PACF pronóstico de serie con variaciones Inter trimestral de las afiliaciones a la seguridad social**



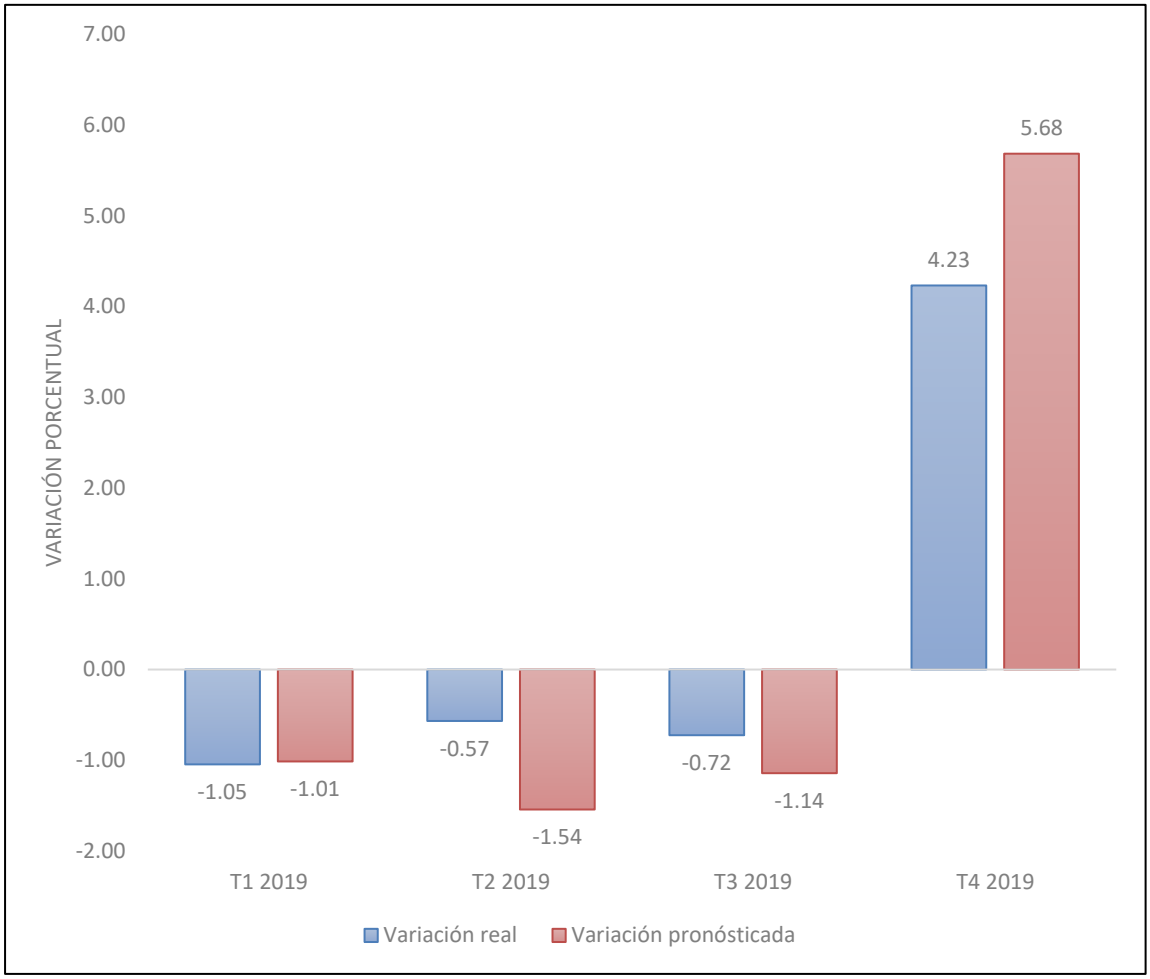
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXVII. **Pronóstico de la variación Inter trimestral de las afiliaciones a la seguridad social 2019**

Modelo		Predicción			
		T1 2019	T2 2019	T3 2019	T4 2019
Variaciones Inter trimestrales de la Seguridad Social	Predicción	-1.01	-1.54	-1.14	5.68
	UCL	2.52	1.99	2.39	9.22
	LCL	-4.55	-5.08	-4.68	2.15

Fuente: elaboración propia, realizado con SPSS 25.

Figura 25. **Comparación de variación real vs pronóstico del año 2019 variaciones inter trimestrales de las afiliaciones a la seguridad social con un nivel de 95 % de confianza**



Fuente: elaboración propia.

De acuerdo con este segundo modelo de pronóstico, se ve un buen ajuste a los datos de la variación Inter trimestral de las afiliaciones a la seguridad social del segundo trimestre de 2001 al cuarto trimestre de 2018, los valores pronosticados son muy cercanos a la realidad evidenciada en los trimestres de ese año.

3.4.2. Pronósticos actividades económicas

Para los pronósticos de las actividades económicas, se tomaron las variaciones Inter trimestrales de las actividades económicas que tuvieron influencia significativa en el tercer modelo de predicción del análisis de regresión múltiple del objetivo 3.

3.4.2.1. Agricultura, Ganadería y caza

Los valores para considerar en el pronóstico de la actividad económica de agricultura, ganadería y caza constituyen las variaciones inter trimestrales del 2001 al 2018.

Tabla XXVIII. **Modelo de pronóstico de la actividad de agricultura**

Descripción del modelo			
			Tipo de modelo
ID de modelo	1. Agricultura, ganadería, caza, silvicultura y pesca	Modelo_1	ARIMA (0,0,3) (0,1,1)

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXIX. **Estadísticos de ajuste del modelo ARIMA, actividad económica de agricultura**

Estadístico de ajuste	Media
R cuadrado-estacionaria	.308
R cuadrado	.988
RMSE	2.055
MAPE	53.705
MaxAPE	969.708
MAE	1.521
MaxAE	5.482
BIC normalizado	1.566

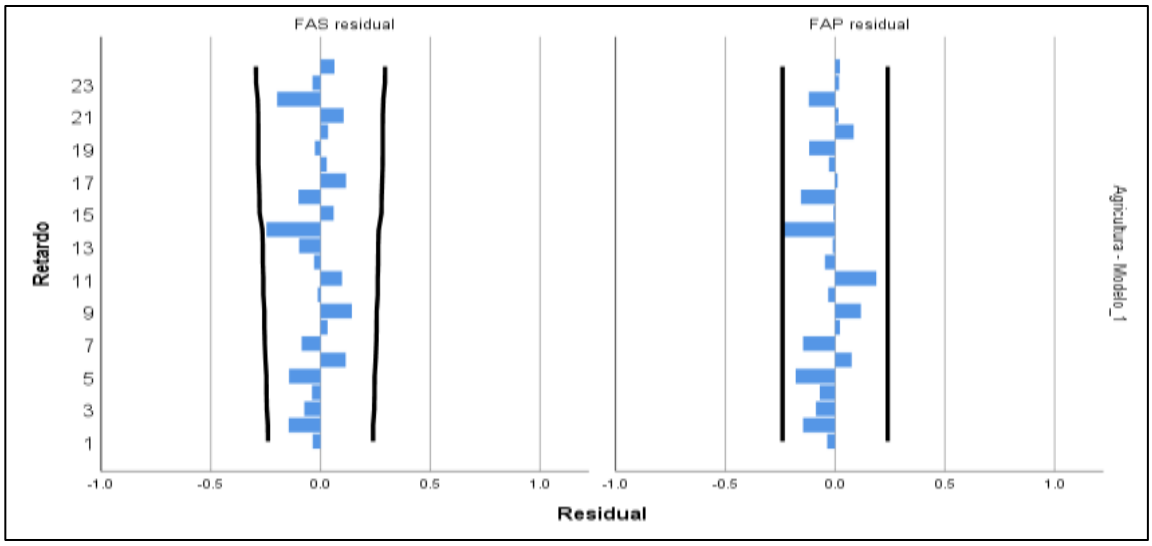
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXX. **Diagnóstico del modelo de predicción, actividad económica de agricultura**

Modelo	Ljung-Box Q (18)			Número de valores atípicos
	Estadísticos	DF	Sig.	
1. Agricultura, ganadería, caza, silvicultura y pesca-Modelo_1	16.448	16	.422	0

Fuente: elaboración propia, realizado con SPSS 25.

Figura 26. **ACF y PACF de residuales del modelo de pronóstico de la actividad económica de agricultura**



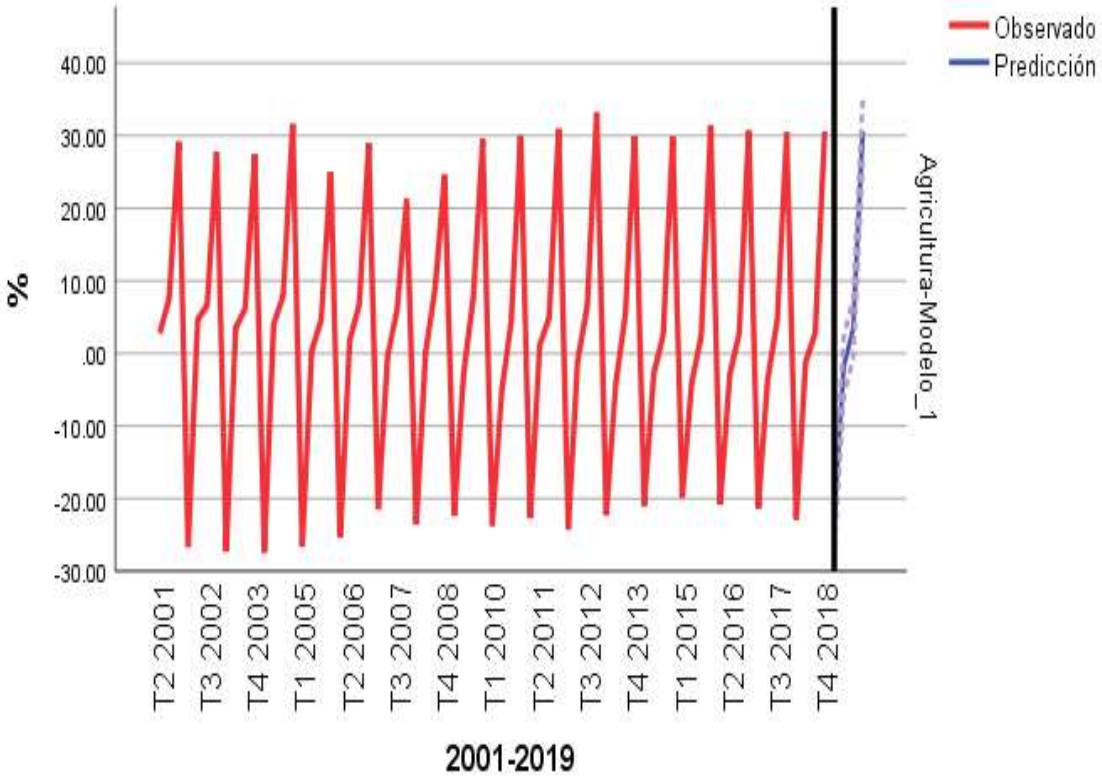
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXI. **Pronóstico de la actividad económica de agricultura del año 2019**

Modelo		T1 2019	T2 2019	T3 2019	T4 2019
1. Agricultura, ganadería, caza, silvicultura y pesca	Predicción	-22.92	-1.79	3.47	30.58
	UCL	-18.84	2.30	7.55	34.91
	LCL	-27.01	-5.87	-.62	26.25

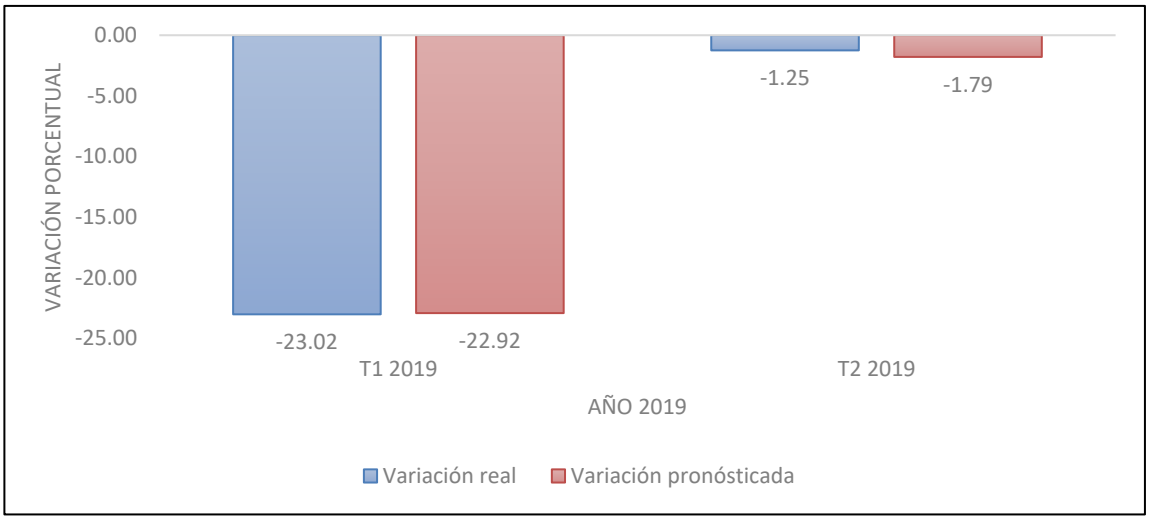
Fuente: elaboración propia, realizado con SPSS 25.

Figura 27. **Pronóstico de la actividad económica de agricultura para el año 2019**



Fuente: elaboración propia, realizado con SPSS 25.

Figura 28. **Comparación de valores reales y valores pronosticados de las variaciones de la actividad económica agricultura 2019 con un nivel de 95 % de confianza**



Fuente: elaboración propia.

La actividad económica de agricultura, ganadería, silvicultura y pesca de acuerdo con los registros de variaciones Inter trimestrales del año 2001 al 2018, se establece un modelo ARIMA de orden (0,0,3) (0,1,1), es decir, con tendencia secular: cero diferencias y autorregresivos y 3 medias móviles; con tendencia o variaciones estacionales fue necesario aplicar cero autorregresivos, una diferencia y una media móvil.

De esa cuenta, se estableció un error porcentual medio (MAPE) de 53.7 % en el modelo; sin embargo, como se mencionó anteriormente, fue necesaria la evaluación de otros criterios, como el caso de la comprobación de ruido blanco de *Ljung-Box* y se estableció un valor p = 0.422 >0.05, existe presencia de ruido blanco en los residuales.

En la figura 27, se visualiza los criterios ACF y PACF donde es posible observar que los residuos no están auto correlacionados en el tiempo.

La tabla XXXI, muestra los valores pronosticados para cada uno de los cuatrimestres del 2019.

La figura 29, permite hacer una comparación de los valores reales mostrados en los primeros 2 trimestres del año 2019. Es importante mencionar que, en los valores del tercer y cuarto trimestre de ese año, el precio de referencia 2001 que el Banco de Guatemala utiliza para el cálculo del PIB, cambió al año de referencia 2013; por dicha razón, no se pueden cotejar las variaciones inter trimestrales de las actividades económicas de esos períodos.

Los valores pronosticados respecto de las cifras reales reportadas en el período de evaluación son muy cercanos entre sí.

3.4.2.2. Industria manufacturera

La industria manufacturera, ha registrado un peso relativo promedio del 18.38 % respecto del PIB trimestral durante los 19 años de estudio, para realizar el pronóstico de corto plazo, se utilizó la variación inter trimestral de dicho período.

Tabla XXXII. **Modelo de pronóstico de industria**

Descripción del modelo			
ID de modelo	3. Industrias manufactureras	Modelo_1	Tipo de modelo Estacional simple

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXIII. **Estadísticos de ajuste del modelo**

Estadístico de ajuste	Media
R cuadrado-estacionaria	0.676
R cuadrado	0.799
RMSE	2.355
MAPE	82.198
MaxAPE	881.033
MAE	1.810
MaxAE	6.858
BIC normalizado	1.833

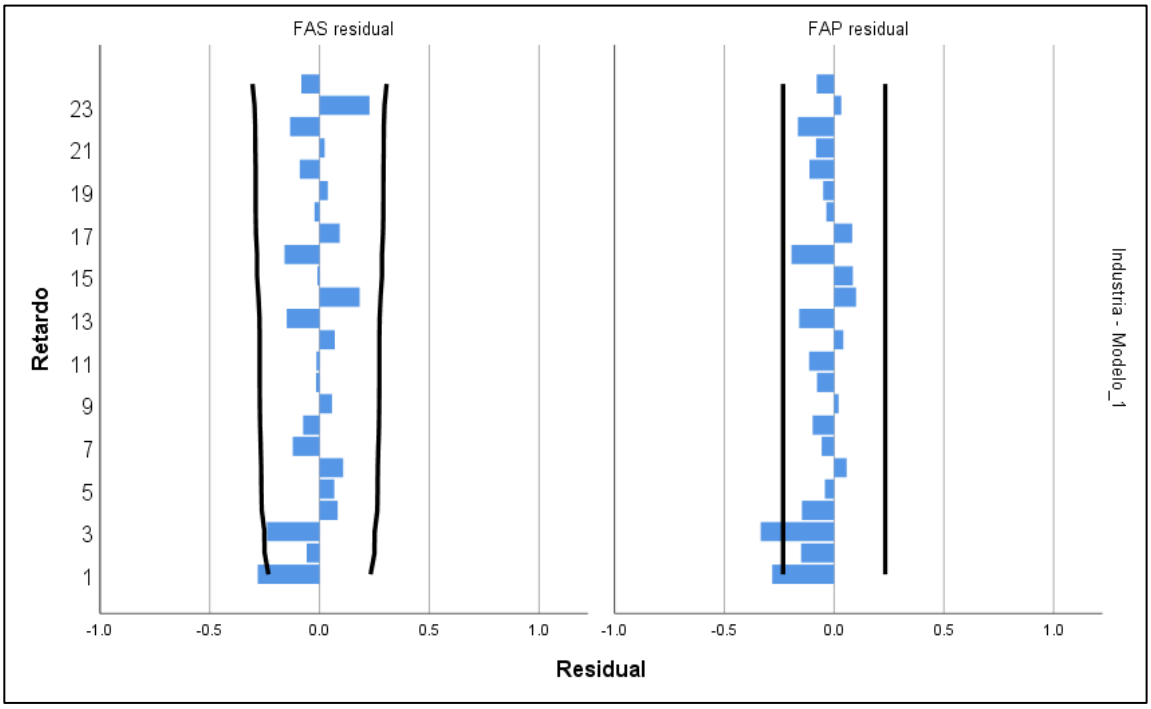
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXIV. **Diagnóstico del modelo de industria manufacturera**

Modelo	Estadísticos del modelo			Número de valores atípicos
	Ljung-Box Q (18)			
	Estadísticos	DF	Sig.	
3. Industrias manufactureras	23.032	16	.113	0

Fuente: elaboración propia, realizado con SPSS 25.

Figura 30. **ACF y PACF residuales del modelo, industria manufacturera**



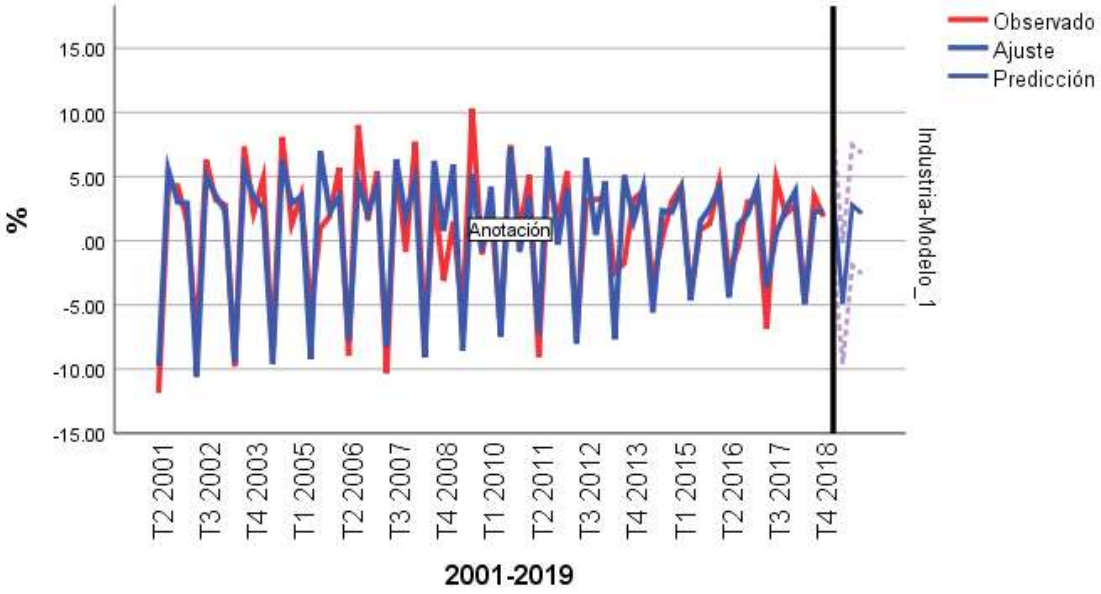
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXV. **Pronóstico año 2019, industria manufacturera**

Modelo		Predicción			
		T1 2019	T2 2019	T3 2019	T4 2019
3. Industrias manufactureras-Modelo_1	Predicción	3.42	-4.87	2.76	2.13
	UCL	8.12	-.17	7.46	6.83
	LCL	-1.27	-9.57	-1.94	-2.57

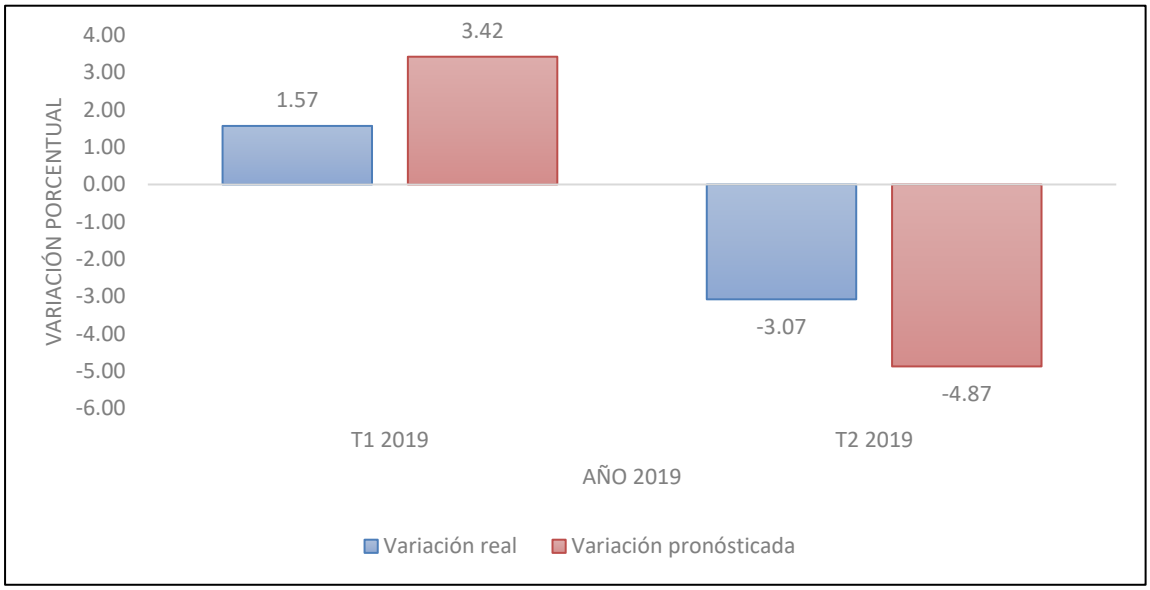
Fuente: elaboración propia, realizado con SPSS 25.

Figura 31. **Pronóstico 2019 industria manufacturera**



Fuente: elaboración propia, realizado con SPSS 25.

Figura 32. **Comparación de valores reales y valores pronosticados de las variaciones de la actividad económica industria manufacturera 2019 con un nivel del 95 % de confianza**



Fuente: elaboración propia.

Según los resultados de la estimación del modelo de pronóstico de la actividad económica industria manufacturera, el modelo que mejor su ajusta a las variaciones inter trimestrales de 2001 a 2018, es el modelo estacional simple, el cual se adapta a una serie sin tendencia y con constante efecto estacional.

De esa cuenta, el diagnóstico del modelo a través del estadístico *Ljung-Box* = 0 .113 >0.05 se dice que, la serie presenta ruido blanco, es decir media constante igual a cero y no autocorrelación.

En la figura 30 de la función de autocorrelación (ACF) y autocorrelación parcial (PACF), se muestran los residuales en 24 retardos donde no hay presencia de autocorrelación significativa.

Por último, en las figuras 31 y 32 se visualiza el pronóstico 2019, en esta última, se compara el primer y segundo trimestre del 2019 con las cifras reales durante esos meses, para el caso de los siguientes trimestres (2 y 4), el Banco de Guatemala registró los valores a precios de referencia del año 2013; por tal razón, no es posible comparar objetivamente dichos resultados, dado que las series están expresadas con los valores de referencia del 2001.

3.4.2.3. Comercio al por mayor y al por menor

La actividad económica del comercio al por mayor y menor, ha representado el 12.02 % del PIB trimestral, para la estimación del pronóstico de corto plazo, se utilizaron las variaciones inter trimestrales de 2001 al 2018.

Tabla XXXVI. **Modelo de pronóstico de la actividad económica del comercio**

Descripción del modelo			
ID de modelo	6. Comercio al por mayor y al por menor	Modelo_1	Tipo de modelo ARIMA (0,0,1) (0,1,1)

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXVII. **Estadísticos de ajuste del modelo de pronóstico del comercio**

Estadístico de ajuste	Media
R cuadrado-estacionaria	0.450
R cuadrado	0.968
RMSE	2.083
MAPE	53.414
MaxAPE	500.832
MAE	1.438
MaxAE	7.436
BIC normalizado	1.593

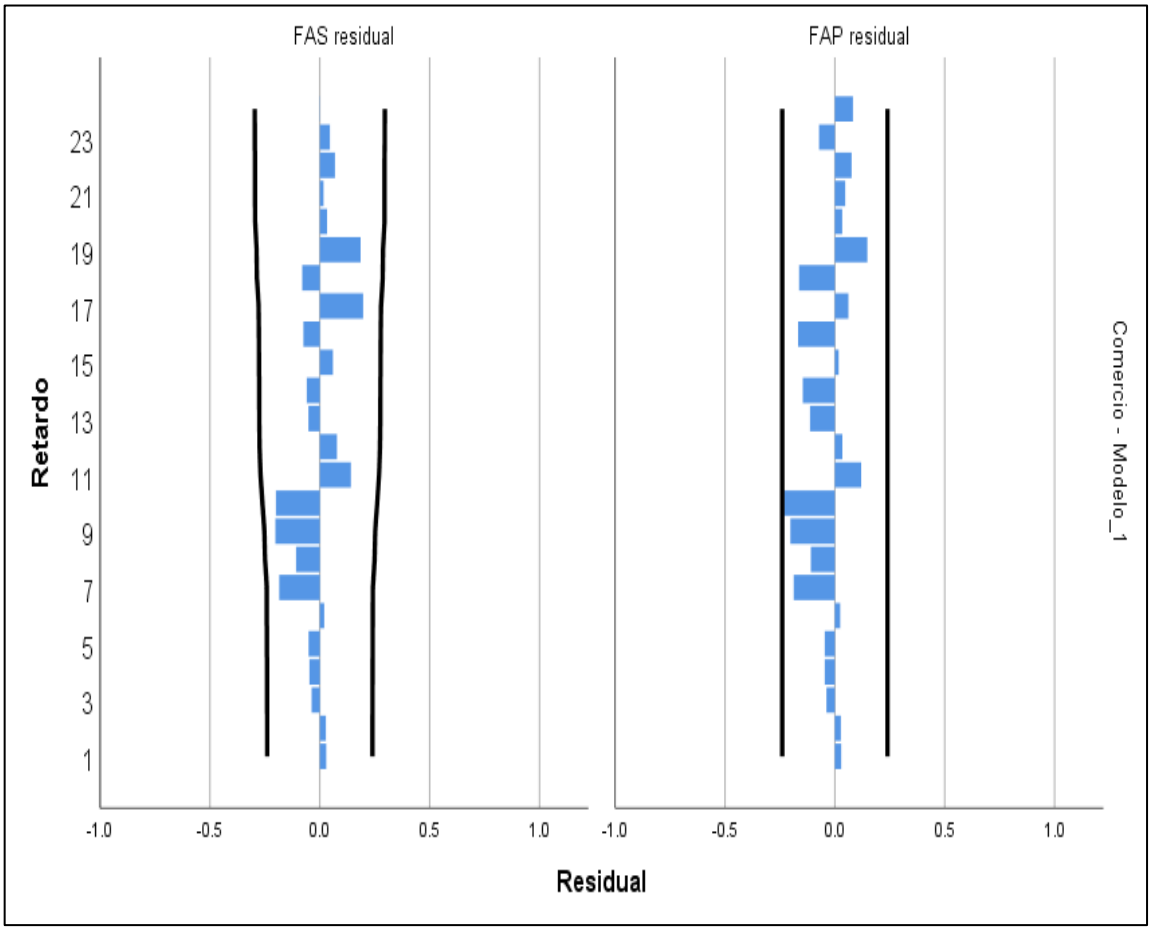
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXVIII. **Diagnóstico del modelo de pronóstico del comercio**

Modelo	Estadísticos del modelo			Número de valores atípicos
	Ljung-Box Q (18)			
	Estadísticos	DF	Sig.	
6. Comercio al por mayor y al por menor	18.372	16	.303	0

Fuente: elaboración propia, realizado conSPSS 25.

Figura 33. **ACF y PACF residuales del modelo, comercio al por mayor y al por menor**



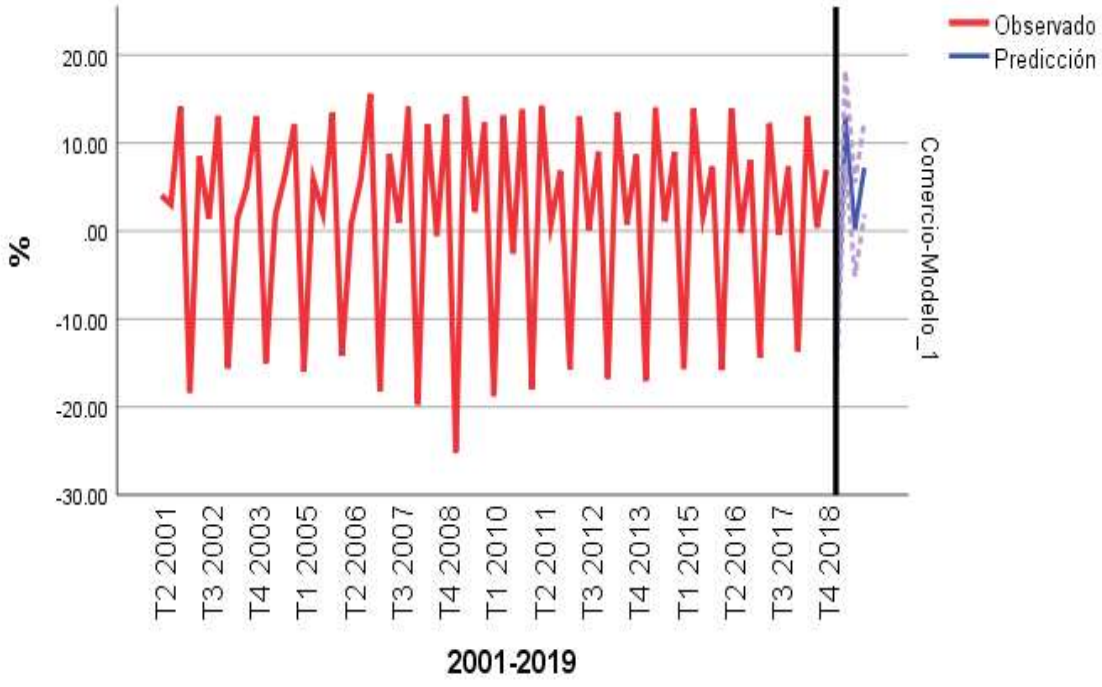
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XXXIX. **Pronóstico año 2019, comercio al por mayor y al por menor**

Modelo		Predicción			
		T1 2019	T2 2019	T3 2019	T4 2019
6. Comercio al por mayor y al por menor-Modelo_1	Predicción	-13.88	12.95	.21	7.19
	UCL	-9.83	18.25	5.51	12.49
	LCL	-17.93	7.65	-5.10	1.89

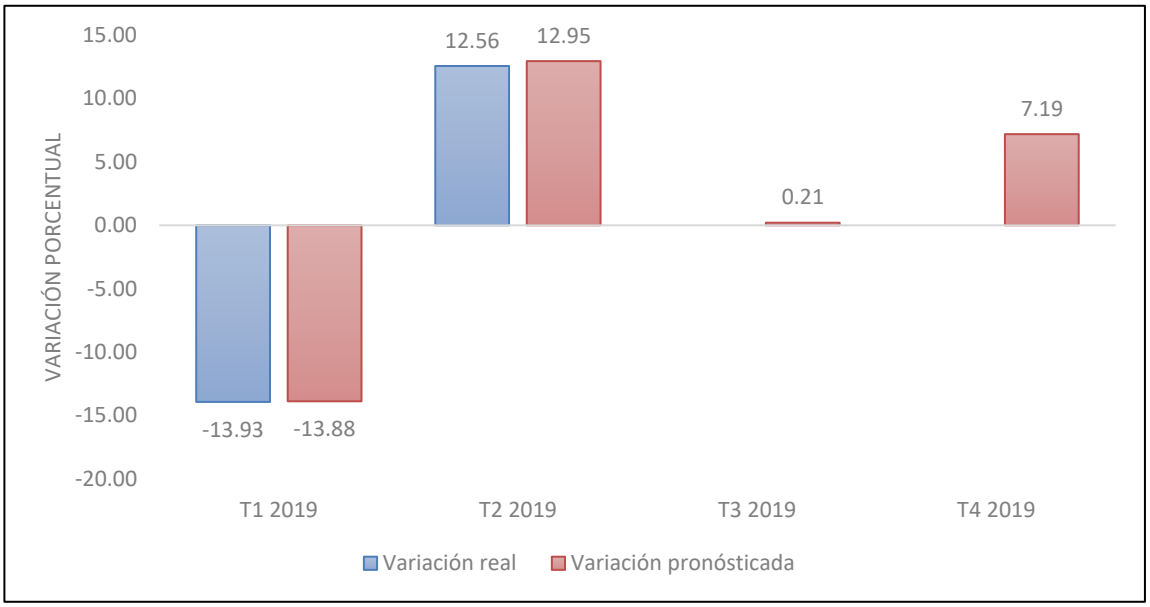
Fuente: elaboración propia, realizado con SPSS 25.

Figura 34. **Pronóstico 2019 de las variaciones de la actividad económica comercio al por mayor y por menor con un nivel del 95 % de confianza**



Fuente: elaboración propia, realizado con SPSS 25.

Figura 35. **Comparación de valores reales vs valores pronosticados de las variaciones de la actividad económica comercio al por mayor y menor 2019, con un nivel de confianza del 95 %**



Fuente: elaboración propia.

El modelo que mejor se ajusta a las variaciones Inter trimestrales de la actividad económica del comercio al por mayor y menor del 2001 al 2018, es el modelo ARIMA de orden (0,0,1) y (0,1,1), con cero autorregresivos y diferencias y con una media móvil en tendencia secular y con el factor estacional, es necesaria una diferencia y una media móvil.

De acuerdo con los estadísticos de ajuste, el modelo presenta un error porcentual absoluto medio (MAPE) del 53.4 %; a pesar de tener un error relativamente alto, el diagnóstico del modelo a través del estadístico *Ljung-Box* = 0.303 > 0.05, los residuales presentan varianza cero y no autocorrelación.

La figura 33 de ACF y PACF de los residuales confirman gráficamente el estadístico *Ljung-Box* mediante 24 rezagos o retardos del modelo.

La tabla XXXIX muestra los valores específicos para cada trimestre del 2019 y sus intervalos a un nivel de confianza del 95 %.

La figura 35, muestra un nivel de exactitud con un excelente ajuste comparado con los valores reales registrados en los primeros dos trimestres de ese año.

Tal como se mencionó anteriormente, no es posible comparar los resultados reales para los trimestres dos y cuatro de ese año, debido a la no disponibilidad de información con los precios de referencia del 2001.

3.4.2.4. Transporte

La actividad económica del transporte, en promedio, representó el 9.07 % del PIB durante los 19 años de estudio, esta actividad constituye una de las cinco que mejor explican las afiliaciones a la seguridad social en Guatemala, se tomaron las variaciones inter trimestrales del 2001 al 2018 para estimar el modelo de pronóstico de mejor ajuste.

Tabla XL. **Modelo de pronóstico de la actividad económica trasporte**

	Descripción del modelo		
			Tipo de modelo
ID de modelo	7. Transporte, almacenamiento y comunicaciones	Modelo_1	Estacional simple

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XLI. **Estadísticos de ajuste modelo de transporte**

Estadístico de ajuste	Media
R cuadrado-estacionaria	0.528
R cuadrado	0.788
RMSE	2.432
MAPE	62.102
MaxAPE	727.514
MAE	1.628
MaxAE	8.079
BIC normalizado	1.897

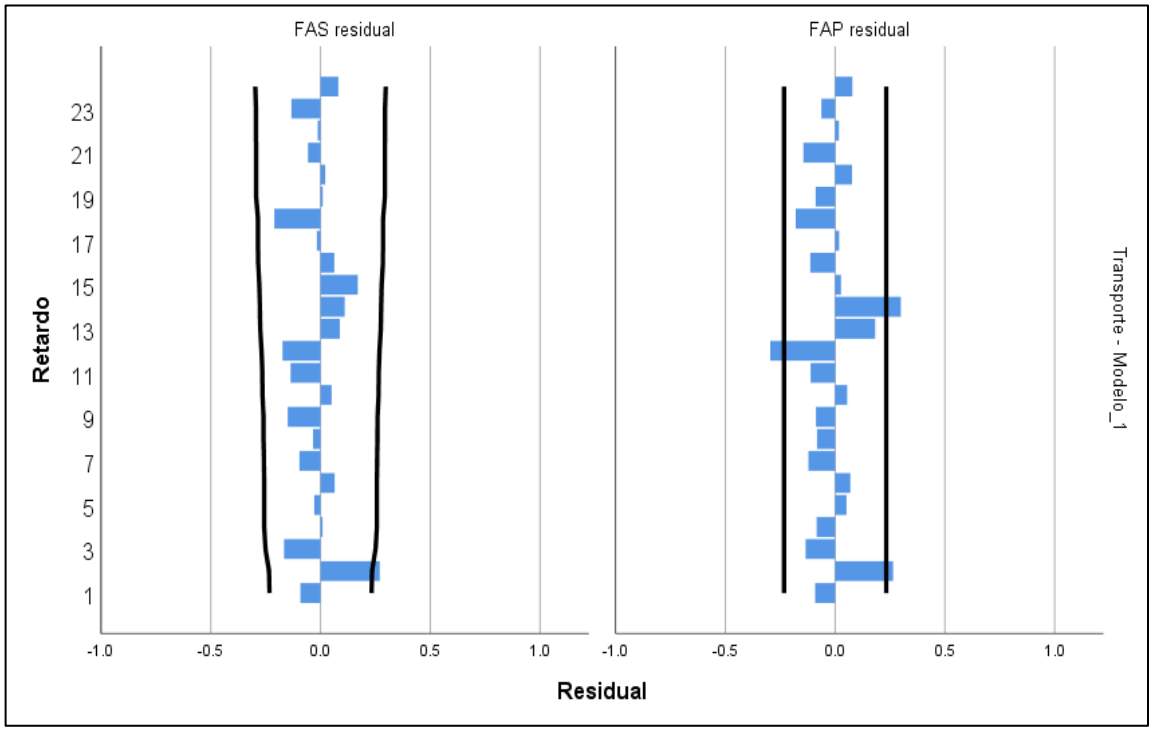
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XLII. **Diagnóstico del modelo de la actividad económica de transporte**

Modelo	Estadísticos del modelo			
	Ljung-Box Q (18)			Número de valores atípicos
	Estadísticos	DF	Sig.	
7. Transporte, almacenamiento y comunicaciones	24.954	16	.071	0

Fuente: elaboración propia, realizado con SPSS 25.

Figura 36. **ACF Y PACF de residuales del modelo de la actividad económica de transporte**



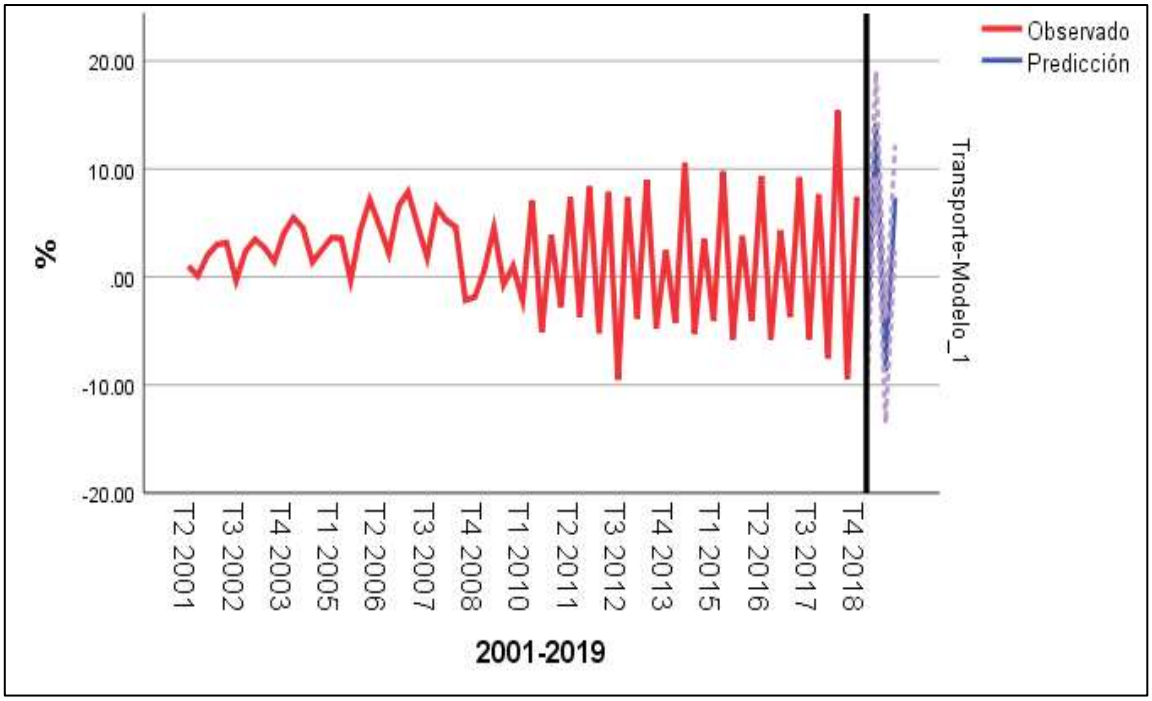
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XLIII. **Pronóstico 2019 de la actividad económica de transporte**

Modelo		Predicción			
		T1 2019	T2 2019	T3 2019	T4 2019
7. Transporte, almacenamiento y comunicaciones	Predicción	-6.75	14.17	-8.70	7.34
	UCL	-1.89	19.02	-3.85	12.20
	LCL	-11.60	9.32	-13.56	2.49

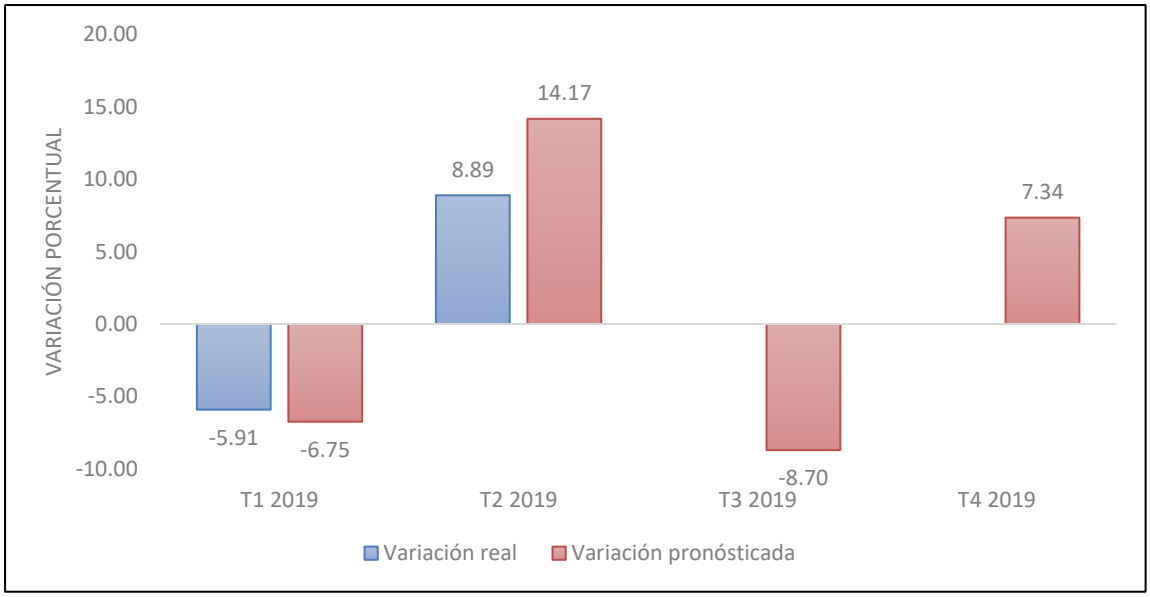
Fuente: elaboración propia, realizado con SPSS 25

Figura 37. **Pronóstico 2019 de las variaciones de la actividad económica de transporte, almacenamiento y comunicaciones con un nivel del 95 % de confianza**



Fuente: elaboración propia, realizado con SPSS 25.

Figura 38. **Comparación de valores reales vs pronóstico 2019 de la actividad económica de transporte, con un nivel del 95 % de confianza**



Fuente: elaboración propia.

Para el caso de la actividad económica de transporte, almacenamiento y comunicación, el modelo que mejor se ajusta a las variaciones inter trimestrales del 2001 al 2018, es el modelo estacional simple. Como se ha mencionado anteriormente, este modelo toma en cuenta la no presencia de tendencia y un factor estacional constante.

Dentro del análisis del modelo, se menciona el error porcentual absoluto medio (MAPE) con un nivel de 62.10 %, el cual es un nivel de error alto para el modelo; sin embargo, mediante el análisis de otros criterios de diagnóstico, se menciona el estadístico *Ljung-Box* =0.071 >0.05, para este caso, el valor de P es muy cercano al nivel de significancia estadística, no obstante, los valores caen dentro del rango de aceptación.

La figura 36 de ACF y PACF contribuye al análisis, dado que, en el retardo 13 y 15 se visualiza autocorrelación en la serie, lo que hace en cierta medida que el modelo carezca de un nivel de incertidumbre que considerar en los pronósticos efectuados.

En la figura 38, la cual es comparativa con los valores reales registrados en los primeros dos trimestres del 2019, se visualiza diferencias significativas, principalmente en el segundo trimestre en donde existe un contraste de 5.28 puntos porcentuales.

Para el presente caso, dada la particularidad de la serie, no se recomienda este tipo de modelo para realizar pronósticos de corto plazo.

3.4.2.5. Administración pública y defensa

La actividad económica de administración pública y defensa representó el 7.25 % promedio del PIB trimestral, es la última actividad económica del modelo de predicción de las afiliaciones a la seguridad social en el período de estudio. Para la estimación del pronóstico de corto plazo, se utilizó la variación inter trimestral de dicha actividad económica del 2001 al 2018.

Tabla XLIV. **Modelo de pronóstico de la actividad económica de administración pública**

Descripción del modelo			
			Tipo de modelo
ID de modelo	11. Administración pública y defensa	Modelo_1	ARIMA (2,0,1) (0,1,1)

Fuente: elaboración propia, realizado con SPSS 25.

Tabla XLV. **Estadísticos de ajuste de la actividad económica de administración pública**

Estadístico de ajuste	Media
R cuadrado-estacionaria	0.556
R cuadrado	0.691
RMSE	3.789
MAPE	194.640
MaxAPE	5052.111
MAE	2.472
MaxAE	13.326
BIC normalizado	3.041

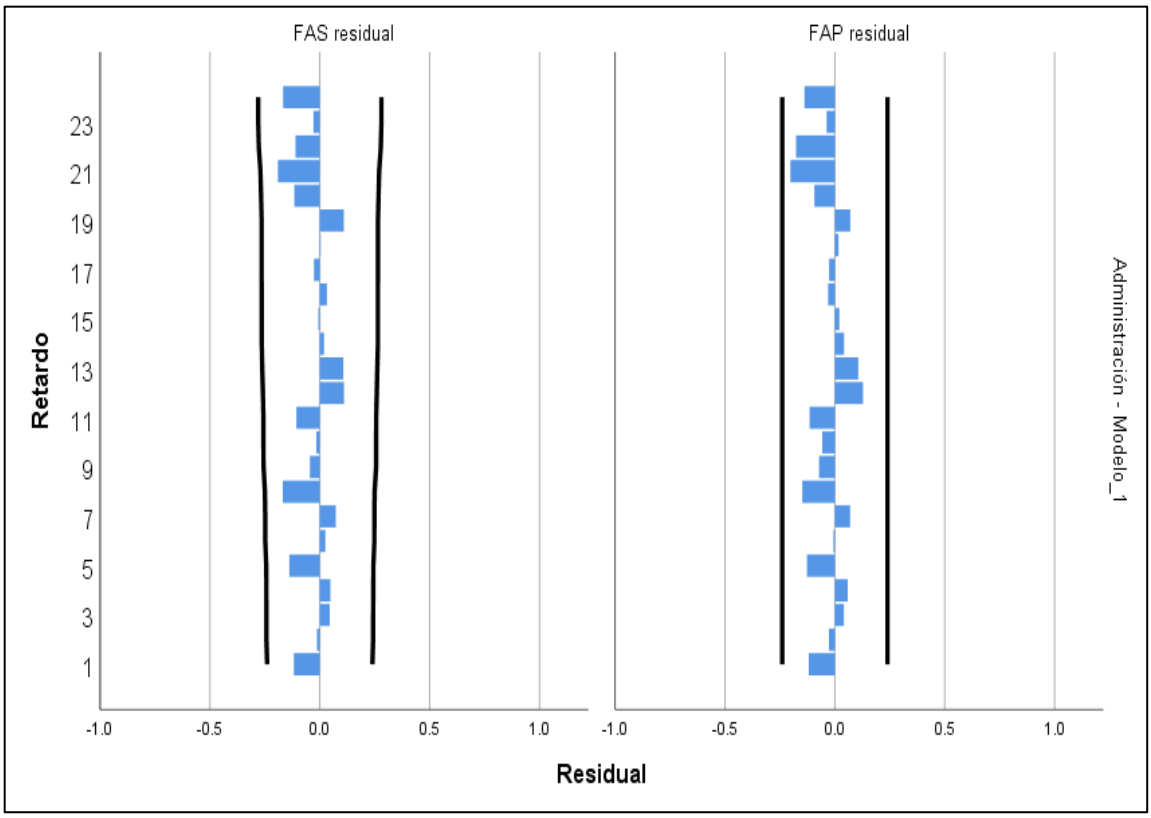
Fuente: elaboración propia, realizado con SPSS 25.

Tabla XLVI. **Diagnóstico del modelo de la administración pública**

Estadísticos del modelo				
Modelo	Ljung-Box Q (18)			Número de valores atípicos
	Estadísticos	DF	Sig.	
11. Administración pública y defensa-Modelo_1	8.745	14	0.847	0

Fuente: elaboración propia, realizado con SPSS 25.

Figura 39. **ACF Y PACF de residuales en modelo de la administración pública y defensa**



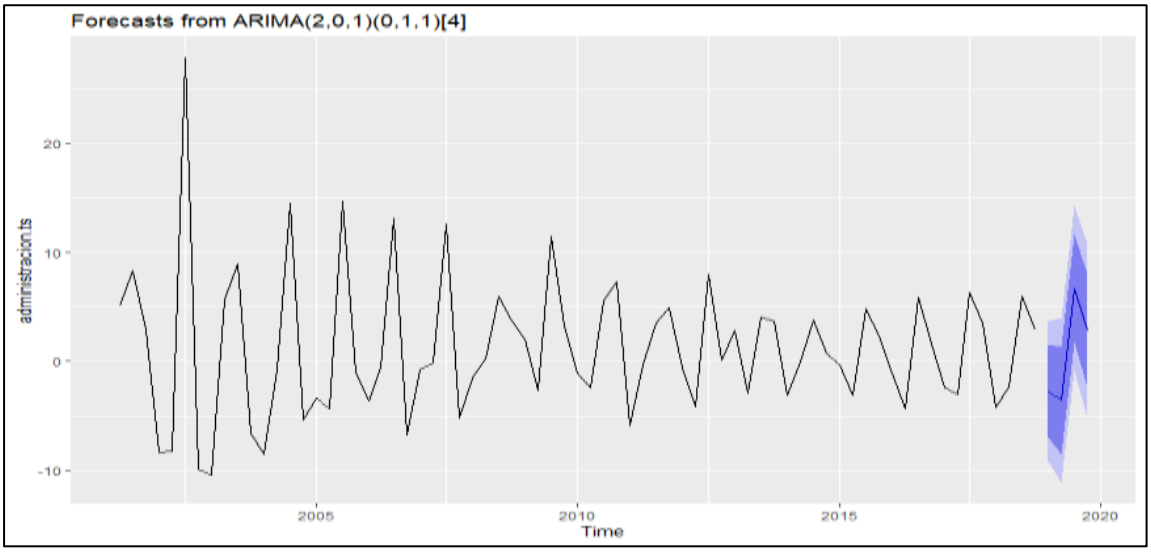
Fuente: elaboración propia, realizado con SPSS 25.

Figura 40. **Pronóstico 2019 de la actividad económica de administración pública y defensa**

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2019 Q1	-2.716973	-6.921333	1.487387	-9.1469868	3.713040
2019 Q2	-3.562420	-8.493721	1.368881	-11.1041945	3.979355
2019 Q3	6.677915	1.743019	11.612810	-0.8693564	14.225186
2019 Q4	2.784592	-2.430892	8.000077	-5.1918028	10.760987

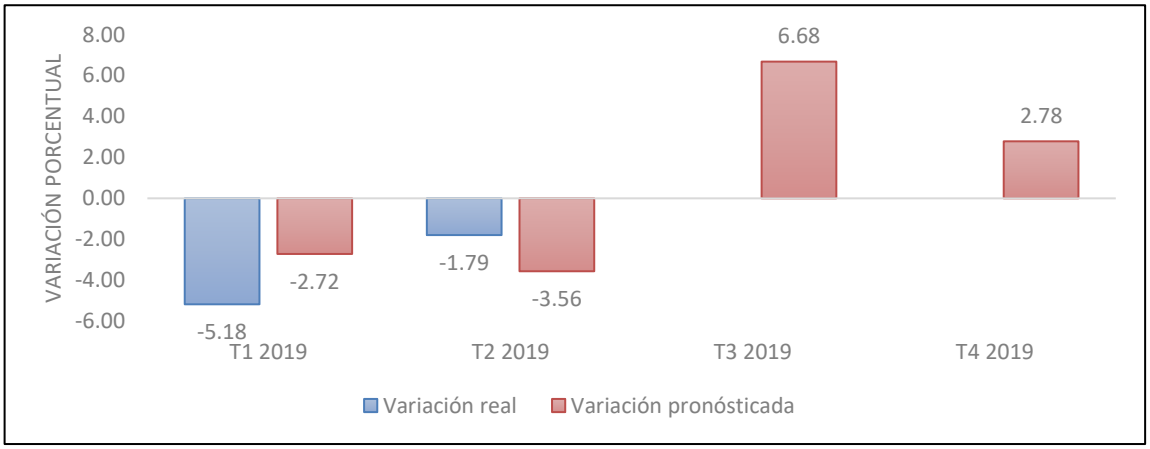
Fuente: elaboración propia, realizado con *Rstudio.*

Figura 41. **Pronóstico 2019 actividad económica de administración pública y defensa con un nivel del 95 % de confianza**



Fuente: elaboración propia, realizado con *Rstudio.*

Figura 42. **Comparación de valores reales vs pronóstico de 2019 de la actividad económica administración pública y defensa**



Fuente: elaboración propia.

La actividad económica de administración pública y defensa constituye la quinta actividad económica que explica las afiliaciones a la seguridad social, de acuerdo con el modelo de predicción presentado en el objetivo 3.

Para la elaboración de este modelo, se determinó el modelo ARIMA de orden 2 autorregresivos, 0 diferencias, 1 media móvil en tendencia secular, y orden 0 autorregresivos,1 diferencia, 1 media móvil en tendencia estacional.

Al comparar los resultados del pronóstico para el año 2019 vs las variaciones raleas registradas en los primeros dos trimestres, se visualiza una diferencia de 2.46 % en el primero y 1.77 % en el segundo. De acuerdo con las distintas pruebas realizadas en los modelos, este modelo representa el que mejor se ajusta a los datos del 2001 al 2018.

3.5. Objetivo general

Estimar el grado de asociación y de dependencia estadística de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala, del 2001 al 2018, mediante el análisis de regresión múltiple y de series de tiempo para construir los modelos de mejor ajuste de predicción y pronósticos en el corto plazo (períodos menores a un año).

De acuerdo con el objetivo general, se pudo constar que, existe asociación estadística (87 %) entre las variaciones inter trimestrales de las afiliaciones a la seguridad social y las variaciones de las 11 actividades económicas de Guatemala, el grado de dependencia estadística es moderadamente alto (65 %) en el modelo final estimado, tomando en consideración las variaciones inter trimestrales de las actividades económicas de agricultura, industria, comercio, transporte y administración pública durante el período 2001-2018.

Cabe mencionar que, al modelo de regresión lineal múltiple estimado, mediante la prueba de linealidad de Ramsey, la cual establece el planteamiento de H0: existe linealidad en el modelo y H1: no existe linealidad en el modelo, se realizó diagnóstico de linealidad donde se obtuvo el siguiente resultado:

Un valor P = 0.02445, el cual es menor a 0.05; por tal razón, el modelo estimado es una aproximación a la realidad que debe de profundizarse mediante un análisis de modelos no lineales.

De acuerdo con los modelos de pronóstico estimados para las afiliaciones a la seguridad social y las cinco actividades económicas de Guatemala que mejor predicen las afiliaciones a la seguridad social, se pudieron comprobar los resultados de los pronósticos realizados vs los registros de los trimestres del año 2019, encontrando cifras muy cercanas entre sí.

4. DISCUSIÓN DE RESULTADOS

El objetivo de la investigación fue estimar modelos de pronóstico de corto plazo de las variables que intervienen en el primer modelo de predicción, con el fin de contar con información estadística oportuna, basada en el conocimiento de los posibles escenarios en la cobertura de la seguridad social en función del crecimiento económico dado a través de las actividades económicas.

4.1. Análisis interno

En la primera fase del análisis exploratorio y descriptivo de las variables de estudio, fue posible conocer el comportamiento de las variables, su distribución, los estadísticos que mejor describen la serie, la prueba de datos atípicos que podrían ocasionar complicaciones en las predicciones y pronósticos finales con sus medidas de tendencia central.

Utilizando la estadística descriptiva, en los datos de las once actividades del PIB; se pudo constatar que, las variaciones inter trimestrales de las actividades económicas de minas y canteras, suministro de electricidad y captación de agua y transporte mostraron distribución normal en sus datos.

Para el establecimiento del grado de fuerza y asociación entre las variaciones inter trimestrales de las afiliaciones de la seguridad social y las variaciones de las 11 actividades económicas, se infiere que existe una fuerte asociación lineal expresada en el coeficiente de *Pearson* (R) de 0.872.

Se construyó el modelo de regresión lineal múltiple que explica con la menor cantidad de variables regresoras posibles, la mayor cantidad de información sobre la variación de la variable dependiente (principio de parsimonia).

En ese orden de ideas, se propone el modelo de predicción de mejor ajuste que incluye cinco variaciones Inter trimestrales de las actividades económicas como variables que explican las variaciones promedio de las afiliaciones a la seguridad social en 2001 al 2018.

Es importante mencionar que, durante las pruebas finales del diagnóstico del modelo de regresión lineal múltiple; a través de la prueba de Ramsey que permite identificar las especificaciones del modelo, particularmente la linealidad; se pudo establecer que un modelo de regresión lineal múltiple no es el mejor para establecer predicciones; no obstante, es una aproximación a la realidad muy valiosa que debe considerarse para hacer un análisis comparativo y profundizar en modelos no lineales.

En consecuencia, posterior a establecer el modelo de predicción, se estimaron modelos de pronósticos de corto plazo, en períodos menores a un año de acuerdo con las variables de mayor importancia en los modelos previamente definidos.

Para el caso de las afiliaciones de la seguridad social, se estimaron dos pronósticos, el primero tomando las series acumuladas, logrando hacer comparaciones entre variables del mismo tipo, la ventaja de este tipo de análisis es que, existe información disponible de años posteriores al estudio, que permitieron comparar los resultados del pronóstico con los registros reales.

El segundo modelo, se estimó con base en las variaciones inter trimestrales de las afiliaciones de la seguridad social, logrando así, pronósticos muy cercanos a los datos reales del 2019.

Para el caso de la comparación de los pronósticos de las actividades económicas, se presentó la dificultad de información; dado que, durante los trimestres posteriores al segundo trimestre 2019, las estimaciones del PIB que realiza el Banco de Guatemala por el origen de la producción registraron cambios en los precios de referencia, calculando los rubros de las actividades económicas a precios de referencia del año 2013.

Por la razón antes expuesta, solo fue posible comparar los resultados en los primeros dos trimestres del 2019; no obstante, los resultados de los pronósticos estimados fueron muy cercanos a los valores reales de esos trimestres.

4.2. Análisis externo

En el contexto del análisis externo, se menciona las oportunidades y herramientas disponibles en la actualidad, la potencia de los Software estadísticos es tal que, grandes y complejos cálculos se pueden realizar en pocos segundos; sin embargo, es importante que el conocimiento del investigador sea sólido en el uso de las herramientas para la correcta interpretación de los resultados.

Por ello, para el presente trabajo de investigación se recurrió al reforzamiento de la teoría y la exposición de trabajos de académicos clásicos y de actualidad para la interpretación de las diferentes pruebas realizadas.

Para el caso de la seguridad social, al registro desagregado por actividad económica de los afiliados no está contemplado en el presente estudio, dado que no fue la intención establecer regresiones canónicas, es decir cuando existe más de una variable dependiente.

En ese sentido, puede considerarse un reto de corto plazo, realizar estudios dende se incluya el registro de afiliados por actividad económica y la diferencia entre el crecimiento de cada actividad del sector productivo del país.

En el caso de las variaciones de las actividades económicas, se analizó en su conjunto la explicación a todas las afiliaciones a la seguridad; no obstante, se pudo determinar que no todas las actividades agregaron información al modelo, algunas resultaron sobre explicadas, y dos actividades que resultaron significativas, su relación fue inversa; es decir, a medida que aumentaban sus variaciones inter trimestrales disminuía las variaciones de los afiliados al seguro social.

En lo que respecta a la comparación de trabajos previos sobre el análisis de regresión múltiple, se cita el trabajo desarrollado por Julio Meneses (2019), en donde se consideró importante el cumplimiento de cada supuesto, mediante ello, es de mucha relevancia no obviar los resultados obtenidos, para no incurrir en errores de interpretación.

Así mismo, las comparativas realizadas por Rosas et, al (2006), en cuanto a los coeficientes de determinación (R^2 Ajustado) y los estadísticos de *Cp de Mallows* ayudaron a seleccionar la mejor modelo dada las comparaciones realizadas entre tres diferentes.

Para la estimación de los pronósticos, fue de utilidad el trabajo de Páez, et, al (2016), en el proceso de comparaciones entro los coeficientes de cada modelo, tales como: el R^2 ajustado, el error cuadrado medio (MSE), los criterios de información de *Akaike* y *Bayesiano* y el error absoluto porcentual. Principalmente el error porcentual absoluto o MAPE, la teoría indica que, este estadígrafo, muestra el error porcentual del modelo; sin embargo, al comparar las cifras reales vs el pronóstico, no siempre se cumplió el supuesto que a menor error mayor precisión en las cifras pronosticadas; en ese sentido, el criterio de selección de un modelo de acuerdo con el MAPE es recomendable cuando se trabajan con cifras con tendencia serial (acumuladas) más no, con variaciones inter trimestrales.

CONCLUSIONES

1. Existe una fuerte asociación lineal entre las variaciones inter trimestrales de las afiliaciones a la seguridad social y variaciones de las actividades económicas de Guatemala durante el período 2001-2018, expresado por el coeficiente de correlación de *Pearson* (0.872). La actividad económica que tiene mayor asociación lineal con las afiliaciones a la seguridad social es la agricultura, ganadería y caza; así mismo, existe grado moderadamente fuerte (R^2 Ajustado = 0.68) de dependencia estadística entre las variaciones inter trimestrales de las afiliaciones a la seguridad social respecto de las variaciones de las once actividades económicas de Guatemala en el período 2001-2018; de las cuales, se mencionan las siguientes: Minas y Canteras, Comercio al por Mayor y Menor, Transporte, Almacenamiento y Comunicaciones e Industria Manufacturera.

2. Según la muestra analizada, se infiere que el modelo general entre las afiliaciones a la seguridad social en Guatemala respecto de las actividades económicas de Guatemala en el período 2001-2018; cumple con los supuestos en los residuos. Así mismo, cumplen los supuestos de análisis de series de tiempo; no obstante, se evidencia que no todas las variables de estudio registran estacionariedad, independencia y homogeneidad.

3. El modelo de mejor ajuste de predicción de las variaciones inter trimestrales de las afiliaciones a la seguridad social respecto de las variaciones inter trimestrales de las actividades económicas de Guatemala está expresado por las variaciones de las actividades económicas de

agricultura, industria, comercio, transporte y administración pública como variables independientes.

4. Los mejores modelos de pronósticos construidos para las seis variables finales de estudio fueron comparados con las cifras reales de las variaciones inter trimestrales de los cuatro trimestres del 2019 en lo que respecta a la seguridad social y de dos trimestres de las variaciones inter trimestrales de las actividades económicas, encontrando aproximaciones valiosas a lo registrado durante esos períodos.

5. De acuerdo con las estimaciones realizadas, existe una fuerte asociación y dependencia estadística lineal en las afiliaciones a la seguridad social y las actividades económicas de Guatemala en el período 2001-2018, el modelo de mejor ajuste de predicción fue a través de 5 de las 11 actividades económicas del estudio y, los pronósticos realizados a 6 variables finales de estudio brindaron valiosa información cercana al comportamiento real de los trimestres del 2019.

RECOMENDACIONES

1. Tomar en consideración el grado de asociación y de dependencia estadística de las variaciones inter trimestrales de las afiliaciones a la seguridad social respecto a las variaciones inter trimestrales de las actividades económicas de Guatemala en el entendido que, se evidencia una fuerte asociación y dependencia estadística que puede permitir ampliar los estudios según lo consideren conveniente para al Instituto Guatemalteco de Seguridad Social.

2. Utilizar el presente estudio como base para que los estudiantes, centros académicos, tanques de pensamiento y autoridades del Instituto Guatemalteco de Seguridad Social, profundicen en estimación de modelos de predicción más allá del análisis de regresión lineal múltiple que considere los diferentes escenarios bajo los supuestos de la no linealidad de los parámetros.

3. Impulsar los estudios en el instituto Guatemalteco de Seguridad Social, de las causas que establecen la relación positiva entre las variaciones inter trimestrales de las actividades económicas de la agricultura, industria y transporte; así mismo, identificar las posibles causas del porqué ante el incremento de las variaciones relativas del comercio y administración pública, existe una disminución de las variaciones relativas de afiliados a la seguridad social.

4. Considerar los modelos utilizados en el presente estudio, para que el Instituto Guatemalteco de Seguridad Social, replique los pronósticos de

períodos necesarios, a fin de incluir dentro de sus proyectos de políticas internas, sustento técnico que permita la ampliación de la cobertura de la seguridad social en Guatemala.

5. Utilizar el presente estudio como base, para que el Instituto Guatemalteco de Seguridad Social, estudiantes, centros académicos y tanques de pensamiento, puedan realizar predicciones y pronósticos de corto plazo; sobre las predicciones, utilizar modelos de no linealidad en los parámetros que permitan comparar cifras entre resultados y contar con aproximaciones que permitan ampliar los conocimientos técnicos hacia ampliar la cobertura de la seguridad social de la población guatemalteca Instituto Guatemalteco de Seguridad Social, estudiantes, centros académicos y tanques de pensamiento.

REFERENCIAS

1. Abuín, J. R. (2007). *Regresión lineal múltiple*. Estadística, Editor, 32.

2. Alaminos, E., & Ayuso Gutiérrez, M. (2015). *Una estimación actuarial del coste individual de las pensiones de jubilación y viudedad: Concurrencia de pensiones del Sistema de la Seguridad Social español*. Estudios de Economía Aplicada, 33(3), 817–838.

3. Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2008). *Estadística para administración y economía, Cengace Learning Editores, S.A*. recuperado de https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/estadistica-para-administracion-y-economia_anderson_sweeney_y_williams.pdf

4. Arroyo Delgado, I. (2019). *Predicción de precipitaciones con dispositivos IoT mediante análisis de series temporales*. B.S. Thesis, Universidad Politécnica de Madrid (2019).

5. Asamblea Nacional Constituyente. (1985). *Constitución Política de la República de Guatemala, reformada por Cónsula Popular* (Acuerdo Legislativo 18-93). Congreso de La República de Guatemala, 18, 76. Recuperado de https://www.oas.org/juridico/mla/sp/gtm/sp_gtm-int-text-const.pdf

6. Avella, M., & Fergusson, L. (2004). *El ciclo económico: enfoques e ilustraciones. Los ciclos económicos de Estados Unidos y Colombia. Borradores de Economía*; No. 284, 1–78. Recuperado de http://repositorio.banrep.gov.co/handle/20.500.12134/5302

7. Balsells, E. (2020). *La enfermedad renal crónica de causa no tradicional (ERCnT): una visión desde la protección social y la economía preventiva. Ciencia, Tecnología y Salud*, 7(1). Recuperado de https://doi.org/10.36829/63cts.v7i1.896

8. Berlanga, S. C. (2014). *La imputación múltiple y su aplicación a series temporales financieras*.

9. Briscoe, I. (2009). *El Estado y la seguridad en Guatemala.* Fride Working Paper 88, February, 36.

10. Brugger Methol, A., & Merello Sáder, L. (2010). *Determinantes socioeconómicos del suicidio en Uruguay: Una Aproximación a través de series temporales*.

11. Caridad Walkiria Guerra Bustillo, D. I., Magaly Herrera Villafranca, D. I., Yasser Vázquez Alfonso III, D., & Atenaida Beatriz Quintero Bueno, D. (2014). *Contribución de la Estadística al análisis de variables categóricas: Aplicación del Análisis de Regresión Categórica en las Ciencias Agropecuarias* Statistical contribution to the analysis of categorical variables: Application of Categorical Regression Ana. Revista Ciencias Técnicas Agropecuarias, 23(1), 68–73. Recuperado de http://scielo.sld.cu/pdf/rcta/v23n1/rcta12114.pdf

12. CARLOS MANTEROLA, D., VIVIANA PINEDA, N., & MINCIR, G. (2008). *El valor de "p" y la "significación estadística". Aspectos generales y su valor en la práctica clínica:* Interpretation of medical statistics. Revista Chilena de Cirugía, 60(1), 86–89. Recuperado de https://doi.org/10.4067/S0718-40262008000100018

13. Castro, E. M. M. (2019*). Bioestadística aplicada en investigación clínica: conceptos básicos. Revista Médica Clínica Las Condes*, 30(1), 50–65. Recuperado de https://doi.org/10.1016/j.rmclc.2018.12.002

14. Cienfuegos, M., & Cienfuegos, A. (2016*). Lo cuantitativo y cualitativo en la investigación. Un apoyo a su enseñanza. Revista Iberoamericana Para La Investigación y El Desarrollo Educativo*, 7(13), 1–11. Recuperado de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-74672016000200015 %0Ahttp://www.scielo.org.mx/pdf/ride/v7n13/2007-7467-ride-7-13-00015.pdf

15. Curcio, J., Goldschmit, A., Velásquez Pinto, M. D., & Leguizamón, G. B. (n.d.). Este documento fue preparado por Oscar Cetrángolo, experto de la CEPAL.

16. Del Valle Moreno, V., Walkiria, C., & Bustillo, G. (2012). *La Multicolinealidad en modelos de Regresión Lineal Múltiple. Revista Ciencias Técnicas Agropecuarias*, 21(4), 80–83.

17. Devore, J. (2008). *Probabilidad y estadística para ingeniería y ciencias*. In Cengage Learning: Vol. 7th Ed.

18. Diaz-Tendero Bollain, A. (2017). *Pensiones y estratificación social en América Central y las Antillas Mayores*. Latinoamérica. Revista de Estudios Latinoamericanos, 65. Recuperado de https://doi.org/10.22201/cialc.24486914e.2017.65.56842

19. Enrique, J., & Grajeda, B. (2020). *Construcción de series temporales para evaluar un sistema de pensiones en México* Time series construction for assessing a pension system in Mexico. 71–84.

20. Eslava, J. (2001*). El sistema general de seguridad social en salud*. Revista de La Facultad de Medicina, 49(1).

21. Espinoza, S. D. M. (2015). *Modelos de regresión para el pronóstico de series temporales con estacionalidad creciente. Computación y Sistemas*, 18(4), 821–831. Recuperado de https://doi.org/10.13053/CyS-18-4-1552

22. Exterior, C. (1992). *Desestacionalización de series de tiempo económicas: ajustes previos.* 42(1), 1042–1053.

23. F., G. L., Padilla, O. U., Sánchez-Pérez, H. J., & León-Cortés, J. L. (2017). *La seguridad social de los trabajadores al servicio del Estado: "crónica de una muerte anunciada."* Estudios Políticos, 41. Recuperado de https://doi.org/10.1016/j.espol.2017.05.002

24. Felipe, H., Correa, V., Luna, L., & Ruiz, F. (2002). *Desestacionalización de Series Económicas: El procedimiento usado por el Banco Central*

de Chile. Documentos de Trabajo, 177, 43. Recuperado de http://si2.bcentral.cl/public/pdf/documentos-trabajo/pdf/dtbc177.pdf

25. Gammage, S. (2009). *Género, pobreza de tiempo y capacidades en Guatemala*: *un análisis multifactorial desde una perspectiva económica*. Cepal.

26. Guerguil, M. (1988). *Algunos alcances sobre la definición del sector informal. Revista de La CEPAL*, 1988(35), 55–63. Recuperado de https://doi.org/10.18356/ad6c44a8-es

27. Guerra, Caridad W., & Cabrera, A., & Fernández, Lucía (2003). *Criterios para la selección de modelos estadísticos en la investigación científica. Revista Cubana de Ciencia Agrícola*, 37(1),3-10. [fecha de Consulta 26 de Julio de 2022]. ISSN: 0034-7485. Disponible en: recuperado de https://www.redalyc.org/articulo.oa?id=193018072001

28. Instituto, A. L., Progreso, E. L., Isaac, P., & Fuentes, G. (2014). Guatemalteco De Seguridad Social, 7008.

29. Juan, P. (2015). *Modelos de estimación del requerimiento de liquidez para el pago de prestaciones en una entidad de seguridad social*. Recuperado de https://repositorio.uasb.edu.ec/handle/10644/4850

30. Katz, J. (2007). *Desarrollo Productivo.* In director (Issue C).

31. *LA SEGURIDAD SOCIAL EN EL DERECHO DE INTEGRACIÓN SUBREGIONAL DE AMÉRICA LATINA Y EL CARIBE*. (2019). Revista Latinoamericana de Derecho Social, 28.

32. MANTEROLA D, C., & PINEDA N, V. (2008). *El valor de p; y la significación estadística: Aspectos generales y su valor en la práctica clínica.* Revista Chilena de Cirugía, 60(1), 86–89. Recuperado de https://doi.org/10.4067/S0718-40262008000100018

33. Márquez Cebrián, M. D. (2002). *Modelo setar aplicado a la volatilidad de la rentabilidad de las acciones: algoritmos para su identificaci*ón. Universidad Politécnica de Catalunya.

34. Mendizábal Bermúdez, G. (2020). *Informe: seguridad social en México*, 2019. Revista Latinoamericana de Derecho Social, 1(37). Recuperado de https://doi.org/10.22201/iij.24487899e.2020.37.14873

35. Mesa Lago, C. (1985). *El Desarrollo de la Seguridad Social en América Latina.* 10, 8126–8127.

36. Mesa-Lago, C., Barrios Escobar, M. A., & Fuentes Aragón, H. (1997). *La seguridad social en Guatemala: diagnóstico y propuesta de reforma*. 149 p.

37. Ministerio del Trabajo. (2021). *¿Qué es la Seguridad Social? Formalización y Protección Del Empleo,* Recuperado de https://www.mintrabajo.gov.co/empleo-y-

pensiones/empleo/subdireccion-de-formalizacion-y-proteccion-del-empleo/que-es-la-seguridad-social

38. Moyn, S., & Moyn, S. (2019). *Derechos Humanos: Derecho y Cambio Social En La Historia,* 235–252. Recuperado de https://doi.org/10.2307/j.ctvr7fcmg.13

39. Nazco Torres, W. (2018). *Almacenamiento y Visualización de Series Temporales.*

40. Neffa, J. C., Panigo, D. T., Pérez, P. E., & Persia, J. (2014). *Actividad, empleo y desempleo: conceptos y definiciones.* Clasco, http://bib, 160. Recuperado de http://biblioteca.clacso.edu.ar/Argentina/ceil-conicet/20171027042035/pdf_461.pdf

41. OECD, of Tax Administrations, I.-A. C., for Latin America, E. C., the Caribbean, & Bank, I.-A. D. (2020). Revenue Statistics in Latin America and the Caribbean 2020. In Organización para la Cooperación y el Desarrollo Económicos.

42. Olivero, N. (1997). *Modelos Arima*. Revista Científica Ciencia Médica, 1, 23–30.

43. Páez, E. G. M., González, F. E. C., & Duarte, C. A. M. (2016). *Aplicación de series de tiempo en la realización de pronósticos de producción, El reventón energético*, 14(1), 79-88.

44. Piloña, G. (2016). *Guía Práctica de Investigación de Campo Piloña*.pdf (pp. 1–336).

45. Pittsburgh, U. De. (2008). *LA EXTENSIÓN DE LA COBERTURA DE LA SEGURIDAD SOCIAL* challenge for Latin America in the 21st century: the extension of Social Security coverage. Security, 67–81.

46. Prof, L., & Rodríguez, C. E. (2009). *Diccionario de economía*. 1–117.

47. Robles, J. N. (2010). *Hacia un nuevo modelo de seguridad social*. Economía-unam, 7(20), 7–33. Recuperado de https://doi.org/10.22201/fe.24488143e.2010.20.166

48. Rodríguez, M. (2001). *Análisis de Regresión Múltiple. Estadística Informática: Casos y Ejemplos Con El SPSS*, 3–17. Recuperado de https://doi.org/84-7908-638-6

49. Rosas, M, Chacín, F, García, J, Ascanio, M, & Cobo, M. (2006). *Modelos de regresión lineal múltiple en presencia de variables cuantitativas y cualitativas para predecir el rendimiento estudiantil*. Revista de la Facultad de Agronomía, 23(2), 197-214. Recuperado en 27 de julio de 2022, de recuperado de http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0378-78182006000200007&lng=es&tlng=es.

50. Santos, F., Amador, L., Amaya, N., & Menjívar, J. S. (2020). *EFICIENCIA DEL GASTO PÚBLICO EN SEGURIDAD CIUDADANA: UN ANÁLISIS PARA CENTROAMÉRICA.* Economía y Administración (E&A), 11(2). Recuperado de https://doi.org/10.5377/eya.v11i2.10518

51. Scheaffer, R., & Mendenhall, W. (2007). *Elementos de Muestreo*. Recuperado de https://issuu.com/hectorm.delossantos-posadas/docs/elementos_de_muestreo_-_schaffer_et

52. *Seguridad social en Guatemala*. (2006). Revista Latinoamericana de Derecho Social, 0(2). Recuperado de https://doi.org/10.22201/iij.24487899e.2006.2.9453

53. Stallings, B., & Peres Núñez, W. (2000). Crecimiento, empleo y equidad: el impacto de las reformas económicas en América Latina y el Caribe. CEPAL.

54. Tusell, F. (2011). *Análisis de Regresión, Introducción Teórica y Práctica basada en R*.

55. Vega Vilca, J. C., & Guzmán, J. (2011). *Regresión PLS y PCA Como Solución al Problema de Multicolinealidad en Regresión Múltiple.* Revista de Matemática: Teoría y Aplicaciones, 18(1), 9. Recuperado de https://doi.org/10.15517/rmta.v18i1.2111

APÉNDICES

Apéndice 1. **Matriz de coherencia**

PROBLEMA	PREGUNTAS DE INVESTIGACIÓN	OBJETIVOS	RESULTADOS
PROBLEMA PRINCIPAL	PREGUNTA GENERAL	GENERAL	GENERAL
No se conocen los modelos de mejor ajuste de predicción y pronóstico, realizado con el análisis de regresión múltiple y de series de tiempo entre las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018.	¿Cuáles son los modelos de mejor ajuste de predicción y pronóstico realizado con el análisis de regresión múltiple y de series de tiempo, para las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018?	Estimar el grado de asociación y dependencia estadística de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala, del 2001 al 2018, mediante el análisis de regresión múltiple y de series de tiempo para construir los modelos de mejor ajuste de predicción y pronósticos en el corto plazo (períodos menores a un año).	De acuerdo con el objetivo general, se pudo constar que, existe asociación estadística (87 %) entre las variaciones inter trimestrales de las afiliaciones a la seguridad social y las variaciones de las 11 actividades económicas de Guatemala, el grado de dependencia estadística es moderadamente alto (65 %) en el modelo final estimado, tomando en consideración las variaciones inter trimestrales de las actividades económicas de

Continuación apéndice 1.

1. ¿Cuál es el grado de asociación y de dependencia estadística entre las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018? 2. ¿Se cumplen los supuestos del análisis de regresión múltiple y de series de tiempo en el registro de las afiliaciones a la seguridad social y el registro de las actividades económicas de Guatemala en el período 2001-2018? 3. ¿Cuál es el modelo de predicción de mejor ajuste de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala en el período 2001-2018?	1.Estimar el grado de fuerza o asociación y dependencia estadística entre las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala entre los años 2001 y 2018 a través del análisis de correlación y de regresión múltiple. 2. Comprobar los supuestos del análisis de regresión múltiple y de series de tiempo de las afiliaciones a la seguridad social y de las actividades económicas de Guatemala del período de 2001 a 2018 a través de las pruebas estadísticas para cada tipo de análisis. 3. Estimar el modelo de predicción de mejor ajuste de las afiliaciones a la seguridad social respecto de las actividades económicas de Guatemala entre los años 2001 y 2018 a través del diagnóstico y validación de cada modelo.	agricultura, industria, 1. Existe una fuerte asociación lineal entre las variaciones inter trimestrales de las afiliaciones a la seguridad social y variaciones de las actividades económicas de Guatemala durante el período 2001-2018, expresados por el coeficiente de correlación de Pearson (0.872). 2. Se cumplen con los supuestos en los residuos. Así mismo, cumplen los supuestos de análisis de series de tiempo; no obstante, se evidencia que no todas las variables de estudio registran estacionariedad, independencia y homogeneidad. 3. El modelo de mejor ajuste está expresado por las variaciones de las actividades económicas de agricultura, industria, comercio, transporte y administración pública como variables independientes.

Continuación anexo 1.

4. ¿Cuál es el modelo de pronóstico de mejor ajuste de corto plazo de las afiliaciones a la seguridad social y de las actividades económicas en Guatemala en el período 2001-2018? CONCLUSIÓN GENERAL Existe una fuerte asociación y dependencia estadística lineal en las afiliaciones a la seguridad social y las actividades económicas de Guatemala en el período 2001-2018, expresado por el coeficiente de Pearson (0.872) el modelo de mejor ajuste de predicción fue a través de 5 de las 11 actividades económicas del estudio y, los pronósticos realizados a 6 variables finales de estudio brindaron valiosa información cercana al comportamiento real de los trimestres del 2019.	4. Estimar el modelo de pronóstico de mejor ajuste a corto plazo de las afiliaciones a la seguridad social y de las actividades económicas en Guatemala según los datos de los años 2001 a 2018 a través del diagnóstico y validación del modelo. RECOMENDACIONE GENERAL Utilizar el presente estudio como base, para que el Instituto Guatemalteco de Seguridad Social, estudiantes, centros académicos y tanques de pensamiento, puedan realizar predicciones y pronósticos de corto plazo; sobre las predicciones, utilizar modelos de no linealidad en los parámetros que permitan comparar cifras entre resultados y contar con aproximaciones que permitan ampliar los conocimientos técnicos hacia ampliar la cobertura de la seguridad social	4. Existen 6 modelos de pronóstico que se ajustan a los datos de las variaciones Inter trimestrales y se Encuentra buena aproximación comparando cifras con los trimestres del 2019

Fuente: elaboración propia.