



Universidad Nacional  
**Federico Villarreal**

**VRIN** | VICERRECTORADO  
DE INVESTIGACIÓN

**ESCUELA UNIVERSITARIA DE POSGRADO**

**EL ANÁLISIS DE REGRESIÓN LINEAL Y LA PRECISIÓN DEL PRONÓSTICO  
DE DEMANDA DE PRODUCTOS HIDROBIOLÓGICOS CONGELADOS EN EL  
PERÚ**

**Línea de Investigación:**

**Competitividad Industrial, Diversificación Productiva y Prospectiva**

**Tesis para Optar el Grado Académico de Maestro en Gerencia de Proyectos de  
Ingeniería**

**Autor:**

**Montoya Manrique, Israel Manuel**

**Asesor:**

**Breña Oré, Jorge Luis**

**(ORCID: 0000-0001-6450-7052)**

**Jurado:**

**Flores Vidal, Higinio Exequiel**

**Vales Carrillo, Jorge Alberto**

**Ochoa Sotomayor, Nancy Alejandra**

**Lima – Perú**

**2022**

**Dedicatoria**

A mis padres, Gloria y Jorge, y a mi tío Hugo, que  
en paz descansen, por todas y cada una de las  
lecciones que me dieron, y aún me siguen dando.

A mi amigo y colega, Ing. Luis Efraín Clemente  
Sanguinetti, que descansa en paz, por su constante,  
incommensurable e incondicional apoyo.

A Giovanna, mi esposa, mi mujer, mi amiga, mi  
compañera, mi cómplice, mi todo; por estar siempre  
a mi lado, amándome, apoyándome y animándome.

### **Agradecimiento**

A mi profesor en la maestría y asesor de tesis,  
Dr. Jorge Luis Breña Oré, por su esmerada  
orientación y minuciosa revisión de la tesis.

Al Dr. Mariano Sergio Gutiérrez Torero, por  
su predisposición y apoyo incondicional para  
revisar la tesis.

Los consejos y sugerencias de ambos  
profesionales mejoraron significativamente la  
redacción de la tesis.

## Índice

Dedicatoria.....	ii
Agradecimiento.....	iii
Índice.....	iv
Lista de Tablas .....	xi
Lista de Figuras.....	xiv
Resumen.....	xxii
Abstract.....	xxiii
I Introducción.....	1
1.1 Planteamiento del Problema .....	1
1.2 Descripción del Problema.....	2
1.3 Formulación del Problema.....	7
1.3.1 Problema General.....	7
1.3.2 Problemas Específicos .....	7
1.4 Antecedentes .....	8
1.4.1 Antecedentes Internacionales.....	8
1.4.2 Antecedentes Nacionales.....	11
1.5 Justificación de la Investigación .....	13
1.6 Limitaciones de la Investigación .....	15
1.7 Objetivos de la Investigación.....	16
1.7.1 Objetivo General .....	16
1.7.2 Objetivos Específicos.....	16
1.8 Hipótesis .....	17
1.8.1 Hipótesis General .....	17
1.8.2 Hipótesis Específicas .....	17

II Marco Teórico .....	18
2.1 Marco Conceptual .....	18
2.1.1 Los Pronósticos .....	18
2.1.2 Demanda .....	18
2.1.3 Pronóstico de Demanda.....	19
2.1.4 Importancia de la Precisión de un Pronóstico de Demanda.....	19
2.1.5 Medidas de Precisión del Pronóstico .....	22
2.1.5.1 Desviación Media Absoluta (DMA) .....	23
2.1.5.2 Error Cuadrático Medio (ECM) .....	23
2.1.5.3 Error Medio Absoluto Porcentual (EMAP).....	23
2.1.5.4 Sesgo.....	24
2.1.6 Horizontes Temporales de los Pronósticos de Demanda .....	26
2.1.7 Clasificación de los Pronósticos.....	27
2.1.8 Métodos de Pronóstico .....	28
2.1.8.1 Métodos Cualitativos.....	28
2.1.8.2 Métodos Cuantitativos.....	29
2.1.8.2.1 Métodos de Series Temporales .....	30
2.1.8.2.2 Métodos Causales o Asociativos .....	37
2.1.9 Elección del Método de Pronóstico.....	39
2.1.10 El Análisis de Regresión Lineal .....	45
2.1.10.1 El Análisis de Regresión Lineal Simple .....	46
2.1.10.2 El Análisis de Regresión Lineal Múltiple .....	47
2.1.10.3 El Principio de Parsimonia .....	48
2.1.10.4 Mínimos Cuadrados Ordinarios (MCO).....	48
2.1.10.5 Tipos de Datos.....	49

2.1.10.6 Supuestos del Análisis de Regresión Lineal con Series Temporales .....	50
2.1.10.6.1 El Modelo Es Lineal .....	50
2.1.10.6.2 No Existe Colinealidad o Multicolinealidad Exacta.....	51
2.1.10.6.3 El Promedio de los Errores Es Cero .....	53
2.1.10.6.4 Los Errores No Están Autocorrelacionados .....	54
2.1.10.6.5 Los Errores Son Homocedásticos .....	56
2.1.10.6.6 Los Errores Se Distribuyen Normalmente.....	57
2.1.10.6.7 Ausencia de Cambio Estructural .....	59
2.1.10.6.8 Especificación Correcta .....	60
2.1.10.7 Propiedades de los Parámetros Estimados Mediante MCO .....	62
2.1.10.8 Errores Tipos I y II, Nivel de Significancia y Potencia.....	63
2.1.10.9 El Valor p .....	63
2.1.10.10 Detección de la Violación de los Supuestos.....	64
2.1.10.10.1 Linealidad: El Multiplicador de Lagrange.....	64
2.1.10.10.2 Colinealidad: El Factor Inflacionario de Varianza (FIV).....	65
2.1.10.10.3 El Promedio de los Errores Es Cero .....	66
2.1.10.10.4 Autocorrelación: La Prueba d de Durbin-Watson (DW).....	67
2.1.10.10.5 Heterocedasticidad: La Prueba de White.....	69
2.1.10.10.6 Normalidad: La Prueba de Doornik-Hansen (DH).....	70
2.1.10.10.7 Cambio Estructural: Las Gráficas CUSUM y CUSUMQ .....	71
2.1.10.10.8 Especificación: La Prueba RESET de Ramsey (RR) .....	73
2.1.10.11 Violación de los Supuestos: Medidas Correctivas .....	73
2.1.10.11.1 El Modelo No Es Lineal .....	73
2.1.10.11.2 Colinealidad.....	76
2.1.10.11.3 El Promedio de los Errores No Es Cero .....	77

2.1.10.11.4 Errores Autocorrelacionados .....	77
2.1.10.11.5 Errores Heterocedásticos .....	78
2.1.10.11.6 Errores No Normales .....	79
2.1.10.11.7 Cambio Estructural .....	81
2.1.10.11.8 Especificación Incorrecta.....	83
2.1.10.12 Significancia de los Coeficientes.....	84
2.1.10.12.1 Significancia Global: La Prueba F.....	84
2.1.10.12.2 Significancia Individual: La Prueba t .....	85
2.1.10.13 El Coeficiente de Determinación ( $r^2$ o $R^2$ ) .....	86
2.1.10.14 El Coeficiente de Determinación Corregido ( $r_c^2$ o $R_c^2$ ).....	87
2.1.10.15 Pronóstico Fuera de la Muestra .....	88
2.1.11 La Industria de Productos Hidrobiológicos Congelados en el Perú.....	88
2.1.11.1 Desembarque para la Producción de Congelados (1980-2017) .....	89
2.1.11.2 Producción de Congelados (1980-2017) .....	89
2.1.11.3 Producción de Congelados Según Especie (1988-2017).....	90
2.1.11.4 Demanda de Congelados (1980-2018).....	93
2.1.11.5 Demanda Per Cápita de Productos Hidrobiológicos (1980 y 2018).....	95
2.1.11.6 Demanda Per Cápita de Congelados (1980-2018) .....	96
2.1.11.7 Exportaciones de Congelados en Volumen (1980-2018).....	97
2.1.11.8 Exportaciones de Congelados en Valor FOB (1991-2018).....	98
2.1.11.9 Exportaciones de Congelados en Valor FOB Promedio (1991-2018) .....	98
2.1.11.10 Capacidad Instalada (1984-2017).....	99
2.1.11.11 Capacidad Instalada Según Región (2017).....	101
2.1.12 El Software Gretl.....	102
III Método .....	103

3.1 Tipo de Investigación.....	103
3.2 Población y Muestra .....	103
3.2.1 Población.....	103
3.2.2. Muestra.....	104
3.3 Operacionalización de Variables .....	105
3.4 Instrumentos.....	105
3.5 Procedimientos.....	106
3.5.1 Construcción, Verificación, Corrección y Significancia de los Modelos Iniciales..	106
3.5.2 Extensión de las Series Temporales y Precisión del Pronóstico .....	107
3.5.3 Número de Variables Independientes y Precisión del Pronóstico .....	107
3.5.4 Modelo de Regresión Lineal Validado y Precisión del Pronóstico.....	108
3.6 Análisis de Datos .....	108
IV Resultados.....	109
4.1 Contratación de Hipótesis .....	109
4.1.1 Hipótesis Específica 1 ( $H_1$ ) .....	109
4.1.2 Hipótesis Específica 2 ( $H_2$ ) .....	110
4.1.3 Hipótesis Específica 3 ( $H_3$ ) .....	111
4.1.4 Hipótesis General .....	113
4.2 Análisis e interpretación .....	114
4.2.1 Construcción, Verificación, Corrección y Significancia de los Modelos Iniciales..	114
4.2.1.1 Construcción de los Modelos 1980-2018 .....	114
4.2.1.2 Verificación del Modelo MCO: CONG vs. PBI (1980-2018) .....	115
4.2.1.3 Corrección del Modelo MCO: CONG vs. PBI (1980-2018).....	120
4.2.1.4 Verificación del Modelo MCO: CONG vs. PBI e INF1 (1980-2018) .....	120
4.2.1.5 Corrección del Modelo MCO: CONG vs. PBI e INF1 (1980-2018) .....	125

4.2.1.6 Verificación del Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018) .....	126
4.2.1.7 Corrección del Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018).....	131
4.2.1.8 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. PBI e INF1 (1980-2018).....	132
4.2.1.9 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. DI (1980-2018).....	134
4.2.1.10 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. DI e INF1 (1980-2018) .....	135
4.2.1.11 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. PBIman e INF1 (1980-2018) .....	136
4.2.1.12 Construcción, Verificación y Corrección de los Modelos 1980-2013 .....	137
4.2.1.13 Significancia de los Modelos PW 1980-2013 .....	139
4.2.2 Extensión de las Series Temporales y Precisión del Pronóstico .....	142
4.2.2.1 Modelos PW: L-CONG vs. DI .....	142
4.2.2.2 Modelos PW: L-CONG vs. DI e INF1 .....	145
4.2.2.3 Modelos PW: L-CONG vs. PBIman e INF1 .....	148
4.2.2.4 Modelos PW: L-CONG vs. PBI .....	151
4.2.2.5 Modelos PW: L-CONG vs. PBI e INF1 .....	155
4.2.3 Número de Variables Independientes y Precisión del Pronóstico .....	159
4.2.3.1 Modelo PW: L-CONG vs. DI Aumentado a PW: L-CONG vs. DI e INF1 .....	160
4.2.3.2 Modelo PW: L-CONG vs. PBI Aumentado a PW: L-CONG vs. PBI e INF1 .....	160
4.2.4 Modelo de Regresión Lineal Validado y Precisión del Pronóstico.....	161
4.2.4.1 Modelo MCO: CONG vs. TIME (1980-2013).....	163
4.2.4.2 Modelos PW: L-CONG vs. DI (1980-2013) y MCO: CONG vs. DI (1980-2013) .....	164

4.2.4.3 Modelos PW: L-CONG vs. DI e INF1 (1980-2013) y MCO: CONG vs. DI e INF1 (1980-2013).....	165
4.2.4.4 Modelos PW: L-CONG vs. PBI (1980-2013) y MCO: CONG vs. PBI (1980-2013).....	166
4.2.4.5 Modelos PW: L-CONG vs. PBI e INF1 (1980-2013) y MCO: CONG vs. PBI e INF1 (1980-2013).....	168
4.2.4.6 Modelos PW: L-CONG vs. PBIman e INF1 (1980-2013) y MCO: CONG vs. PBIman e INF1 (1980-2013).....	169
V Discusión de Resultados .....	171
5.1 Extensión de las Series Temporales y Precisión del Pronóstico .....	171
5.2 Número de Variables Independientes y Precisión del Pronóstico .....	172
5.3 Modelo de Regresión Lineal Validado y Precisión del Pronóstico.....	173
VI Conclusiones.....	175
VII Recomendaciones.....	176
VIII Referencias.....	178
IX Anexos .....	193
Anexo A Matriz de Consistencia .....	193
Anexo B Series Temporales Empleadas (1980-2018).....	194
Anexo C Datos Para Probar la Hipótesis Específica 1 .....	195
Anexo D Datos Para Probar la Hipótesis Específica 2.....	197

## Lista de Tablas

Tabla 1 Precisión del Pronóstico Según el EMAP.....	24
Tabla 2 Duración de los Horizontes Temporales de los Pronósticos Según Autores.....	27
Tabla 3 Elección del Método de Pronóstico Según el Ciclo de Vida de los Productos.....	41
Tabla 4 Elección del Método de Pronóstico Según el Patrón de la Serie Temporal .....	42
Tabla 5 Elección del Método de Pronóstico Según el Portafolio de Productos.....	43
Tabla 6 Elección del Método de Pronóstico Según Área de Aplicación y Precisión Requerida .....	44
Tabla 7 Ejemplos de Modelos Lineales y No Lineales.....	51
Tabla 8 Valores del FIV y su Interpretación Según Autor .....	66
Tabla 9 Modelos No Lineales Transformados en Lineales por Medio de Logaritmos .....	75
Tabla 10 Transformaciones Box-Cox Según el Valor de $\lambda$ .....	80
Tabla 11 Construcción de Variables Dummy (D) .....	82
Tabla 12 Operacionalización de Variables .....	105
Tabla 13 EMAP (2014-2018) de Modelos No Validados, de Modelos Validados y Diferencial.....	111
Tabla 14 Modelos de Regresión Lineal Simple Ordenados Según $r^2$ (1980-2018).....	115
Tabla 15 Codificación de los Supuestos MCO.....	133
Tabla 16 Verificación de los Modelos MCO con Variable Dependiente CONG (1980-2018) .....	133
Tabla 17 Verificación de los Modelos MCO con Variable Dependiente L-CONG (1980- 2018).....	134
Tabla 18 Modelos Corregidos por PW con Variable Dependiente L-CONG (1980-2018) .	134
Tabla 19 Verificación de los Modelos MCO con Variable Dependiente CONG (1980-2013) .....	138

Tabla 20 Verificación de los Modelos MCO con Variable Dependiente L-CONG (1980-2013).....	138
Tabla 21 Modelos PW con Variable Dependiente L-CONG (1980-2013).....	139
Tabla 22 Modelo PW: L-CONG vs. DI (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018) .....	143
Tabla 23 Modelos PW: L-CONG vs. DI. EMAP, Precisión, Sesgo (2014-2018) y Merma Según Extensión de las Series Temporales .....	144
Tabla 24 Modelo PW: L-CONG vs. DI e INF1 (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018) .....	146
Tabla 25 Modelos PW: L-CONG vs. DI e INF1. EMAP, Precisión, Sesgo (2014-2018) y Merma Según Extensión de las Series Temporales.....	147
Tabla 26 Modelo PW: L-CONG vs. PBIman e INF1 (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018) .....	149
Tabla 27 Modelos PW: L-CONG vs. PBIman e INF1. EMAP, Precisión, Sesgo (2014-2018) y Merma Según Extensión de las Series Temporales.....	150
Tabla 28 Modelo PW: L-CONG vs. PBI (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018) .....	152
Tabla 29 Modelos PW: L-CONG vs. PBI. EMAP, Precisión, Sesgo (2014-2018) y Merma Según Extensión de las Series Temporales .....	153
Tabla 30 Modelo PW: L-CONG vs. PBI e INF1 (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018) .....	155
Tabla 31 Modelos PW: L-CONG vs. PBI e INF1. EMAP, Precisión, Sesgo (2014-2018) y Merma Según Extensión de las Series Temporales.....	156
Tabla 32 EMAP, Sesgo y Extensión por Variable Independiente de Modelos PW Seleccionados .....	158

Tabla 33	EMAP, Precisión, Sesgo (2014-2018) y Poder Explicativo de Modelos no Validados y de Modelos Validados (1980-2013).....	162
Tabla 34	EMAP y Sesgo (2014-2018) con el Modelo No Validado MCO: CONG vs. TIME (1980-2013).....	163
Tabla 35	EMAP y Sesgo (2014-2018) con el Modelo Validado PW: L-CONG vs. DI (1980-2013) y con el Modelo No Validado MCO: CONG vs. DI (1980-2013).....	164
Tabla 36	EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. DI e INF1 (1980-2013) y del Modelo No Validado MCO: CONG vs. DI e INF1 (1980-2013) .....	166
Tabla 37	EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. PBI (1980-2013) y del Modelo No Validado MCO: CONG vs. PBI (1980-2013) .....	167
Tabla 38	EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. PBI e INF1 (1980-2013) y del Modelo No Validado MCO: CONG vs. PBI e INF1 (1980-2013) .....	168
Tabla 39	EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. PBIman e INF1 (1980-2013) y del Modelo No Validado MCO: CONG vs. PBIman e INF1 (1980-2013).....	169
Tabla 40	Clasificación del EMAP Propuesta por el Autor .....	176

## Lista de Figuras

Figura 1 Pronóstico con Sesgo (a) Mayor a Cero y (b) Menor a Cero .....	25
Figura 2 Pronóstico Insesgado (a) con Precisión Muy Buena y (b) con Precisión Pobre .....	26
Figura 3 Clasificación de los Métodos de Pronóstico.....	28
Figura 4 Patrones de Demanda: (a) Nivel, (b) Tendencia, (c) Estacional y (d) Ciclo .....	31
Figura 5 Comportamiento Estacional (a) Aditivo y (b) Multiplicativo .....	34
Figura 6 Patrones de una Serie Temporal .....	36
Figura 7 Árbol de Decisión para la Elección del Método de Pronóstico.....	40
Figura 8 Ubicación de Algunos Productos Según el Ciclo de Vida .....	41
Figura 9 Elección del Método de Pronóstico Según Costo Total y Precisión .....	45
Figura 10 La Recta de Regresión.....	46
Figura 11 El Plano de Regresión .....	48
Figura 12 Errores en (a) una Recta de Regresión y (b) un Plano de Regresión .....	49
Figura 13 Variables (a) No Colineales y (b) Colineales .....	52
Figura 14 Errores (a) Con Autocorrelación Positiva, (b) Con Autocorrelación Negativa y (c) Sin Autocorrelación (Errores esféricos) .....	55
Figura 15 Errores (a) Con Autocorrelación Positiva, (b) Con Autocorrelación Negativa y (c) Sin Autocorrelación.....	56
Figura 16 Errores (a) Homocedásticos, (b) Heterocedásticos Crecientes y (c) Heterocedásticos Decrecientes .....	57
Figura 17 La Distribución Normal.....	58
Figura 18 Regresiones (a) Paralelas, (b) Concurrentes y (c) Inconexas .....	60
Figura 19 Patrón de Errores por una Variable Omitida .....	61
Figura 20 Zonas de No Rechazo y Rechazo de $H_0$ para Autocorrelación .....	69

Figura 21 Distribución (a) Asimétrica Derecha o Positiva, (b) Asimétrica Izquierda o Negativa y (c) Simétrica .....	70
Figura 22 Distribución (a) Platicúrtica, (b) Mesocúrtica y (c) Leptocúrtica .....	71
Figura 23 Gráfica CUSUM: (a) Estabilidad Estructural y (b) Cambio Estructural .....	72
Figura 24 Gráfica CUSUMQ: (a) Estabilidad Estructural y (b) Cambio Estructural .....	72
Figura 25 Modelos No Lineales (a y c) Linealizados por Logaritmos (b y d).....	74
Figura 26 Modelo Polinomial (a) Cuadrático y (b) Cúbico .....	76
Figura 27 Perú: Desembarque para la Producción de Congelados (1980-2017).....	89
Figura 28 Perú: Producción de Productos Hidrobiológicos Congelados (1980-2017).....	90
Figura 29 Perú: Producción de Productos Hidrobiológicos Congelados Según Especie (1988-2017).....	91
Figura 30 Perú: Producción de Congelados Según Especies Principales (1988-2017) .....	91
Figura 31 Perú: Participación de Congelados de Pota y Merluza en la Producción Total (1988-2017) .....	92
Figura 32 Pota Congelada Presentada en (a) Filetes Precocidos, (b) Anillas, (c) Botones y (d) Rabas .....	92
Figura 33 Perú: Demanda de Productos Hidrobiológicos Congelados (1980-2018).....	93
Figura 34 Perú: Demanda de Productos Hidrobiológicos y de Congelados (1980-2018).....	94
Figura 35 Perú: Contribución de la Demanda Congelados en la Demanda Total (1980-2018) .....	94
Figura 36 Perú: Demanda Per Cápita de Productos Hidrobiológicos Según Rubros (1980 y 2018).....	95
Figura 37 Perú: Contribución en la Demanda Per Cápita de Productos Hidrobiológicos Según Rubros (a) en 1980 y (b) en 2018 .....	96
Figura 38 Perú: Demanda Per Cápita de Congelados (1980-2018).....	97

Figura 39	Perú: Exportación de Congelados en Volumen (1980-2018).....	97
Figura 40	Perú: Valor FOB de las Exportaciones de Congelados (1991-2018).....	98
Figura 41	Perú: Valor FOB Promedio de las Exportaciones de Congelados (1991-2018) ..	99
Figura 42	Perú: Número de Plantas de Productos Hidrobiológicos Congelados (1984-2017) .....	100
Figura 43	Perú: Capacidad Instalada de la Industria de Productos Hidrobiológicos Congelados (1984-2017) .....	100
Figura 44	Perú: Capacidad Instalada Promedio de la Industria de Productos Hidrobiológicos Congelados (1984-2017) .....	100
Figura 45	Perú: Capacidad Instalada de la Industria de Productos Hidrobiológicos Congelados Según Región (2017) .....	101
Figura 46	Perú: Participación de la Capacidad Instalada de la Industria de Productos Hidrobiológicos Congelados Según Región (2017) .....	101
Figura 47	Perú: Número de Plantas y Capacidad Instalada Promedio de la Industria de Productos Hidrobiológicos Congelados Según Región (2017) .....	102
Figura 48	Diagrama de Dispersión Entre la Extensión de las Series Temporales y la Precisión del Pronóstico de Demanda (EMAP) .....	109
Figura 49	Prueba t de Correlación Entre la Extensión de las Series Temporales y la Precisión del Pronóstico de Demanda .....	110
Figura 50	Diagrama de Dispersión Entre el Número de Variables Independientes y la Precisión del Pronóstico de Demanda (EMAP) .....	110
Figura 51	Prueba t de Correlación Entre el Número de Variables Independientes y la Precisión del Pronóstico de Demanda .....	111
Figura 52	Pruebas de Normalidad para el EMAP Diferencial.....	112
Figura 53	Estadísticos Principales del EMAP Diferencial .....	113

Figura 54 Prueba t de Diferencia de EMAP Medio Entre Modelos Validados y Modelos No Validados .....	113
Figura 55 Modelo MCO: CONG vs. PBI (1980-2018) .....	115
Figura 56 Modelo MCO: CONG vs. PBI (1980-2018). Prueba LM de Linealidad (Cuadrados) .....	116
Figura 57 Modelo MCO: CONG vs. PBI (1980-2018). Prueba LM de Linealidad (Logaritmos) .....	116
Figura 58 Valores de $d_L$ y $d_U$ para la Prueba d de DW de Autocorrelación ( $n=39$ , $k=1$ ) .....	117
Figura 59 Modelo MCO: CONG vs. PBI (1980-2018). Prueba d de DW de Autocorrelación .....	117
Figura 60 Modelo MCO: CONG vs. PBI (1980-2018). Prueba de Heterocedasticidad de White .....	118
Figura 61 Modelo MCO: CONG vs. PBI (1980-2018). Prueba de Normalidad de DH.....	118
Figura 62 Modelo MCO: CONG vs. PBI (1980-2018). Pruebas Gráficas de Cambio Estructural (a) CUSUM y (b) CUSUMQ .....	119
Figura 63 Modelo MCO: CONG vs. PBI (1980-2018). Prueba de Especificación RR.....	119
Figura 64 Modelo MCO: CONG vs. PBI e INF1 (1980-2018) .....	120
Figura 65 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Cuadrados) .....	121
Figura 66 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Logaritmos) .....	121
Figura 67 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Colinealidad: FIV .....	122
Figura 68 Valores de $d_L$ y $d_U$ para la Prueba d de DW de Autocorrelación ( $n=39$ , $k=2$ ) .....	122
Figura 69 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba d de DW de Autocorrelación .....	123

Figura 70 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba de Heterocedasticidad de White.....	123
Figura 71 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba de Normalidad de DH .....	124
Figura 72 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Pruebas Gráficas de Cambio Estructural (a) CUSUM y (b) CUSUMQ .....	124
Figura 73 Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba de Especificación RR .....	125
Figura 74 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018) .....	126
Figura 75 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Cuadrados) .....	126
Figura 76 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Logaritmos) .....	127
Figura 77 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Colinealidad: FIV .....	127
Figura 78 Valores de $d_L$ y $d_U$ para la Prueba d de DW de Autocorrelación ( $n=39$ y $k=2$ ) ...	128
Figura 79 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba d de DW de Autocorrelación .....	128
Figura 80 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba de Heterocedasticidad de White .....	129
Figura 81 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba de Normalidad de DH .....	129
Figura 82 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Pruebas Gráficas de Cambio Estructural (a) CUSUM y (b) CUSUMQ .....	130
Figura 83 Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba de Especificación RR.....	130

Figura 84 Modelo PW: L-CONG vs. PBI e INF1 (1980-2018) .....	131
Figura 85 Modelo PW: L-CONG vs. PBI e INF1 (1980-2018). Prueba d de DW de Autocorrelación .....	131
Figura 86 Modelo PW: L-CONG vs. DI (1980-2018).....	135
Figura 87 Modelo PW: L-CONG vs. DI e INF1 (1980-2018) .....	136
Figura 88 Modelo PW: L-CONG vs. PBIman e INF1 (1980-2018).....	137
Figura 89 Modelo PW: L-CONG vs. DI (1980-2013).....	139
Figura 90 Modelo PW: L-CONG vs. DI e INF1 (1980-2013) .....	140
Figura 91 Modelo PW: L-CONG vs. PBIman e INF1 (1980-2013).....	141
Figura 92 Modelo PW: L-CONG vs. PBI (1980-2013).....	141
Figura 93 Modelo PW: L-CONG vs. PBI e INF1 (1980-2013) .....	142
Figura 94 Modelo PW: L-CONG vs. DI (1980-2013). CONG Real y Pronosticado (2014- 2018).....	143
Figura 95 Modelos PW: L-CONG vs. DI. EMAP (2014-2018) Según Extensión de las Series Temporales .....	145
Figura 96 CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. DI con Extensión de (a) 15 Años (1999-2013) y (b) 28 Años (1986-2013) .....	145
Figura 97 Modelo PW: L-CONG vs. DI e INF1 (1980-2013). CONG Real y Pronosticado (2014-2018) .....	146
Figura 98 Modelos PW: L-CONG vs. DI e INF1. EMAP (2014-2018) Según Extensión de las Series Temporales .....	148
Figura 99 CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. DI e INF1 con Extensión de (a) 32 Años (1982-2013) y (b) 28 Años (1986-2013) ....	148
Figura 100 Modelo PW: L-CONG vs. PBIman e INF1 (1980-2013). CONG Real y Pronosticado (2014-2018) .....	149

Figura 101 Modelos PW: L-CONG vs. PBI <sub>man</sub> e INF1. EMAP (2014-2018) Según Extensión de las Series Temporales .....	151
Figura 102 CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. PBI <sub>man</sub> e INF1 con Extensión de (a) 16 Años (1998-2013) y (b) 26 Años (1988-2013) .....	151
Figura 103 Modelo PW: L-CONG vs. PBI (1980-2013). CONG Real y Pronosticado (2014-2018) .....	152
Figura 104 Modelos PW: L-CONG vs. PBI. EMAP (2014-2018) Según Extensión de las Series Temporales .....	154
Figura 105 CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. PBI con Extensión de (a) 13 Años (2001-2013) y (b) 22 Años (1992-2013) .....	154
Figura 106 Modelo PW: L-CONG vs. PBI e INF1 (1980-2013). CONG Real y Pronosticado (2014-2018).....	155
Figura 107 Modelos PW: L-CONG vs. PBI e INF1. EMAP (2014-2018) Según Extensión de las Series Temporales.....	157
Figura 108 CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. PBI e INF1 con Extensión de (a) 28 Años (1986-2013) y (b) 31 Años (1983-2013) ..	157
Figura 109 CONG Real y Pronosticado (2014-2018) con los Modelos PW (a) L-CONG vs. DI (1999-2013), (b) L-CONG vs. DI e INF1 (1982-2013), (c) L-CONG vs. PBI (2001-2013) y (d) L-CONG vs. PBI e INF1 (1986-2013) .....	159
Figura 110 Modelos PW: L-CONG vs. DI y PW: L-CONG vs. DI e INF1. EMAP (2014-2018) Según Número de Variables Independientes .....	160
Figura 111 Modelos PW: L-CONG vs. PBI y PW: L-CONG vs. PBI e INF1. EMAP (2014-2018) Según Número de Variables Independientes .....	161

Figura 112 EMAP (2014-2018) de Modelos No Validados y de Modelos Validados (1980-2013).....	163
Figura 113 CONG Real y Pronosticado (2014-2018) con el Modelo No validado MCO: CONG vs. TIME (1980-2013).....	164
Figura 114 CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. DI (1980-2013) y con el Modelo No Validado MCO: CONG vs. DI (1980-2013).....	165
Figura 115 CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. DI e INF1 (1980-2013) y con el Modelo No Validado MCO: CONG vs. DI e INF1 (1980-2013).....	166
Figura 116 CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. PBI (1980-2013) y con el Modelo No Validado MCO: CONG vs. PBI (1980-2013).....	167
Figura 117 CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. PBI e INF1 (1980-2013) y con el Modelo No Validado MCO: CONG vs. PBI e INF1 (1980-2013).....	169
Figura 118 CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. PBIman e INF1 (1980-2013) y con el Modelo No Validado MCO: CONG vs. PBIman e INF1 (1980-2013).....	170
Figura 119 CONG Real y Pronosticado (2014-2018) (a) con el Modelo Simple PW: L-CONG vs. DI (1980-2013) y (b) con el Modelo Aumentado PW: L-CONG vs. DI e INF1 (1980-2013).....	172
Figura 120 CONG Real y Pronosticado (2014-2018) (a) con el Modelo Simple PW: L-CONG vs. PBI (1986-2013) y (b) con el Modelo Aumentado PW: L-CONG vs. PBI e INF1 (1986-2013).....	173

## Resumen

Esta tesis pretende evaluar la relación entre el análisis de regresión lineal y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú. La evaluación determinará la existencia y dirección de la relación entre la precisión del pronóstico, tanto con la extensión de las series temporales como con el número de variables independientes y establecerá si la precisión del pronóstico de un modelo validado es mayor frente a un modelo no validado. Se recopiló series temporales de demanda de productos hidrobiológicos congelados (variable dependiente) y de indicadores macroeconómicos (variables independientes). Las series temporales (1980-2018) se partieron en dos. Con la primera parte (1980-2013) se construyeron, verificaron y corrigieron los modelos de regresión lineal, empleando el software Gretl. Con la segunda parte (2014-2018) se realizaron los pronósticos fuera de la muestra y se calcularon sus precisiones por medio del Error Medio Absoluto Porcentual (EMAP). Los resultados fueron mixtos. Primero, no se encontró relación entre la precisión del pronóstico y la extensión de las series temporales. Aunque, se determinó que utilizando extensiones de entre 13 y 16 años, se obtienen los modelos más precisos. Segundo, la relación es positiva entre la precisión del pronóstico y el número de variables independientes. Tercero, la precisión del pronóstico de un modelo validado es mayor frente a un modelo no validado. Y, finalmente, existe una relación positiva entre el análisis de regresión lineal y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

**Palabras clave:** error medio absoluto porcentual (EMAP), Gretl, precisión del pronóstico de demanda, regresión lineal, series temporales.

### Abstract

This thesis aims to evaluate the relationship between linear regression analysis and demand forecasting accuracy of frozen hydrobiological products in Peru. The evaluation will determine existence and direction of relationship between demand forecasting accuracy with both extension of time series and number of independent variables; and will establish whether forecasting accuracy from a validated model is major versus a no validated model. Time series of frozen hydrobiological products demand (dependent variable) and macroeconomic indicators (independent variables) were collected. Time series (1980-2013) were divided into two. With first part (1980-2013) linear regression models were constructed, verified and corrected using Gretl software. With second part (2014-2018), out-of-sample forecasts were made and its accuracies were calculated by means of Mean Absolute Percent Error (MAPE). The results were mixed. First, it was not found relationship between forecasting accuracy and time series extension. Although it was found that relationship is direct for certain ranges of years, whereas for others is inverse. Further, it was determined that using extensions between 13 and 16 years, the most accurate models are obtained. Second, there is a positive relationship between forecasting accuracy and number of independent variables. Third, the forecasting accuracy from a validated model is major versus a no validated model. And, finally, there is a positive relationship between linear regression analysis and demand forecasting accuracy of frozen hydrobiological products in Peru.

**Keywords:** demand forecasting accuracy, Gretl, linear regression, mean absolute percent error (MAPE), time series.

## **I Introducción**

El pronóstico de demanda, del bien a producir o del servicio a ofrecer, es uno de los pilares del proceso de formulación y evaluación de proyectos de inversión. A partir del pronóstico, se obtiene información cuantitativa que, posteriormente, permitirá la construcción de los flujos de caja del proyecto y, por ende, el cálculo de los indicadores de rentabilidad, sobre los cuales se decidirá la conveniencia de la materialización del proyecto de inversión. Los principales indicadores de rentabilidad son el Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR) (Baca, 2013; Sapag et al., 2014).

El análisis de regresión lineal, con series temporales, es una de las técnicas más empleadas en la construcción de modelos que permiten, para cierto horizonte temporal, pronosticar la demanda. Sin embargo, para que el modelo resultante sea de utilidad se requiere el cumplimiento de algunas condiciones. Primero, debe verificarse que el modelo no viole ninguno de los supuestos del método Mínimos Cuadrados Ordinarios (MCO). Luego, debe establecerse la significancia del modelo, tanto globalmente como individualmente. Seguidamente, el modelo debe exhibir suficiente poder explicativo. Y, finalmente, el modelo debe ser lo más preciso posible. Solo después, puede utilizarse el modelo para pronosticar.

Aunque, cumplir escrupulosamente con todas las condiciones descritas puede parecer un proceso trabajoso y engorroso, es necesaria su ejecución, solo así se obtendrán indicadores de rentabilidad confiables y la decisión final que se tome, acerca de la cristalización del proyecto, será la más acertada. Cabe resaltar que la decisión final puede involucrar la inversión de algunos miles de soles o de hasta varios miles de millones de dólares.

### **1.1 Planteamiento del Problema**

Por medio de la aplicación del análisis de regresión lineal con series temporales es posible obtener pronósticos de demanda de gran precisión, siempre y cuando no se viole ninguno de los supuestos bajo los cuales se encontró la ecuación de regresión. Por lo cual, es

imprescindible verificar el cumplimiento de los supuestos, antes de utilizar la ecuación de regresión para pronosticar. No obstante, ante la violación de algún supuesto, es posible aplicar medidas correctivas. Sin embargo, en muchos casos en los que se ha utilizado el análisis de regresión, no se ha verificado el cumplimiento de los supuestos. La no verificación puede deberse al desconocimiento de los supuestos, a ignorar las pruebas para establecer si los supuestos han sido violados, a considerar que la violación de los supuestos no representa un problema y a la falta de conocimiento de las medidas remediales ante la violación de los supuestos (Hoekstra et al., 2012).

## **1.2 Descripción del Problema**

En el Perú la extracción pesquera tiene dos destinos, el Consumo Humano Directo (CHD) y el Consumo Humano Indirecto (CHI). Por un lado, el CHI se refiere a la pesca para la producción de harina y aceite de pescado. Por otro lado, el CHD se subdivide a su vez en cuatro rubros, pesca para consumo fresco y pesca para la producción de enlatados, congelados y curados. Desde el año 2012 en el Perú se viene experimentando un crecimiento sostenido de la demanda de productos hidrobiológicos congelados. Según el Ministerio de la Producción (PRODUCE, 2019) en 2012 se consumieron 73,5 miles t y en 2018 el consumo llegó a 140,2 miles de t, es decir un crecimiento de poco más de 91%. Este comportamiento ha significado que el rubro crezca en importancia, en 1980 representó el 6,7% del CHD, en 2012 significó el 13,2% y en 2018 su participación llegó al 21,8%. Por lo expuesto, es evidente que el rubro congelados se encuentra en franca expansión, crecimiento que ha originado la inversión de capitales, lo cual se refleja en el crecimiento sostenido de la capacidad instalada del sector. En 2012 existían 117 plantas que totalizaban una capacidad instalada de 6 804 t/día con un promedio de 58,15 t/día. Mientras que en 2017 las plantas sumaban 127 y podían procesar 9 696 t/día, con una media de 76,35 t/día. De lo expuesto se desprende algunas interrogantes, durante los próximos años, ¿continuará la tendencia

expansiva de la demanda de congelados? o por el contrario ¿se revertirá?, ¿en qué momento? y, en cualquier caso, ¿cuántas toneladas demandarán los consumidores durante los años siguientes?

El análisis de regresión es el método apropiado para realizar pronósticos de corto, mediano y largo plazos (Hanke y Wichern, 2014). El método posibilita la construcción de modelos que permiten pronosticar la demanda de un producto individual o de un grupo de productos. El pronóstico cuantificará el volumen de demanda para cierto horizonte temporal y mostrará su comportamiento general futuro (creciente, estable o decreciente). En particular, un comportamiento futuro creciente de la demanda genera incentivos para la ampliación de la oferta del bien en cuestión. Esta ampliación se puede dar de dos formas. Por un lado, las empresas en operación podrían incrementar su capacidad instalada. Por otro lado, se podría instalar una nueva planta productora. En cualquiera de los dos casos se debe formular y evaluar un proyecto de inversión para verificar la rentabilidad de la ampliación de la oferta.

Sin embargo, un pronóstico poco preciso de demanda origina que la determinación de la rentabilidad del proyecto sea poco fiable. La rentabilidad de un proyecto se establece a través de los indicadores de rentabilidad, principalmente el VAN y la TIR. Ambos indicadores se calculan a partir de los flujos de caja del proyecto, cuyos componentes básicos son inversión inicial, ingresos operacionales y egresos operacionales.

Los estudios que se realizan en la etapa de formulación de un proyecto, proporcionan la información necesaria para la construcción de los flujos de caja. Ciertamente, el estudio de mercado suministra, entre otros ítems, el pronóstico de demanda, elemento principal sobre el cual se establece el pronóstico de demanda del proyecto, el cual resulta en un insumo importante en la elaboración de los estudios técnico y económico.

En el estudio técnico el pronóstico de demanda del proyecto, es la base para determinar el tamaño de planta, el que a su vez se convierte en un insumo relevante en la

preparación del estudio económico. En el estudio económico, el tamaño de planta es fundamental en la cuantificación de la inversión inicial y mediante el pronóstico de demanda del proyecto, se establecen los ingresos y los egresos operacionales en los que se incurrirán durante el horizonte de evaluación del proyecto.

Con la información evacuada del estudio económico es posible la elaboración de los flujos de caja, y por ende, la determinación de los indicadores de rentabilidad del proyecto. En suma, un pronóstico de demanda lo suficientemente preciso, permitirá obtener indicadores de rentabilidad confiables, en base a los cuales se decidirá la materialización, o no, del proyecto de inversión.

Acerca de la precisión de los pronósticos, Ballou (2004) refirió que normalmente los pronósticos no son cien por ciento precisos, porque lo que ocurrirá no es un reflejo exacto de lo ocurrido. Por su parte Heizer et al. (2017) y Reid y Sanders (2013) manifestaron la existencia de una relación negativa entre precisión y horizonte del pronóstico, en el corto plazo los cambios en los datos no son importantes, sin embargo, es más probable que los patrones y las relaciones cambien en plazos mayores.

La calidad de la información también influye en la precisión de los pronósticos. En ese sentido, Sapag et al. (2014) afirmaron que la calidad de los datos que fueron base del pronóstico afecta su validez. Y Shahabuddin (2009) manifestó que se compromete la precisión del pronóstico cuando se utilizan datos que son inconsistentes y débiles en el sentido estadístico. Por lo cual es necesario que la información provenga de fuentes oficiales tales como el Ministerio de la Producción (PRODUCE), el Ministerio de Economía y Finanzas (MEF), el Instituto Nacional de Estadística e Informática (INEI), el Banco Central de Reserva del Perú (BCRP), la Comisión de Promoción del Perú para la Exportación y el Turismo (PROMPERÚ), entre otras.

Además, Reid y Sanders (2013) señalaron que los pronósticos son altamente precisos cuando se realiza para un conjunto de productos que cuando se hace para productos específicos. Asimismo, es posible que al prorratear el pronóstico global entre los productos específicos se reserve una precisión considerable, que sería mayor a la obtenida del pronósticos de productos específicos (Ballou, 2004). Por lo tanto, el pronóstico de demanda conjunta de productos hidrobiológicos congelados será más preciso que el pronóstico de la demanda de productos individuales tales como filete precocido de pota, de colas de langostino o de conchas de abanico.

Por otro lado, el análisis de regresión con series temporales es muy utilizado para realizar pronósticos (Hanke y Wichern, 2014; Montgomery et al., 2015; Wilson y Keating, 2009). En general, cuando se utiliza el análisis de regresión lineal con series temporales, la precisión de los pronósticos puede tener relación con la extensión de las series temporales, con el número de variables independientes y con la validez del modelo.

En lo que se respecta a la extensión de las series temporales, Heizer et al. (2017) concluyeron que no se puede hacer pronósticos para los 10 años siguientes a partir de un modelo que considera información de los cinco años previos. Así mismo, Groebner et al. (2018) afirmaron que los errores se minimizan cuando se utilizan tamaños de muestra más grandes. De igual manera, Stock y Watson (2020) aseguraron que un tamaño muestral reducido puede afectar los intervalos de confianza y las pruebas de hipótesis. Además, Wheelen et al. (2018) sostuvieron que será mayor la confianza en el pronóstico, cuando se logre encontrar datos relevantes de mucho más tiempo atrás. A su turno, Armstrong (2002) señaló que el incremento del número datos mejora la precisión de los pronósticos de largo plazo. También, Darlington y Hayes (2017) manifestaron que en muestras grandes se pueden aplicar una serie de pruebas potencialmente útiles, las mismas que no son aplicables en muestras pequeñas.

Concretamente, Chase (2013) aconsejó al menos cinco datos por cada variable independiente. A su turno, Behar y Grima (2018), Doane y Seward (2016), Hanke y Wichern (2014) y Wilson y Keating (2009) fueron más conservadores y propusieron 10 observaciones como mínimo por cada variable independiente. Sin embargo, Keith (2019) fue más exigente y sugirió de 10 a 20 datos por cada variable independiente.

En cuanto al número de variables independientes, los modelos de regresión simple pueden resultar insuficientes para explicar, de manera significativa, el comportamiento de la variable dependiente. Sobre el particular, Carlberg (2016) planteó la existencia de una relación positiva entre el número de variables independientes y la precisión del pronóstico. Asimismo, Levine et al. (2017) afirmaron que frecuentemente los pronósticos son mejores al utilizar más de una variable independiente. Por su parte Walpole et al. (2016) precisaron la necesidad de más de una variable independiente en el modelo de regresión. Adicionalmente, Anderson et al. (2017) refirieron que un modelo de regresión múltiple, al considerar más de una variable independiente, ofrece mejores estimaciones que un modelo de regresión simple. También, Lilja (2017) manifestó que un modelo con muy pocas variables independientes puede originar pronósticos sesgados. En el mismo sentido Stock y Watson (2020) señalaron que los modelos de regresión simple son susceptibles al sesgo de la variable no incluida, si la variable es relevante, entonces el valor de la pendiente será sesgado; el análisis de regresión múltiple minimiza el sesgo al poder introducir en el modelo la variable no incluida. Más en concreto, Baca (2013) sostuvo la poca utilidad del tiempo como única variable independiente, el tiempo es insuficiente para explicar el comportamiento de la demanda, por tanto es necesario incluir otras variables.

Finalmente, la validez del modelo de regresión lineal también puede influir en la precisión de los pronósticos de demanda. Un modelo de regresión lineal validado es aquel

que no viola ninguno de los supuestos inherentes al análisis de regresión y que es significativo, tanto a nivel global como individual.

Sobre el particular, la violación del supuesto de ausencia de cambio estructural es la causa primordial de errores en los pronósticos (Castle y Hendry, 2019; Ghysels y Marcellino, 2018; Marcellino, 2016). De igual manera, la violación del supuesto de linealidad produce pronósticos erróneos (Young, 2017). Igualmente, para pronosticar es muy importante que no se viole el supuesto de normalidad de los errores (Gujarati y Porter, 2010; Pedace, 2013; Welc y Rodriguez, 2018). Asimismo, los pronósticos pierden precisión, cuando se violan los supuestos de errores no autocorrelacionados y homocedásticos (Rosales et al., 2013; Welc y Rodriguez, 2018). También, la capacidad de pronóstico depende de la no violación del supuesto de especificación correcta (Welc y Rodriguez, 2018).

Además, (Lind et al. (2018) precisaron que la capacidad de pronóstico también depende de que los coeficientes del modelo sean estadísticamente diferentes de cero. Caso contrario, las variables independientes no son útiles para pronosticar la variable dependiente (Steele et al., 2016; Vu, 2015).

### **1.3 Formulación del Problema**

#### ***1.3.1 Problema General***

¿Qué relación existe entre el análisis de regresión lineal y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú?

#### ***1.3.2 Problemas Específicos***

**Problema Específico 1.** ¿Cómo se relaciona la extensión de las series temporales con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú?

**Problema Específico 2.** ¿En qué forma se relaciona el número de variables independientes con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú?

**Problema Específico 3.** ¿Es mayor la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú, de un modelo de regresión lineal validado frente a un modelo de regresión lineal no validado?

#### 1.4 Antecedentes

##### 1.4.1 Antecedentes Internacionales

Sánchez et al. (2011) en su estudio “Estimación de la demanda de importaciones de limón persa (*Citrus latifolia* Tanaka) en Estados Unidos procedentes de México (1994-2008)”, se plantearon el objetivo de establecer la asociación entre la demanda norteamericana de limón persa importado de México (DLP) y las variables precio unitario de importación (PLP), ingreso norteamericano (PIBU) y tipo de cambio real (TCR).

Se utilizó el análisis de regresión múltiple con series temporales anuales. Las series tuvieron una extensión de 15 años (1994-2008). El modelo superó las pruebas de significancia individual, de significancia global, de autocorrelación, de multicolinealidad y de heterocedasticidad. El poder explicativo del modelo se estableció en 90%, es decir las variables Ln PLP, Ln PIBU y Ln TCR explicaron el 90% del comportamiento de la variable Ln DLP. El modelo construido fue el siguiente:

$$\text{Ln DLP} = -21,402516 - 0,66151 \text{ Ln PLP} + 3,800415 \text{ Ln PIBU} + 0,837033 \text{ Ln TCR}$$

Según el modelo, si el precio unitario de importación (PLP) crece en 1%, la demanda norteamericana de limón persa importado de México (DLP) decrece en 0,66%. Si el ingreso norteamericano (PIBU) se expande en 1%, la DLP aumenta en 3,80%. Si tipo de cambio real (TCR) aumenta en 1%, la DLP crece en 0,84%.

El proyecto final de Nayal (2015) “Regression analysis to forecast the demand of new single family houses in USA”, tuvo como objetivo la construcción de un modelo para pronosticar la demanda de mercado de casas unifamiliares nuevas en los Estados Unidos, usando un método cuantitativo. El método aplicado fue el análisis de regresión múltiple con

series temporales anuales. Las series tuvieron una extensión de 21 años (1994-2014). En el estudio se examinaron las variables independientes PBI (Producto Bruto Interno), consumo personal, inflación, tasa de desempleo, tasa de interés hipotecario y población.

Antes de construir los modelos de regresión y para descartar variables, el investigador verificó el supuesto de linealidad mediante diagramas de dispersión, entre la variable dependiente y cada una de las variables independientes.

Se construyeron seis modelos. El modelo 1 se descartó por la presencia de multicolinealidad entre las variables consumo personal y población. Los modelos 2 y 3 tuvieron bajos coeficientes de determinación corregido de 60,3% y 49,4%, respectivamente. Los modelos 4 y 5, mejoraron el poder explicativo y tuvieron coeficientes de determinación corregido de 70,3% y 76,3%, respectivamente. El modelo 6 se construyó con las variables independientes consumo personal, tasa de desempleo y dos variables dummy (periodo 1 y periodo 2). El modelo presentó un alto poder explicativo, las variables independientes explicaron en un 82,9% el comportamiento de la demanda de casas. Además, el modelo superó las pruebas de multicolinealidad, de significancia individual, de significancia global, de homocedasticidad y de normalidad. El estudio concluyó en que la demanda de casas es afectada positivamente por el consumo personal y es afectada negativamente por la tasa de desempleo. A continuación, se presenta el modelo 6:

$$\text{Demanda de casas} = 8\,579\,583 + 1\,904 \text{ Consumo personal} - 895\,621 \text{ Tasa de desempleo} + \\ 9\,187\,261 \text{ Periodo 1} + 2\,890\,891 \text{ Periodo 2}$$

Finalmente, si el consumo personal se incrementa en una unidad, la demanda se incrementa en 1 904 casas unifamiliares nuevas. Si tasa de desempleo se expande en un punto porcentual, la demanda decrece en 895 621 casas unifamiliares nuevas. Si el valor de la variable dummy periodo 1 es cero, la demanda se contrae en 9 187 261 casas unifamiliares

nuevas. Si el valor de la variable dummy periodo 2 es cero, la demanda de casas unifamiliares nuevas decrece en 2 890 891 unidades.

Soto (2015) en su artículo “Modelo econométrico para estimar la demanda de sandías en USA, 2000 – 2010” utilizó el análisis de regresión múltiple con series temporales anuales para pronosticar la demanda de sandías en los Estados Unidos (DT). Las series temporales tuvieron una extensión de 21 años (1990-2010). El modelo se construyó utilizando las variables independientes población USA (PEUA), inflación USA (IEUA), precio promedio por cada 100 libras de melón cantaloupe en dólares de 2005 (PPD2005M), precio promedio por cada 100 libras de sandía en dólares de 2005 (PPD2005S), índice de precios al consumidor (IPC) y PIB per cápita en dólares de 2005 (PIBPC2005).

El modelo resultante tuvo un alto poder explicativo, las variables independientes PEUA, IEUA, PPD2005, PPD2005S, IPC y PIBPO2005 explicaron en un 89,44% el comportamiento de la demanda de sandías en los Estados Unidos. El modelo superó las pruebas de significancia individual, de significancia conjunta, de autocorrelación y de multicolinealidad. El modelo estimado fue:

$$DT = 10\,827,74 - 5,77 \times 10^{-5} PEUA - 20,89 IEUA + 80,69 PPD2005M - 70,51 PPD2005S + 119,76 IPC - 0,05 PIBPC2005$$

Según el modelo, si la PEUA aumenta en una unidad, la demanda de sandía decrece en 0,00005769 millones de libras. Si la IEUA crece en un punto porcentual, la demanda de sandía se contrae en 20,89 millones de libras. Si el PPD2005M se incrementa en un dólar, la demanda de sandía se expande en 80,69 millones de libras. Si el PPD2005S se expande en un dólar, la demanda de sandía decrece en 70,51 millones de libras. Si el IPC se expande en una unidad, la demanda de sandía aumenta en 119,76 millones de libras. Si el PIBPC2005 se eleva en un dólar, la demanda de sandía disminuye en 0,05 millones de libras.

### 1.4.2 Antecedentes Nacionales

Mendez (2018) en su tesis “Modelo matemático predictivo del comportamiento de la exportación de fresa (*Fragaria vesca* L.) congelada del Perú al mercado asiático frente a factores externos e internos”, se planteó el objetivo de construir un modelo matemático para pronosticar las exportaciones peruanas de fresa congelada. Para tal fin se hizo uso del análisis de regresión múltiple con series temporales anuales. Las series tuvieron una extensión de cinco años (2012-2016). Se tomaron en cuenta tres variables independientes cuantitativas y dos cualitativas: importación (t), precio (US\$/kg), consumo per cápita (kg/persona), barrera sanitaria (Con/Sin) y país (Japón/China).

El modelo superó las pruebas de significancia global, de autocorrelación y de normalidad. El poder explicativo del modelo se estableció en 89,9%, es decir el conjunto de variables independientes explicaron el 89,9% del comportamiento de la exportación de fresa congelada peruana al mercado asiático. El modelo construido fue el siguiente:

$$\text{Exportación} = - 1\,086,725 - 0,23 \text{ Importación} \times \text{Barrera} + 1\,287,344 \text{ Precio} \times \text{Barrera} + 1\,891,864 \text{ Consumo} \times \text{Barrera} + 0,015 \text{ Importación} \times \text{Precio} \times \text{Consumo} - 3\,913,899 \text{ País}$$

Barreto (2018) realizó el estudio “Modelo de regresión del Consumo final privado y el Producto bruto interno nacional del Perú durante el periodo 2000-2013”. El primer objetivo específico fue la construcción de una ecuación de regresión para relacionar el consumo final privado (variable dependiente) y el PBI (variable independiente). El segundo objetivo específico fue pronosticar, para el periodo 2014-2016, el consumo final privado en base al PBI. Se utilizó el análisis de regresión simple con series temporales anuales. Las series temporales tuvieron una extensión de 14 años (2000-2013). Se construyeron cuatro modelos: exponencial, potencial, logarítmico e inverso. De los modelos construidos se seleccionó al modelo exponencial:

$$\text{Consumo final privado} = 73\,936,56 e^{0,00003003 \times \text{PBI}}$$

El modelo superó las pruebas de significancia individual de  $\beta_0$  y  $\beta_1$ . El modelo presentó el más bajo error de estimación (0,009) y el más alto poder explicativo, el PBI explicó en un 99,9% la variación del consumo final privado.

Palacios et al. (2021) en su tesis “Determinantes del incremento del número de clientes residenciales de gas natural en la concesión de Lima y Callao: 2007-2019”, construyeron un modelo log-log para pronosticar del número de clientes residenciales de gas natural en Lima y Callao para el periodo 2007-2019 (DCLIENTES). Se utilizó el análisis de regresión múltiple con series temporales trimestrales. Las series temporales tuvieron una extensión de 49 trimestres (III trimestre 2007-III trimestre 2019). Las variables independientes fueron el aumento de las redes de distribución de gas natural (DREDESTOT), el costo de acceso a la red de gas natural (CTOACCESO), el ahorro por reemplazar el GLP por el gas natural (AHORROGGLP), el ahorro por reemplazar la energía eléctrica por el gas natural (AHORROGEE), el ingreso mensual familiar de hogares con redes de gas natural (INGRESODCR) y una variable dummy para el periodo en que se dejó de subsidiar la conexión (DUMMYSUBSIDIO).

El modelo alcanzó un alto poder explicativo, las variables independientes explicaron el 96,86% de la variación del número de clientes de gas natural en Lima y Callao. Además el modelo superó las pruebas de significancia global, de significancia individual, de multicolinealidad, de autocorrelación, de normalidad, de estabilidad paramétrica, de cointegración y de heterocedasticidad. El modelo resultante fue el siguiente:

$$\begin{aligned} \text{Log DCLIENTES} = & - 13,30272 + 0,557043 \text{ Log DREDESTOT} - 0,597441 \text{ Log} \\ & \text{CTOACCESO} + 1,798054 \text{ Log AHORROGGLP} + 2,236270 \text{ Log AHORROGEE} + \\ & 1,307878 \text{ Log INGRESODCR} + 0,583032 \text{ DUMMYSUBSIDIO} \end{aligned}$$

Según el modelo si DREDESTOT se incrementa en 1%, DCLIENTES crece en 0,56%. Si CTOACCESO se expande en 1%, DCLIENTES se contrae en 0,60%. Si

AHORROGLP aumenta en 1%, DCLIENTES crece en 1,80%. Si AHORROGEE se eleva en 1%, DCLIENTES sube en 2,24%. Si INGRESODCR mejora en 1%, DCLIENTES se incrementa en 1,31%.

### **1.5 Justificación de la Investigación**

El análisis de regresión lineal es una herramienta potente que, entre otras aplicaciones, nos permite pronosticar la demanda en el ámbito de los proyectos de inversión. El pronóstico de demanda, en los proyectos de inversión, suele hacerse para un horizonte temporal de cinco años, y puede ser bastante preciso, siempre y cuando el modelo construido cumpla con todos los supuestos inherentes al método de mínimos cuadrados ordinarios y, además, sea significativo. Sin embargo, a pesar de todo lo estudiado y escrito acerca del tema, en múltiples estudios no se le ha dado la debida importancia y atención al pronóstico de demanda; algunos ejemplos se presentan a continuación:

Chávez y Juscamaita (2013), pronosticaron la demanda estadounidense de cacao para ocho años (2011-2018), aplicando el análisis de regresión simple. Se usó el modelo lineal, tomando al tiempo como variable independiente. Se utilizó una serie temporal compuesta de ocho años (2003-2010). No se validó el modelo ni se calculó la precisión del pronóstico.

De La Cruz (2014), utilizó el análisis de regresión simple con una serie temporal de cinco años (2010-2014) para pronosticar la demanda, en Lima Metropolitana, de un complemento nutricional. El horizonte del pronóstico fue de cinco años (2015-2019). Se optó por un modelo polinómico de segundo grado, siendo el tiempo la variable independiente. No se validó el modelo ni se calculó la precisión del pronóstico.

Castillo y Torres (2016), aplicaron el análisis de regresión simple con una serie temporal compuesta de seis años (2009-2014) para pronosticar la demanda de jabón líquido. El horizonte del pronóstico fue de siete años (2015-2021). Se seleccionó un modelo lineal,

siendo el tiempo la variable independiente. No se validó el modelo ni se calculó la precisión del pronóstico.

Pino (2017), hizo uso del análisis de regresión simple con una serie temporal con 25 años de extensión (1991-2015), para pronosticar la demanda norteamericana de aceitunas rellenas de pimiento. El horizonte del pronóstico fue de seis años (2016-2021). Se eligió un modelo potencial, siendo el tiempo la variable independiente. No se validó el modelo ni se calculó la precisión del pronóstico.

Curo e Ybañez (2017), utilizaron el análisis de regresión simple con una serie temporal con extensión de 16 años (2000-2015), para pronosticar la demanda de néctar de fruta en Lima Metropolitana. El horizonte del pronóstico fue de siete años (2016-2022). Se eligió un modelo lineal, siendo el tiempo la variable independiente. No se validó el modelo ni se calculó la precisión del pronóstico.

Como ya se ha mencionado, el pronóstico de la demanda es la piedra angular en la formulación y evaluación de proyectos de inversión. El pronóstico de la demanda posibilita la determinación de la capacidad de producción adicional a instalar o ampliar, con lo que es posible presupuestar la inversión inicial en activos fijos, los ingresos por ventas y los costos de producción. La información mencionada es esencial en la construcción de los flujos de caja del proyecto y, por ende, en el cálculo del VAN y de la TIR, sobre los cuales se decide la ejecución del proyecto. Es decir, el pronóstico de la demanda debe ser lo suficientemente preciso, de tal manera que al final se calcule un VAN y una TIR confiables.

De los argumentos se deduce que el pronóstico de demanda debe ser lo más preciso posible, con la finalidad de minimizar el riesgo de invertir importantes sumas de dinero y tomar una decisión correcta acerca de la materialización de un proyecto de inversión. Cabe resaltar que se requiere invertir entre US\$5 y US\$7 millones para instalar una planta de productos hidrobiológicos congelados de 100 t/día de producción (Del Castillo, 2016).

Por las consideraciones expuestas, la investigación se justifica en razón a que pretende poner en evidencia, utilizando el análisis de regresión lineal, los factores que se relacionan con la precisión de los pronósticos de productos hidrobiológicos congelados en el Perú. Además, los hallazgos de esta investigación podrían extrapolarse al pronóstico de la demanda de otros productos, o bien podrían ser la base de futuras investigaciones que aborden el mismo tema. Finalmente, un anhelo de esta investigación es incrementar el conocimiento de un campo poco estudiado como es el pronóstico de demanda en la formulación de proyectos de inversión, utilizando el método de regresión lineal con series temporales.

### **1.6 Limitaciones de la Investigación**

Para la realización de esta investigación se han podido identificar las limitantes siguientes:

- Considerando que el pronóstico de la demanda es una variable compleja y dinámica, la presente investigación se limita a determinar, de qué manera tres factores podrían afectar la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú. Estos factores son el número de variables independientes, la extensión de las series temporales y la validación del modelo.
- Para determinar la precisión de los pronósticos, se construyeron modelos de regresión lineal con series temporales anuales. Una limitación del estudio es que las series temporales, empleadas para la construcción de los modelos, se circunscribieron al periodo entre 1980-2013. Y la data del periodo 2014-2018 se utilizó para realizar los pronósticos y establecer su precisión.
- Las variables independientes que sirvieron para la construcción de los modelos, solo explican el comportamiento de la demanda de productos hidrobiológicos congelados en el

Perú. No se debe asumir que, necesariamente este modelo, con las mismas variables tendrán similar poder explicativo en la demanda de otros productos.

- Otra limitación para el desarrollo del presente estudio es que algunos organismos oficiales tardan en procesar y publicar sus estadísticas más recientes. En diciembre de 2018, el PRODUCE publicó la versión 2017 de su anuario estadístico pesquero y acuícola. A inicios de enero de 2020, aún no se había publicado la versión 2018 del anuario. Por lo cual, en este caso, no se tiene acceso oportuno a información estadística oficial.
- Finalmente, y por ello no menos importante, es que no se encontraron investigaciones específicas referentes a la precisión de los pronósticos de demanda para periodos anuales, una limitación que no permite una contrastación con investigaciones de igual naturaleza.

## **1.7 Objetivos de la Investigación**

### ***1.7.1 Objetivo General***

Evaluar la relación existente entre el análisis de regresión lineal y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

### ***1.7.2 Objetivos Específicos***

**Objetivo Específico 1.** Precisar la relación que existe entre la extensión de las series temporales y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

**Objetivo Específico 2.** Determinar la relación del número de variables independientes con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

**Objetivo Específico 3.** Comparar la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú, entre un modelo de regresión lineal validado y un modelo de regresión lineal no validado.

## 1.8 Hipótesis

Las hipótesis de investigación, hipótesis de trabajo o simplemente hipótesis ( $H_i$  o  $H_1$ ,  $H_2$ ,  $H_3$ , etc., cuando hay más de una) son enunciados que manifiestan la posibilidad que, por lo menos dos variables, se relacionen de alguna manera. Y, las hipótesis nulas ( $H_0$ ), son enunciados que contradicen lo manifestado por las hipótesis de investigación (Hernández Sampieri y Mendoza, 2018).

### 1.8.1 Hipótesis General

El análisis de regresión lineal se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

### 1.8.2 Hipótesis Específicas

**Hipótesis Específica 1 ( $H_1$ ).** Existe una relación positiva entre la extensión de las series temporales y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

**Hipótesis Específica 2 ( $H_2$ ).** El número de variables independientes se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

**Hipótesis Específica 3 ( $H_3$ ).** Un modelo de regresión lineal validado presenta mayor precisión en el pronóstico de demanda de productos hidrobiológicos congelados en el Perú, frente a un modelo de regresión lineal no validado.

## II Marco Teórico

### 2.1 Marco Conceptual

#### 2.1.1 *Los Pronósticos*

Un pronóstico es una predicción de lo que ocurrirá, resulta del procesamiento y análisis de información y datos disponibles (Goodwin, 2018). Sin embargo, no solo se debe pronosticar el hecho, si no también el momento de la ocurrencia y su magnitud (Kumar y Suresh, 2009). Y se debe pronosticar con cierto nivel de precisión (González-Rivera, 2016).

Realizar pronósticos se define como el arte-ciencia que consiste en vaticinar acontecimientos que ocurrirán en el futuro (Heizer et al., 2017; Sankaran et al., 2019). En ese sentido, la obtención de buenos pronósticos frecuentemente requiere combinar ciencia e intuición (Stevenson, 2018). También se podría afirmar que pronosticar es una especie de arte o que depende del azar, y no una ciencia ciento por ciento precisa (Pirc et al., 2016). Aunque, antes de la invención de la computadora y de programas especializados, pronosticar era mucho más arte que ciencia (Schroeder y Goldstein, 2018).

Por un lado, pronosticar es un arte porque reconoce cuando es necesario y como incorporar información cualitativa y cuantitativa en el proceso y, por otro lado, es una ciencia porque la construcción de un modelo se fundamenta en principios científicos (Hoshmand, 2010). A veces la especificación del modelo parece ser un arte, mientras que su estimación es ciencia (Elff, 2015). Pronosticar también es un arte por que a medida que los pronosticadores adquieren experiencia, mejora la precisión de sus pronósticos (Harris, 2015; Shim, 2009).

Los fines primordiales de los pronósticos son minimizar la incertidumbre y tratar de obtener las mejores estimaciones de los acontecimientos futuros (Render et al., 2018).

#### 2.1.2 *Demanda*

En términos económicos, la demanda es la cantidad de cierto bien o servicio que una unidad familiar podría comprar durante cierto periodo, si pudiera comprar todo lo deseado al

precio actual de mercado, mientras que la demanda de mercado es el total del bien o servicio, que es demandado por periodo por el total de unidades familiares, que compran en el mercado de aquel bien o servicio (Case et al., 2017). Y la demanda agregada es la cantidad real de producción interna que, de manera global, los consumidores internos y externos, las empresas y el gobierno pretender comprar según el precio (Guell, 2018).

En el contexto de los proyectos de inversión, la demanda se puede clasificar en demanda de mercado y demanda para el proyecto. La demanda de mercado es la cantidad demandada por un grupo de consumidores por unidad temporal en cierto territorio, mientras que la demanda para el proyecto es la porción de la demanda de mercado que el proyecto abastecerá (Sapag et al., 2014).

En el mismo contexto, la demanda también se puede clasificar en potencial, efectiva e insatisfecha. La demanda potencial son los requerimientos latentes que no necesariamente se materializan, mientras que la demanda efectiva es la porción de la demanda potencial que se materializa y la demanda insatisfecha es la que no logra ser abastecida por la oferta (Carbonel, 2016).

### ***2.1.3 Pronóstico de Demanda***

Pronosticar la demanda implica predecir por adelantado su propio comportamiento futuro (Shahabuddin, 2009). Una empresa puede pronosticar su demanda de dos formas, puede pronosticar la demanda de mercado (pronóstico de mercado) y determinar qué porcentaje abastecerá, o bien puede pronosticar directamente su demanda (Hollensen, 2017).

### ***2.1.4 Importancia de la Precisión de un Pronóstico de Demanda***

La precisión del pronóstico determinará la calidad de las decisiones y por ende la materialización de los objetivos y metas de las empresas (Sankaran et al., 2019), por lo cual, un pronóstico de demanda preciso y oportuno determinará el éxito de los negocios (Min, 2015). Por tanto, el pronóstico de demanda tiene vital importancia en todos los contextos y

áreas empresariales, tanto en la creación de nuevas empresas, como en su crecimiento y también en el desarrollo normal de sus actividades. Un pronóstico preciso de demanda para un espacio temporal apropiado, hace posible que las empresas inviertan en la construcción o ampliación de fábricas, en adquisición de maquinarias, materias primas e insumos y en la contratación de personal (Gilliland, 2015).

Por un lado, la decisión de crear una nueva empresa o la de expandir las actividades de una existente, debe basarse en los resultados obtenidos de la formulación y evaluación de un proyecto de inversión. En general, el pronóstico de demanda juega un rol crucial en los proyectos de inversión, debido a sus repercusiones en la rentabilidad (Sapag et al., 2014), a medida que disminuye la precisión del pronóstico, aumenta el riesgo de la inversión (Jobber y Lancaster, 2015). El pronóstico de demanda es un insumo vital para la construcción de los flujos de caja, a partir de los cuales se obtienen los indicadores de rentabilidad del proyecto, y sobre los que se decide la materialización o no del proyecto. Además, los flujos de caja pronosticados de manera precisa, permiten conocer dificultades de antemano y por tanto planear las medidas correctivas pertinentes (Ferrell et al., 2018).

En particular, el pronóstico de demanda, es el factor más relevante en la determinación de la capacidad de producción de la futura instalación (Carbonel, 2016; Sapag et al., 2014). En ese sentido, un pronóstico de demanda poco preciso puede propiciar dos situaciones extremas no deseadas. Puede subestimar la capacidad y ésta quizás no sea capaz de producir todo lo demandado, lo que ocasionaría pérdida de clientes; o puede sobreestimar la capacidad, incrementando los costos operativos debido a la excesiva capacidad ociosa (Ferrell et al., 2018).

Otro aspecto relacionado con los proyectos de inversión, es que el nivel de precisión requerido de un pronóstico está directamente relacionado con el valor monetario de la decisión (Hoshmand, 2010). Si una decisión involucra la inversión de grandes montos de

dinero, entonces el pronóstico debe ser suficientemente preciso (Jacobs y Chase, 2018), a pesar de que los pronósticos de alta precisión normalmente son más onerosos para modelar y aplicar (Kumar y Suresh, 2009; Reid y Sanders, 2013) y solo son convenientes para pronósticos de mediano-largo plazo de series temporales agregadas (Silver et al., 2017). En suma, la creación de nuevas empresas o la ampliación de las existentes, requiere la inversión de considerables recursos monetarios, por lo que es necesario contar con un pronóstico de demanda lo más preciso posible.

Por otro lado, en el contexto de empresas en operación, la planificación de sus actividades también requiere de pronósticos suficientemente precisos (Kumar y Suresh, 2009). Es decir, no se puede hacer un buen plan si no se cuenta con un pronóstico razonablemente preciso (Vu, 2015). Por tanto, luego de construir pronósticos precisos de demanda, las empresas pueden empezar a planear las actividades propias del negocio (Bozarth y Handfield, 2019). Es así que, el desarrollo exitoso de actividades tales como planeación de inventarios, planeación logística, programación de la producción, proyección de flujos de caja, decisiones de compras; pueden depender exclusivamente del pronóstico de demanda (Goodwin, 2018).

Obtenido el pronóstico de demanda, la gerencia puede calcular los costos y beneficios proyectados del producto, que incluyen los costos de mercadeo, investigación y desarrollo, operaciones, contabilidad y finanzas (Armstrong y Kotler, 2017). También, un pronóstico de demanda preciso hace posible que el departamento de compras solicite la cantidad correcta de materias primas e insumos, que operaciones manufacture la cantidad adecuada de producto y que logística envíe la cantidad apropiada de producto (Wisner et al., 2019).

En concreto, si el pronóstico tiene sesgo optimista, la empresa no podrá vender todo el inventario; pero si el sesgo es pesimista, la empresa puede perder oportunidades de mercado por no tener suficiente inventario (Hollensen, 2017); y si la situación persiste, puede perder

clientes (Harrison et al., 2014). Por lo que una compañía que incrementa la precisión de sus pronósticos y que factura US\$ 1 000 millones, podría reducir sus inventarios entre US\$ 5-10 millones; además los errores evitables en los pronósticos pueden incrementar los costos de producción entre 2% y 4% (Goodwin, 2018).

Y aunque una pequeña variación pareciera no ser trascendente, Bunn y Farmer (1985) encontraron que un aumento de apenas un punto porcentual en el error del pronóstico, pudo provocar el incremento de los costos operacionales de una empresa británica de servicios públicos en 10 millones de libras (citado por Dannecker, 2015). Y suponiendo un error de pronóstico de 5%, Hobbs et al. (1999), concluyeron que las pérdidas financieras anuales de dos empresas norteamericanas de servicios públicos, se situarían entre 1,8 y 4,5 millones de dólares (citado por Dannecker, 2015).

### ***2.1.5 Medidas de Precisión del Pronóstico***

Un pronóstico siempre se desviará de los datos reales en mayor o menor medida, por lo que es necesario medir la precisión del pronóstico. La medición del desempeño del pronóstico se realiza para cuantificar que tan bien predice y para comparar diferentes modelos estadísticos y establecer cual se ajusta mejor a la demanda histórica y por tanto pronostica mejor la demanda futura (Chase, 2013).

La Desviación Media Absoluta (DMA), el Error Cuadrado Medio (ECM) y el Error Medio Absoluto Porcentual (EMAP) son las medidas más comunes para evaluar la precisión de un pronóstico (Díaz et al., 2011; Evans, 2017; González-Rivera, 2016; Stevenson, 2018; Taylor, 2016). Todas las medidas mencionadas parten del error del pronóstico. En particular, cuando se pronostica la demanda de algún producto, el error se calcula como sigue:

$$\text{Error de pronóstico } (e_t) = \text{Demanda real} - \text{Demanda pronosticada}$$

Un error grande puede revelar el uso de un método de pronóstico inadecuado o que el método necesita ser ajustado (Taylor, 2016).

**2.1.5.1 Desviación Media Absoluta (DMA).** La DMA o MAD (Mean Absolute Deviation) se calcula promediando la sumatoria de los valores absolutos de los errores. La MAD está en las mismas unidades que la serie original. Mientras más pequeño es la DMA, más grande es la precisión del pronóstico. Matemáticamente la DMA se expresa como:

$$DMA = \frac{\sum |e_t|}{n}$$

**2.1.5.2 Error Cuadrático Medio (ECM).** El ECM o MSE (Mean Squared Error) se calcula promediando los cuadrados de los errores. Un inconveniente del ECM es que acentúa los errores más grandes debido al término al cuadrado. Mientras más pequeño es el ECM, más grande es la precisión del pronóstico. Matemáticamente el ECM se expresa como:

$$ECM = \frac{\sum e_t^2}{n}$$

**2.1.5.3 Error Medio Absoluto Porcentual (EMAP).** Un problema con la DMA y con el ECM es que están influenciados por la magnitud de los datos. Si los datos se expresan en miles, la DMA y el ECM pueden ser muy grandes. Para evitar este problema se usan medidas de precisión porcentuales como el EMAP (MAPE, Mean Absolute Percent Error).

El EMAP establece que tan grande es el error del pronóstico (Bozarth y Handfield, 2019) y mide mejor su precisión (Vu, 2015), por lo cual es la medida de precisión más ampliamente utilizada (Chase, 2016; Dhrymes, 2017; Kolassa y Siemsen, 2016; Konasani y Kadre, 2015; Tofallis, 2015) y es la recomendada por la mayoría de libros de texto (Lee y Lee, 2017).

El EMAP es de gran utilidad cuando los valores de la variable dependiente son grandes y es posible emplearlo para cotejar la precisión de un mismo o de varios métodos (Hanke y Wichern, 2014). Además, se entiende fácilmente y se puede emplear para dirimir entre pronósticos, cuando DMA y ECM son contradictorios (Ivanov et al., 2017).

El EMAP se calcula al expresar en términos porcentuales el promedio de los valores absolutos de los errores. Matemáticamente el EMAP se expresa como:

$$\text{EMAP} = \frac{(\sum |e_t| / D_t) (100)}{n}$$

Aunque ha habido cierta discusión acerca de si el denominador del error debería ser la demanda pronosticada en lugar de la demanda real, la mayoría de programas emplean la demanda real como divisor (Goodwin, 2018).

EMAPs pequeños implica que los datos históricos se ajustan bien al modelo y por tanto los pronósticos son más precisos (Gee, 2015; Taylor, 2016), por lo cual es deseable un EMAP máximo de 5% (Konasani y Kadre, 2015). Sin embargo, cuando en el modelo se omiten variables independientes relevantes, el EMAP tiende a crecer a lo largo del tiempo (Pasupathy, 2010).

Ghiani et al. (2013) clasificaron en cuatro categorías la precisión del pronóstico, según el valor alcanzado por el EMAP, tal como se presenta en la Tabla 1.

**Tabla 1**

*Precisión del Pronóstico Según el EMAP*

Rango	Precisión
$0 < \text{EMAP} \leq 10\%$	Muy buena
$10\% < \text{EMAP} \leq 20\%$	Buena
$20\% < \text{EMAP} \leq 30\%$	Moderada
$\text{EMAP} > 30\%$	Pobre

*Nota:* Adaptado de *Evaluation of the quality of a forecasting method by means of the mean absolute percentage deviation* (p. 99), por Ghiani et al., 2013, *Introduction to logistics systems management* (2nd ed.), Wiley.

**2.1.5.4 Sesgo.** El sesgo consiste en un error sistemático, en donde el pronóstico de demanda siempre se encuentra o por encima o por debajo de la demanda observada

(Chapman et al., 2017). El sesgo ocurre, entre otras causas, por incluir variables irrelevantes y por especificar mal la relación entre las variables (Jacobs y Chase, 2018).

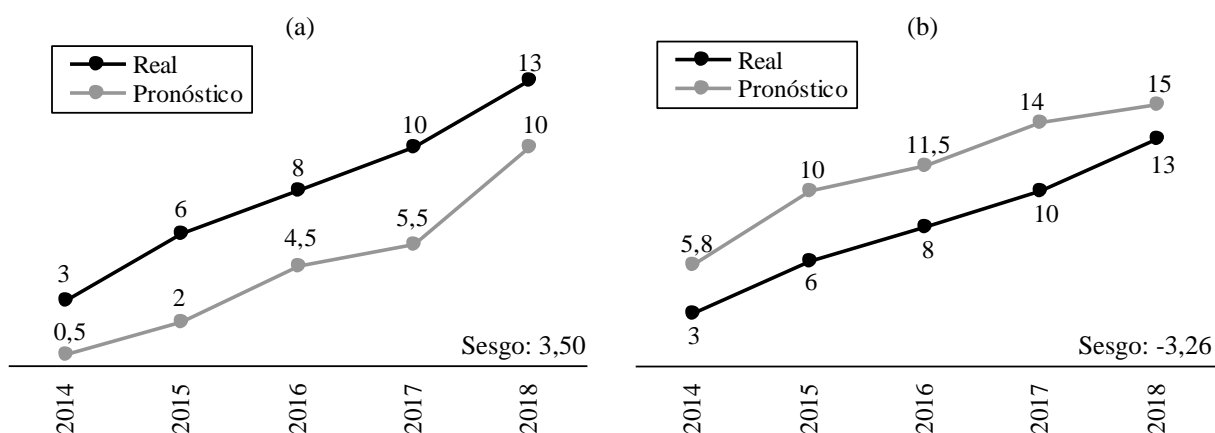
Cuando el pronóstico no presenta sesgo, el error flota aleatoriamente por encima y por debajo de cero (Nahmias y Olsen, 2015). Ante la presencia de sesgo, el pronóstico debe ser revisado y posiblemente se deba cambiar por uno que incremente su precisión (Chapman et al., 2017). Matemáticamente el sesgo se expresa como:

$$\text{Sesgo} = \frac{\sum e_t}{n}$$

El sesgo puede tomar cualquier valor. Cuando es mayor a cero el pronóstico tiende a subestimar la demanda (Figura 1a) y cuando es menor a cero el pronóstico tiende a sobreestimarla (Figura 1b) (Groebner et al., 2018).

### Figura 1

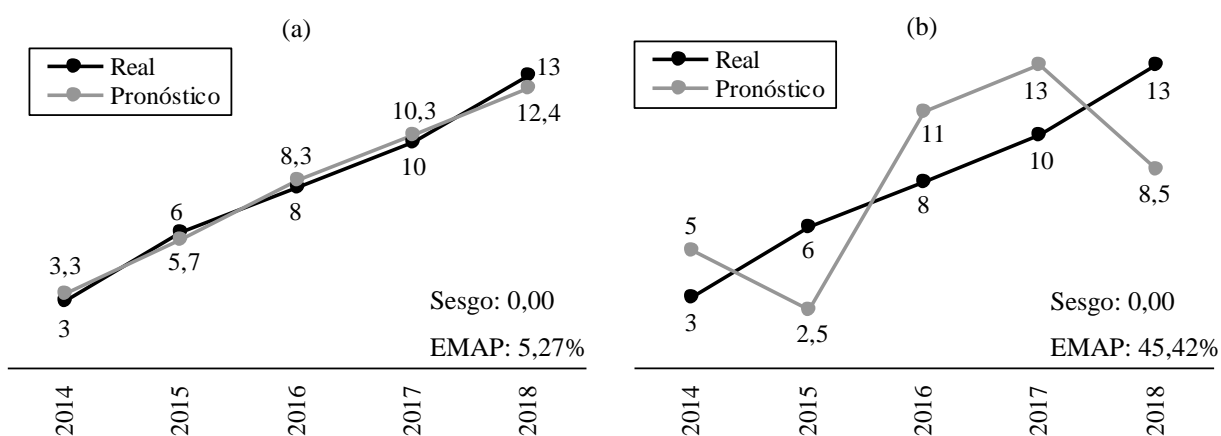
*Pronóstico con Sesgo (a) Mayor a Cero y (b) Menor a Cero*



Finalmente, cuando el sesgo toma un valor de cero o es muy próximo a él, entonces el pronóstico no tiene sesgo (Hanke y Wichern, 2014). Sin embargo, pronósticos insesgados pueden presentar muy poca precisión (Clements, 2005). Es decir, es posible que un pronóstico con sesgo igual a cero sea muy preciso (Figura 2a), o bien otro pronóstico, igualmente insesgado, sea muy impreciso (Figura 2b).

**Figura 2**

*Pronóstico Inseguro (a) con Precisión Muy Buena y (b) con Precisión Pobre*



### **2.1.6 Horizontes Temporales de los Pronósticos de Demanda**

Todos los autores coinciden en que los horizontes temporales de los pronósticos son tres: corto, mediano y largo. Los pronósticos de corto plazo se utilizan en la planificación de compras, tamaño de la fuerza laboral, programación y asignación laboral y en establecer los niveles de producción. Mientras que los pronósticos de mediano plazo son útiles en la planificación de ventas, planificación de la producción, presupuesto de caja y análisis de planes operativos. Y los pronósticos de largo plazo se usan en la planificación de nuevos productos, desembolsos de capital, localización o ampliación de fábricas y en investigación y desarrollo (Heizer et al., 2017).

En lo que no están de acuerdo los autores es en la duración de los horizontes temporales, aunque la mayoría coincide en que el corto plazo equivale a menos de tres meses, el mediano plazo comprende entre tres meses y dos años y el largo plazo es para periodos mayores a los dos años (Tabla 2).

**Tabla 2***Duración de los Horizontes Temporales de los Pronósticos Según Autores*

Corto Plazo	Mediano Plazo	Largo Plazo	Autores
Menos de 2 meses	2 meses - 1 año	Más de 1 año	Taylor (2016, p. 720)
			Groebner et al. (2018, p. 661)
			Jacobs y Chase (2018, p. 448)
			Krajewski et al. (2008, p. 525)
Menos de 3 meses	3 meses - 2 años	Más de 2 años	Lysons y Farrington (2016, pp. 305-306)
			Min (2015, p. 102)
			Shim (2009, p. 117)
			Waters (2011, p. 250)
Menos de 6 meses	6 meses - 2 años	Más de 2 años	Hoshmand (2010, p. 19)
			Schroeder y Goldstein (2018, p. 189)
Menos de 3 meses	3 meses - 3 años	Más de 3 años	Heizer et al. (2017, p. 108)
			Ivanov et al. (2017, p. 304)
Menos de 1 año	1 - 5 años	Más de 5 años	Kumar y Suresh (2009, p. 56, 105)
			Nash y Nash (2001, pp. 10-11)

**2.1.7 Clasificación de los Pronósticos**

Jacobs y Chase (2018) clasificaron a los pronósticos, según su propósito, en estratégicos y tácticos. Por un lado, los autores afirmaron que los pronósticos estratégicos ayudan a implantar la estrategia para satisfacer la demanda agregada; y que son los más convenientes para tomar decisiones sobre tamaño de planta, diseño de procesos de producción y de servicio, localización y distribución de planta, y planeación de ventas y operaciones. Por otro lado, los investigadores señalaron que los pronósticos tácticos se utilizan para procesos rutinarios, cuantifican la demanda para varias semanas o meses y aseguran que el tiempo de espera de los compradores esté acorde con sus expectativas.

Por su parte, Heizer et al. (2017) afirmaron que las organizaciones emplean tres grandes tipos de pronósticos: económicos, tecnológicos y de demanda. Los autores sostuvieron que los pronósticos de indicadores económicos tales como inflación, oferta monetaria, producto bruto interno y otros, direccionan el ciclo de los negocios. También los

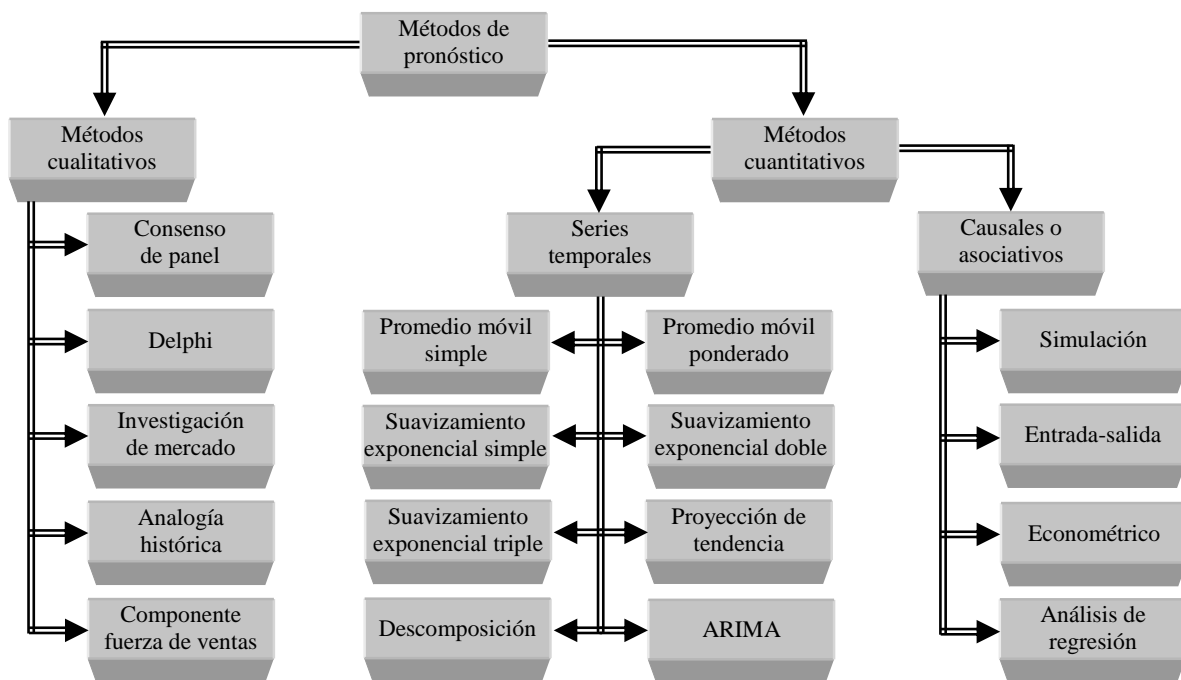
investigadores manifestaron que mediante los pronósticos tecnológicos es posible predecir el avance de la tecnología, el cual puede generar innovación productiva, la que a su vez requerirá la construcción de fábricas y el diseño de equipos. Por último, los autores precisaron que el nivel de producción, la capacidad instalada y los sistemas de programación son conducidos por el pronóstico de demanda, que su a vez se convierte en un insumo para las áreas de finanzas, marketing y personal.

### 2.1.8 Métodos de Pronóstico

Los métodos de pronósticos se dividen en dos grandes grupos: métodos cualitativos y métodos cuantitativos (Figura 3).

**Figura 3**

*Clasificación de los Métodos de Pronóstico*



**2.1.8.1 Métodos Cualitativos.** Mediante los métodos cualitativos se obtiene un pronóstico a partir de aspectos subjetivos tales como la intuición, las emociones y la experiencia de los tomadores las decisiones (Heizer et al., 2017). Dentro de esta categoría están:

**Consenso de Panel.** Un grupo de empleados de diversos grados de responsabilidad de una empresa, se reúnen abiertamente y libremente para intercambiar ideas, sin embargo, es posible que los aportes de los empleados de los más bajos niveles no sean considerados por los empleados de los niveles más altos (Jacobs y Chase, 2018).

**El Método Delphi.** Un grupo de expertos en un campo determinado se reúnen en un sitio predeterminado, ellos no son de la misma empresa, no se conocen o aún pueden ser extranjeros; y pueden no estar de acuerdo en ciertas cosas, pero si concuerdan en que probablemente sucedan, los consensos logrados de esa manera son usados como pronósticos (Reid y Sanders, 2013).

**Investigación de Mercado.** El método recaba información de clientes y potenciales clientes acerca de lo que pretenden comprar en el futuro; la información proporcionada puede ser útil en la preparación de un pronóstico, en el rediseño del producto y en la creación de nuevos productos (Render et al., 2018).

**Analogía Histórica.** Cuando se desea introducir en el mercado un producto nuevo, es posible pronosticar su demanda utilizando, como modelo, el comportamiento de la demanda histórica de un producto parecido o de un producto genérico que ya existe en el mercado (Jacobs y Chase, 2018).

**Componente Fuerza de Ventas.** En este método cada vendedor estima cuánto venderá en su zona o región, luego los estimados de cada vendedor se agregan a nivel distrital, regional o nacional, para formar el pronóstico total (Heizer et al., 2017).

**2.1.8.2 Métodos Cuantitativos.** Los métodos cuantitativos se fundamentan en el uso de modelos matemáticos, lo que los hace consistentes. Es decir, cada vez que se parta del mismo modelo y se utilice el mismo conjunto de datos, se obtendrá exactamente el mismo pronóstico. Una desventaja es que pueden requerir de mucha información (Reid y Sanders,

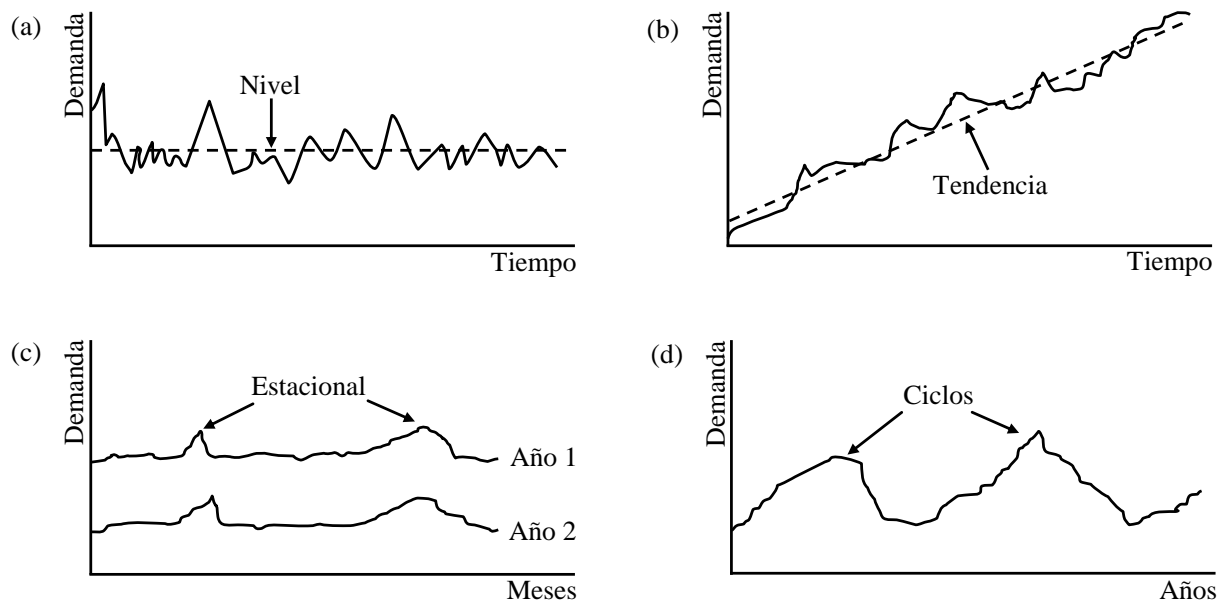
2013). Los métodos cuantitativos se dividen a su vez en métodos de series temporales y en métodos causales o asociativos.

**2.1.8.2.1 Métodos de Series Temporales.** Los métodos de series temporales pretenden pronosticar el futuro utilizando información histórica. Estos métodos asumen que lo que vaya a suceder en el futuro está relacionado con lo que ha ocurrido en el pasado (Render et al., 2018). La elección del método depende del comportamiento de la serie temporal, por ejemplo la demanda. La demanda puede presentar hasta cinco patrones o componentes:

- El patrón nivel u horizontal (L, Level) presenta movimientos en torno a una media constante (Figura 4a). Este es el patrón más fácil de pronosticar, es común en productos que se encuentran en la etapa de madurez de su ciclo de vida (Reid y Sanders, 2013).
- El patrón tendencia (T, Trend) muestra movimientos graduales ascendentes o descendentes en el tiempo (Figura 4b). A menudo estos movimientos son originados por cambios en la población, por variación en los ingresos y por cambios culturales (Stevenson, 2018).
- El patrón estacional (S, Seasonal) exhibe un comportamiento iterativo creciente y decreciente de corto plazo y de duración constante, puede darse en forma diaria, semanal, mensual, etc. (Figura 4c).
- El patrón ciclo (C, Cycle) es parecido al patrón estacional, la diferencia radica en que la duración de un ciclo es variable y de largo plazo, puede durar años o décadas (Figura 4d), y es poco pronosticable. Este patrón es originado por el ciclo de vida de los negocios y por el ciclo de vida del producto o servicio (Krajewski et al., 2016).
- El patrón aleatorio es una variación inexplicable que no puede ser pronosticada, no presenta un comportamiento reconocible.

### Figura 4

*Patrones de Demanda: (a) Nivel, (b) Tendencia, (c) Estacional y (d) Ciclo*



*Nota.* Adaptado de *Patterns of Demand* (p. 298), por Krajewski et al., 2016, *Operations Management. Processes and supply chains* (11th ed.), Pearson.

Si la demanda presenta nivel, se emplean el promedio móvil simple, promedio móvil ponderado y suavizamiento exponencial simple; si existe tendencia los métodos son proyección de tendencia y suavizamiento exponencial doble, y si es estacional se recomienda el suavizamiento exponencial triple (Ivanov et al., 2017).

Sin embargo, el promedio móvil y el suavizamiento exponencial no son apropiados en la formulación de proyectos, debido a que únicamente pronostican la demanda del periodo siguiente, y lo que se requiere es un pronóstico mínimo de cinco años (Baca, 2013).

- **Promedio Móvil Simple.** Este método obtiene el pronóstico ( $P_{t+1}$ ) al promediar la demanda ( $D$ ) de los últimos  $n$  periodos; luego calcula el pronóstico del periodo siguiente eliminando el dato más antiguo del promedio anterior y adicionando el de la demanda ya conocida. La expresión matemática del método es:

$$P_{t+1} = \frac{D_t + D_{t-1} + \dots + D_{t-n+1}}{n}$$

- **Promedio Móvil Ponderado.** Este método asigna ponderaciones ( $W_n$ ) a los datos utilizados en la obtención del pronóstico, la asignación es decreciente, al dato más reciente se les asigna el mayor peso ( $W_1 > W_2 > \dots > W_n$ ), la sumatoria de pesos siempre es uno. La expresión matemática del método es:

$$P_{t+1} = W_1D_t + W_2D_{t-1} + \dots + W_nD_{t-n+1}$$

- **Suavizamiento Exponencial Simple (SES).** Conocido también como suavizamiento de primer orden, en este método cada pronóstico se calcula a partir de dos sumados, el primero se obtiene multiplicando la constante de suavizamiento ( $\alpha$ ) y la demanda previa ( $D_t$ ), mientras que el segundo es el producto del complemento de  $\alpha$  y el pronóstico anterior ( $P_t$ ). La elección de  $\alpha$  se puede realizar mediante prueba-error, con el error del pronóstico como criterio de decisión. Usualmente  $\alpha$  está entre 0,05 y 0,50, los valores pequeños son para series con tendencias estables, mientras que los valores grandes son para series con tendencias cambiantes (Stevenson, 2018). Frecuentemente se asigna un  $\alpha$  de 0,1 para productos consolidados y la asignación es de 0,5 para productos que se encuentran al principio o al final de su ciclo de vida (Schönsleben, 2016). Otro criterio es elegir como óptimo a aquel  $\alpha$  que entregue el EMAP más pequeño (Chase, 2013; Hanke y Wichern, 2014; Wilson y Keating, 2009). La fórmula del método es:

$$P_{t+1} = \alpha D_t + (1 - \alpha) P_t$$

- **Suavizamiento Exponencial Doble (SED).** A este método se le conoce también como suavizamiento exponencial ajustado para tendencia, suavizamiento exponencial lineal, suavizamiento de segundo orden o método de Holt. El método es útil cuando la serie temporal presenta tendencia. Cuando una serie presenta tendencia, el pronóstico del método SES siempre será muy bajo si la tendencia es creciente y siempre será muy alto si la tendencia es decreciente (Goodwin, 2018). El método suaviza los componentes de las series mediante dos ecuaciones ( $L_t$  y  $T_t$ ) y dos constantes de suavizamiento ( $\alpha$  y  $\beta$ ), de tal manera

que los datos históricos se promedian y se elimina la aleatoriedad. El pronóstico ( $P_t$ ) es la suma de las dos ecuaciones. Las ecuaciones son:

$$L_t = \alpha D_t + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

$$P_{t+p} = L_t + pT_t$$

Donde:  $L_t$  = Suavizamiento exponencial del componente nivel en el periodo  $t$

$T_t$  = Suavizamiento exponencial del componente tendencia en el periodo  $t$

$\alpha$  = Constante de suavizamiento del componente nivel ( $0 \leq \alpha \leq 1$ )

$\beta$  = Constante de suavizamiento del componente tendencia ( $0 \leq \beta \leq 1$ )

$D_t$  = Demanda real en el periodo  $t$

$p$  = Períodos a pronosticar

$P_{t+p}$  = Pronóstico de demanda para  $p$  periodos

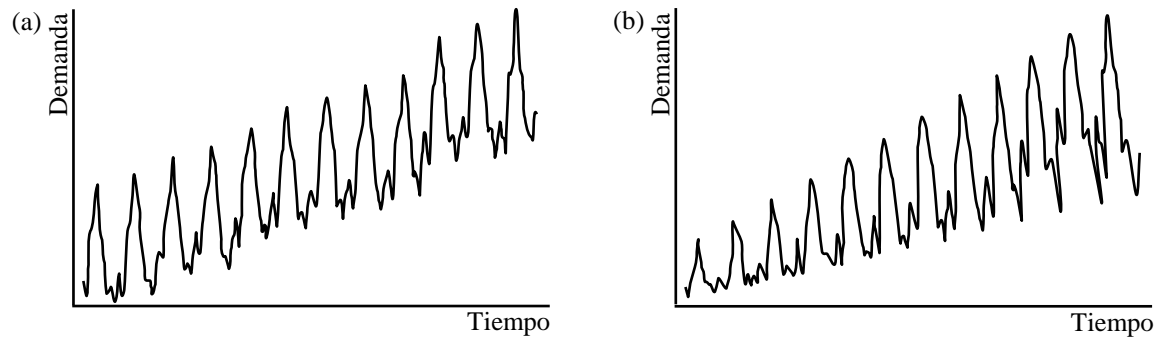
La elección de  $\alpha$  y  $\beta$  se puede realizar buscando la combinación que minimice el EMAP (Chase, 2013; Hanke y Wichern, 2014), los valores grandes originan variaciones muy raudas en las series y los valores pequeños generan variaciones menos raudas (Hanke y Wichern, 2014). Usualmente los valores de  $\alpha$  y  $\beta$  son muy pequeños, se ubican entre 0,1 y 0,2 (Albright y Winston, 2017; Reid y Sanders, 2013) o entre 0,1 y 0,3 (Jacobs y Chase, 2018); siendo  $\beta \leq \alpha$  para la mayoría de usos (Nahmias y Olsen, 2015). Y cuando  $\alpha = \beta$  se origina el suavizamiento exponencial doble de Brown (Goodwin, 2018).

- **Suavizamiento Exponencial Triple (SET).** Este método también es conocido como suavizamiento exponencial ajustado para tendencia y variación estacional, suavizamiento de tercer orden, método de Winter o método de Holt-Winter. El método SET se emplea en series temporales que presentan tendencia y estacionalidad. El método tiene un modelo para cada uno de los dos tipos de estacionalidad (Goodwin, 2018). El modelo de estacionalidad aditiva (Figura 5a) se recomienda cuando la estacionalidad es prácticamente constante en el

tiempo, y se emplea el modelo de estacionalidad multiplicativa (Figura 5b) cuando la estacionalidad varía en el tiempo (Hyndman y Athanasopoulos, 2018).

### Figura 5

*Comportamiento Estacional (a) Aditivo y (b) Multiplicativo*



Además, y según el ciclo de vida, el modelo aditivo se emplea para productos maduros y el modelo multiplicativo es el indicado para productos en crecimiento (Kolassa y Siemsen, 2016). De los dos, el modelo multiplicativo es más sencillo y ha tenido buen desempeño (Albright y Winston, 2017) por lo que normalmente al elegir, el modelo aditivo no es la primera opción (Chase, 2013).

El método SET suaviza los componentes de las series mediante tres ecuaciones ( $L_t$ ,  $T_t$  y  $S_t$ ) y tres constantes de suavizamiento ( $\alpha$ ,  $\beta$  y  $\gamma$ ). El pronóstico, en el modelo multiplicativo, es la suma de las dos primeras ecuaciones multiplicado por la tercera. Las ecuaciones del modelo multiplicativo son:

$$L_t = \alpha \frac{D_t}{S_{t-s}} + (1-\alpha) (L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma \frac{D_t}{L_t} + (1-\gamma) S_{t-s}$$

$$P_{t+p} = (L_t + pT_t) S_{t-s+p}$$

Donde:  $L_t$  = Suavizamiento exponencial del componente nivel en el periodo  $t$

$T_t$  = Suavizamiento exponencial del componente tendencia en el periodo t

$S_t$  = Suavizamiento exponencial del componente estacional en el periodo t

$D_t$  = Demanda real en el periodo t

$\alpha$  = Constante de suavizamiento del componente nivel ( $0 \leq \alpha \leq 1$ )

$\beta$  = Constante de suavizamiento del componente tendencia ( $0 \leq \beta \leq 1$ )

$\gamma$  = Constante de suavizamiento del componente estacional ( $0 \leq \gamma \leq 1$ )

p = Períodos a pronosticar

s = Longitud de la estacionalidad

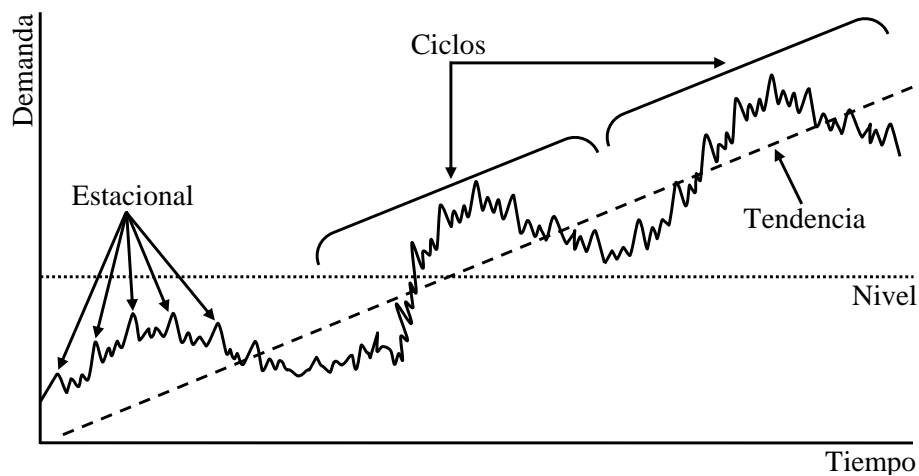
$P_{t+p}$  = Pronóstico de demanda para p periodos

La elección de  $\alpha$ ,  $\beta$  y  $\gamma$  se puede realizar buscando la combinación que minimice el EMAP (Chase, 2013). Para este fin, lo más utilizado es un algoritmo de optimización que calcule las constantes óptimas de suavizamiento (Hanke y Wichern, 2014). Sin embargo, si se quiere asegurar un pronóstico estable, las constantes de suavizamiento deberían tomar valores entre 0,1 y 0,2 (Nahmias y Olsen, 2015).

- **Proyección de Tendencia.** Este método ajusta una línea de tendencia a la serie temporal mediante los Mínimos Cuadrados Ordinarios (MCO) para calcular los parámetros de la recta ( $b_0$  y  $b_1$ ), y teniendo a la demanda (D) como variable dependiente y al tiempo (t) como variable independiente. El método se emplea para realizar pronósticos de mediano-largo plazo (Render et al., 2018). La expresión matemática del método es:

$$D = b_0 + b_1t$$

- **Descomposición.** El método consiste en aislar cada patrón de la serie temporal (Figura 6), luego realizar por separado el pronóstico de cada patrón y por último integrar los pronósticos individuales para obtener el pronóstico de la serie. El método se emplea para realizar pronósticos de corto y largo plazo (Hanke y Wichern, 2014).

**Figura 6***Patrones de una Serie Temporal*

*Nota.* Adaptado de *Demand Charted over 4 Years, with a Growth Trend and Seasonality Indicated* (p. 113), por Heizer et al., 2017, *Operations management. Sustainability and supply chain Management* (12th ed.), Pearson.

- **ARIMA (Autoregressive Integrated Moving Average).** Los modelos ARIMA presentan tres componentes. El primero es el componente autoregresivo (AR), en el que la demanda en cierto momento ( $D_t$ ) está influenciada por las demandas de momentos precedentes ( $D_{t-n}$ ) (Kolassa y Siemsen, 2016). La expresión matemática del modelo más simple, conocido como AR(1) es:

$$D_t = b_0 + b_1 D_{t-1} + \text{Error}_t$$

Y la expresión matemática del modelo AR(2) es:

$$D_t = b_0 + b_1 D_{t-1} + b_2 D_{t-2} + \text{Error}_t$$

El segundo es el componente integración (I), esto es, previo al análisis, se calculan las primeras diferencias ( $\Delta = D_t - D_{t-1}$ ) de la demanda, es decir el modelo se construye con los cambios en la demanda y no con la demanda en si (Kolassa y Siemsen, 2016). La expresión matemática del modelo AR(1)I(1) es:

$$D_t - D_{t-1} = b_0 + b_1 D_{t-1} - D_{t-2} + \text{Error}_t$$

$$\Delta D_t = b_0 + b_1 \Delta D_{t-1} + \text{Error}_t$$

El tercero es el componente media móvil (MA), en el que la demanda en cierto momento ( $D_t$ ) está influenciada por el error del momento anterior ( $\text{Error}_{t-1}$ ) (Kolassa y Siemsen, 2016). La expresión matemática del modelo MA(1) es:

$$D_t = b_0 + b_1 \text{Error}_{t-1} + \text{Error}_t$$

Sin embargo, el método ARIMA no es el indicado para todas las series temporales, lo es para aquellas en donde la tendencia y la estacionalidad no son los factores preponderantes (Chatfield, 2015). Ejemplos de tales series son el desempleo y la inflación (Elliott y Timmermann, 2016).

Una extensión de los modelos ARIMA son los modelos ARIMAX (Autoregressive Integrated Moving Average with eXogenous input). En estos modelos es posible incorporar variables independientes que expliquen significativamente el comportamiento de la demanda (Chase, 2016), la inclusión de las variables se puede realizar de manera aditiva o multiplicativa (Dannecker, 2015).

**2.1.8.2.2 Métodos Causales o Asociativos.** Los métodos causales o asociativos asumen que la variable que se pretende pronosticar es función de una o más variables. La tarea del pronosticador consiste en determinar las variables que matemáticamente se relacionan y utilizar esta información para realizar el pronóstico (Reid y Sanders, 2013). Estos métodos se desempeñan bien en pronósticos de largo plazo (Krajewski et al., 2016). Los métodos causales se pueden presentar en forma descriptiva como la simulación y el modelo entrada-salida; y estadística, como el modelo econométrico y el análisis de regresión (Ballou, 2004).

- **Simulación.** Este método imita en el tiempo el impacto de la demanda de bienes y/o servicios finales, sobre las necesidades en los distintos eslabones de la cadena de distribución y suministros (Ballou, 2004).

- **Entrada-Salida.** El método busca establecer interrelaciones entre industrias y sectores de una economía, de tal manera que la demanda de un producto puede ser pronosticada mediante el pronóstico de demanda de otros productos (Ghiani et al., 2004). Por ejemplo el pronóstico de demanda de alimentos enlatados puede ser de utilidad para pronosticar la demanda de envases de hojalata.
- **Econométrico.** Este método hace uso de la matemática, la estadística y la economía para construir y validar modelos que midan, expliquen y pronostiquen el comportamiento de variables económicas tales como la demanda. El modelo resultante no solo debe tener sentido estadístico, sino también económico. La técnica más utilizada en la construcción de modelos econométricos es el análisis de regresión (Brooks, 2014).
- **Análisis de Regresión.** De todos los métodos de los que dispone la estadística, el análisis de regresión es el más potente y útil (Siegel, 2016), por lo que es quizás el más ampliamente empleado (Montgomery et al., 2012). El método modela matemáticamente la relación funcional existente entre una variable dependiente y una o más variables independientes, de tal manera que sea posible explicar y pronosticar el comportamiento de la variable dependiente en función de una o más variables independientes. El método entregará un buen pronóstico siempre y cuando se logre reconocer y obtener datos de variables independientes que se relacionen estrechamente con la variable dependiente (Camm et al., 2019). La variable dependiente puede tomar el nombre de variable explicada, variable de respuesta, variable endógena o regresando. Mientras que las variables independientes pueden tomar el nombre de variables explicativas, variables de control, variables exógenas o regresores. Según Welc y Rodríguez (2018) la clasificación de los modelos de regresión se puede establecer por la forma funcional del modelo (lineal y no lineal), por el número de variables independientes (simple y múltiple), por el tipo de datos (series temporales y corte transversal), por la relación entre las variables (causal y sintomático), por la naturaleza de

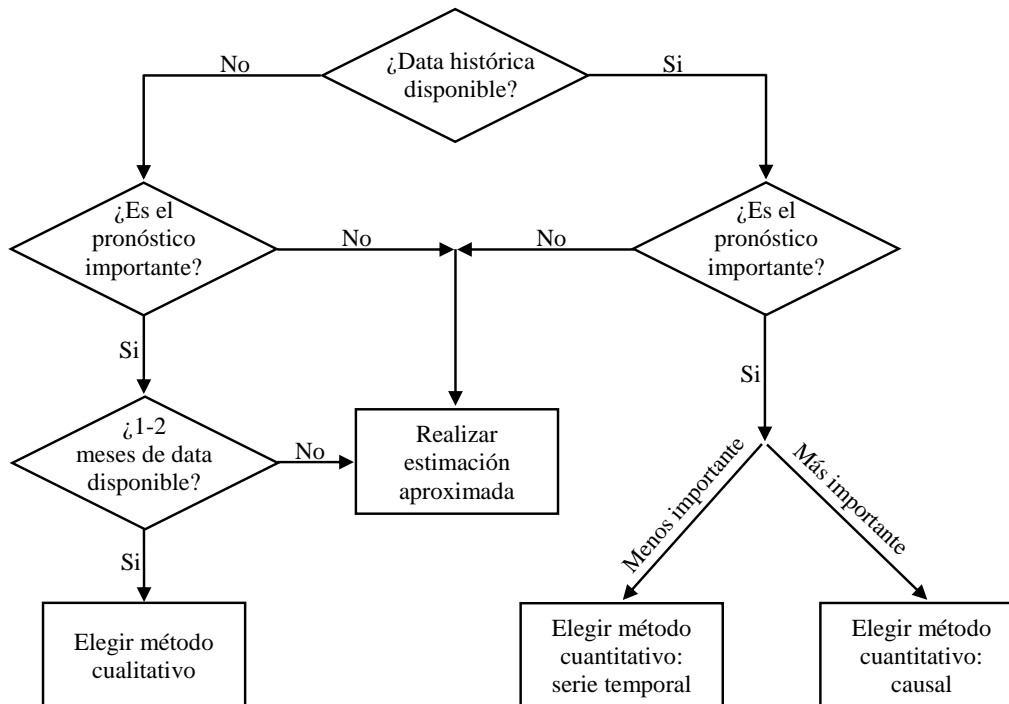
las variables (continuos y discretos) y por el número de ecuaciones (uniecual y multiecual).

Respecto al génesis del análisis de regresión, Gutiérrez (2012) sostuvo que la palabra regresión fue usada por primera vez por Francis Galton (1822-1911), quien como resultado de una investigación encontró que, aunque naturalmente existía una relación directa entre la estatura de los padres y la de sus hijos, también se daba una relación inversa, por lo cual la estatura media muestral regresaba hacia la estatura media poblacional. El autor agregó que la investigación de Galton originó la ley de regresión universal, la cual fue corroborada por Karl Pearson (1857-1936) y que es la base del análisis de regresión.

Acerca del aporte de Karl Pearson, Gujarati y Porter (2010) afirmaron que Pearson acopió más de un millar de tallas y encontró una relación inversa entre la estatura de los padres y la de sus hijos. Además, McClave y Sincich (2018) señalaron que George Yule (1871-1951) fue el primero en utilizar el método de mínimos cuadrados, luego desarrollaría la teoría de regresión múltiple y más tarde crearía el análisis de series temporales.

### ***2.1.9 Elección del Método de Pronóstico***

Existen diversos criterios para la elección del método de pronóstico apropiado. Lawrence et al. (2009) desarrollaron un árbol de decisión para la elección del método general de pronóstico (Figura 7), con base en la disponibilidad de información y en la importancia del pronóstico.

**Figura 7***Árbol de Decisión para la Elección del Método de Pronóstico*

*Nota.* Adaptado de *The Forecasting Method Decision Tree* (p. 4), por Lawrence et al., 2009, *Fundamentals of forecasting using Excel*, Industrial Press.

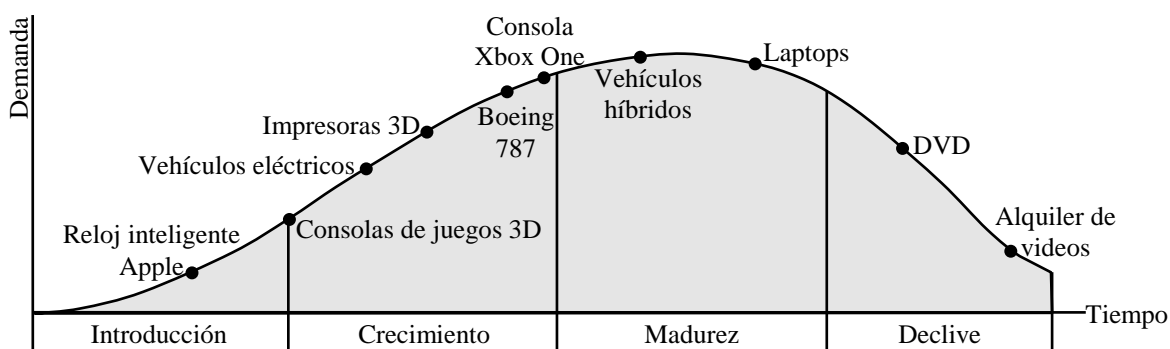
Por su lado Shim (2009), planteó la utilización de las etapas del ciclo de vida de los productos, como criterio de elección del método de pronóstico (Tabla 3). El ciclo de vida del producto se refiere a la duración temporal de su vida comercial y se compone de cuatro etapas: introducción, crecimiento, madurez y declive (Ebert et al., 2020). La gran mayoría de productos atraviesa un ciclo de vida y en cada etapa la demanda experimenta cambios (Stevenson, 2018), aunque la duración y forma del ciclo no necesariamente es la misma para todos los productos (Cooper y Schindler, 2014) y algunos productos como papel, lápices, clavos, leche, azúcar y harina no atraviesan un ciclo de vida (Reid y Sanders, 2013).

**Tabla 3***Elección del Método de Pronóstico Según el Ciclo de Vida de los Productos*

Etapa	Información	Horizonte	Método de pronóstico
Introducción	Sin información disponible	Largo	Cualitativo
Crecimiento	Poca información disponible	Largo	Investigación de mercado Regresión Series temporales
Madurez	Considerable información disponible	Corto-Largo	Regresión Series temporales Suavizamiento exponencial simple
Declinación	Abundante información disponible	Corto	Regresión Series temporales Suavizamiento exponencial simple Cualitativo

*Nota:* Adaptado de *Life Cycle Effects on Forecasting Methodology* (p. 116), por Shim, 2009, Strategic business forecasting, Global Professional Publishing.

Sobre el particular, Heizer et al. (2017) dieron ejemplos de algunos productos según su ubicación en el ciclo de vida (Figura 8).

**Figura 8***Ubicación de Algunos Productos Según el Ciclo de Vida*

*Nota.* Adaptado de *Strategy and Issues During a Product's Life* (p. 41), por Heizer et al., 2017, Operations management. Sustainability and supply chain Management (12th ed.), Pearson.

Para Hanke y Wichern (2014), el principal factor en la elección del método cuantitativo de pronóstico es el patrón presente en la serie temporal. Además, los autores afirmaron que otro factor a considerar es el horizonte temporal del pronóstico. Sobre la base de estos criterios, los investigadores construyeron una guía para elegir, entre métodos de series temporales y causales, el método de pronóstico apropiado (Tabla 4).

**Tabla 4**

*Elección del Método de Pronóstico Según el Patrón de la Serie Temporal*

Método de pronóstico	Patrón	Horizonte
Serie temporal		
Promedio móvil simple	Nivel	Corto
Suavizamiento exponencial simple	Nivel	Corto
Suavizamiento exponencial doble	Tendencia	Corto
Suavizamiento exponencial triple	Estacional	Corto
Descomposición	Estacional	Corto
ARIMA	Nivel, Tendencia, Estacional, Ciclo	Corto
Causal		
Econométrico	Ciclo	Corto
Regresión lineal simple	Tendencia	Mediano
Regresión lineal múltiple	Ciclo, Estacional	Mediano
Regresión lineal múltiple con series temporales	Tendencia, Estacional	Mediano-Largo

*Nota:* Adaptado de *Choosing a Forecasting Technique* (p. 34), por Hanke y Wichern, 2014, Business forecasting (9th ed.), Pearson.

A su turno Chase (2016), utilizó los segmentos del portafolio empresarial de marcas y/o productos para la elección del método de pronóstico (Tabla 5). Para el autor los criterios de segmentación son el valor que las marcas y/o productos representan para la empresa (Alto o bajo) y su pronosticabilidad (Alta o baja). Pronosticabilidad es un indicativo de que algo se puede pronosticar con precisión.

**Tabla 5***Elección del Método de Pronóstico Según el Portafolio de Productos*

Segmento	Valor empresarial	Pronosticabilidad	Métodos de pronóstico
Productos nuevos			Juicio estructurado
Escasa data de demanda histórica o similar a productos existentes	Alto	Baja	Consenso de panel Componente fuerza de ventas Delphi
Productos dinámicos			ARIMAX
Demanda altamente correlacionada con factores causales	Alto	Alta	Regresión lineal múltiple Regresión dinámica
Productos lentos			Modelos combinados ponderados
Demanda fragmentada en mercados y consumidores objetivo	Bajo	Baja	Suavizamiento exponencial no estacional Suavizamiento exponencial doble Promedio móvil simple
Productos estables			ARIMA
Amplia y estable data de demanda histórica, con diferentes patrones de series temporales	Bajo	Alta	Descomposición Suavizamiento exponencial triple Suavizamiento exponencial simple

*Nota:* Adaptado de *Plotting Statistical Methods Based on Segmentation and Portfolio Management Principles* (p. 165), por Chase, 2016, Next generation demand management. People, process, analytics and technology, Wiley.

Por su parte, Schroeder y Goldstein (2018) seleccionaron al método de pronóstico según el área de aplicación, el horizonte temporal y la precisión requerida (Tabla 6).

**Tabla 6***Elección del Método de Pronóstico Según Área de Aplicación y Precisión Requerida*

Área de aplicación	Horizonte	Precisión requerida	Método de pronóstico
<b>Decisiones operativas</b>			
Diseño de procesos	Largo	Media	Cualitativo o causal
Capacidad instalada de fábricas	Largo	Media	Cualitativo y causal
Planificación agregada	Mediano	Alta	Causal y series temporales
Programación de la producción	Corto	Muy alta	Series temporales
Administración de inventarios	Corto	Muy alta	Series temporales
<b>Marketing, finanzas y RRHH</b>			
Programas de marketing a largo plazo	Largo	Media	Cualitativo
Fijación de precios	Corto	Alta	Series temporales
Introducción de nuevos productos	Mediano	Media	Cualitativo y causal
Estimación de costos	Corto	Alta	Series temporales
Presupuesto de capital	Mediano	Alta	Causal y series temporales
Planificación de la fuerza laboral	Mediano	Media	Cualitativo y series temporales

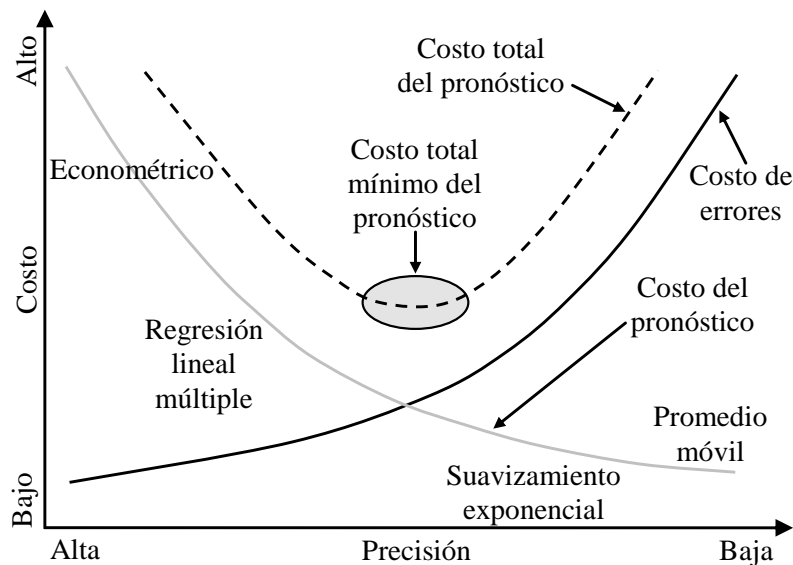
*Nota:* Adaptado de *Forecasting Uses and Methods* (p. 189), por Schroeder y Goldstein, 2018, *Operations management in the supply chain. Decisions and cases* (7th ed.), McGraw-Hill.

Finalmente, Sankaran et al. (2019) propusieron considerar al costo total del pronóstico y la precisión, como criterios para la elección de un método cuantitativo de pronóstico. Los autores afirmaron que el costo total del pronóstico es la suma del costo del pronóstico y el costo de los errores, y que el costo del pronóstico se compone de los costos de obtención de la información, de su modelación y de su monitoreo.

Los investigadores manifestaron también la existencia de un intercambio entre el costo de un método y su precisión. En un extremo, el promedio móvil tiene un costo bajo (baja complejidad) pero un alto costo de errores (baja precisión). En el otro extremo, el método econométrico tiene un costo alto (alta complejidad), pero un bajo costo de errores (alta precisión). Sin embargo, ambos extremos producen un costo total alto, por lo cual los autores propusieron elegir al método que involucre un costo total mínimo (Figura 9).

**Figura 9**

*Elección del Método de Pronóstico Según Costo Total y Precisión*



*Nota.* Adaptado de *Select an Appropriate Forecast Model* (p. 70), por Sankaran et al., 2019, *Improving forecasts with integrated business planning*, Springer.

### 2.1.10 El Análisis de Regresión Lineal

El análisis de regresión lineal es el tipo de regresión más ampliamente empleado, es la base para comprender la construcción y el funcionamiento de prácticamente todos los demás modelos de regresión (Weisberg, 2014), incluyendo a modelos muy sofisticados y, además, es esencial en una diversidad de métodos cuantitativos (Paoletta, 2019). El método, aunque requiere procesar más información que otros métodos, puede manejar los patrones tendencia, estacionalidad y ciclo; además pronostica la demanda con mayor precisión para cualquier espacio temporal de pronóstico (Chase, 2013).

El método construye un modelo matemático lineal, el cual explica y pronostica el comportamiento de una variable dependiente en función de una o más variables independientes (Darlington y Hayes, 2017). Según el número de variables independientes, se tiene el análisis de regresión lineal simple (análisis de regresión univariante) y el análisis de regresión lineal múltiple (análisis de regresión multivariante).

**2.1.10.1 El Análisis de Regresión Lineal Simple.** El análisis de regresión simple utiliza solo una variable independiente para explicar y pronosticar el valor de una variable dependiente. La ecuación que relaciona ambas variables se denomina modelo de regresión simple. El modelo de regresión poblacional simple es:

$$Y = \beta_0 + \beta_1 X_t + \mu_t$$

Donde: Y = Variable dependiente

$\beta_0$  = Intercepto

$\beta_1$  = Parámetro o coeficiente asociado a X

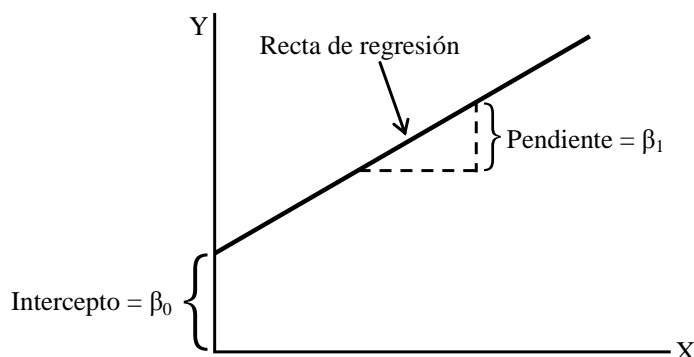
$X_t$  = Variable independiente

$\mu_t$  = Error, término de error, perturbación

El parámetro  $\beta_0$  (intercepto o constante) indica el punto por donde la recta de regresión corta al eje Y y el parámetro  $\beta_1$  es la pendiente (inclinación) de la recta de regresión (Figura 10). La letra griega  $\mu$  representa variables distintas a X que influyen en Y y el subíndice t indica que los datos son series temporales. En el modelo de regresión muestral, los residuos son las diferencias entre el valor observado de la variable dependiente Y y su pronóstico, realizado mediante el modelo de regresión muestral.

### Figura 10

*La Recta de Regresión*



Dado que, es irrealizable la construcción de un modelo de regresión poblacional simple, lo que se hace es una estimación por medio del modelo de regresión muestral simple:

$$\hat{Y} = b_0 + b_1X_t + u_t$$

**2.1.10.2 El Análisis de Regresión Lineal Múltiple.** El análisis de regresión múltiple utiliza dos o más variables independientes para explicar y pronosticar el valor de una variable dependiente. Ballou (2004), precisó que el método permite la construcción de un modelo que permite pronosticar la demanda. El modelo de regresión poblacional múltiple con k variables independientes es:

$$Y = \beta_0 + \beta_1X_{1t} + \beta_2X_{2t} + \dots + \beta_kX_{kt} + \mu_t$$

Donde: Y = Variable dependiente

$\beta_0$  = Intercepto

$\beta_1$  = Parámetro o coeficiente asociado a  $X_{1t}$

$\beta_2$  = Parámetro o coeficiente asociado a  $X_{2t}$

$\beta_k$  = Parámetro o coeficiente asociado a  $X_{kt}$

$X_{1t}$  = Variable independiente 1

$X_{2t}$  = Variable independiente 2

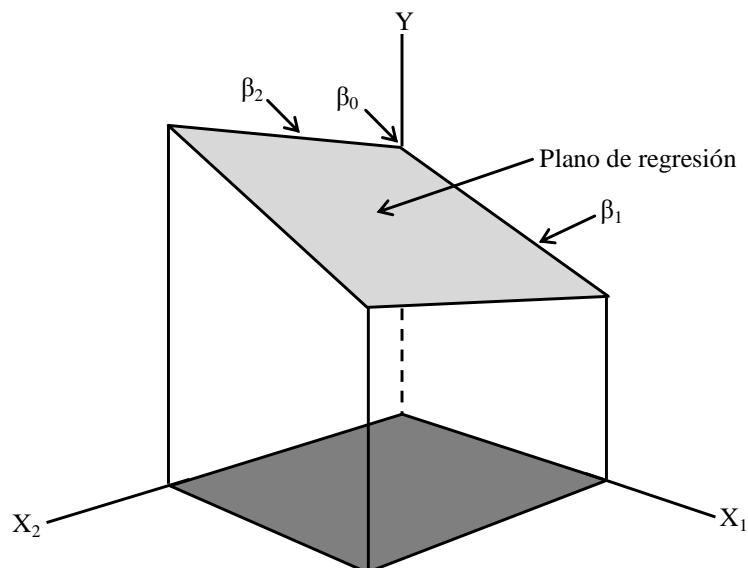
$X_{kt}$  = Variable independiente k

$\mu_t$  = Error, término de error, perturbación

En el modelo de regresión múltiple con dos variables independientes, el parámetro  $\beta_0$  indica el punto por donde el plano de regresión corta al eje Y, el parámetro  $\beta_1$  es la pendiente de  $X_1$  y el parámetro  $\beta_2$  es la pendiente de  $X_2$  (Figura 11).

## Figura 11

### El Plano de Regresión



*Nota.* Adaptado de *Multiple Regression Hyperplane for Population* (p. 600), por Groebner et al., 2018, *Business statistics: A decision-making approach* (10th ed.), Pearson.

Por la misma razón que en el caso del análisis de regresión simple, el modelo de regresión muestral múltiple con  $k$  variables independientes es:

$$\hat{Y} = b_0 + b_1X_{1t} + b_2X_{2t} + b_3X_{3t} + \dots + b_kX_{kt} + u_t$$

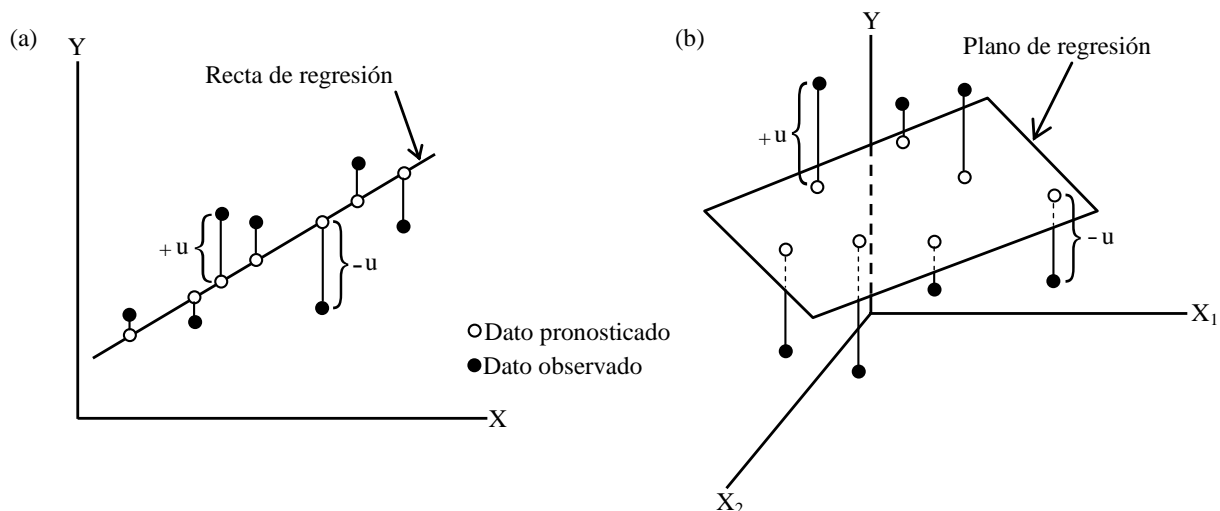
**2.1.10.3 El Principio de Parsimonia.** En muchas ocasiones los modelos de regresión múltiple son superiores a los modelos simples. La cuestión es saber el número de variables independientes que debe contener el modelo múltiple. El principio de parsimonia establece que, al disponer de varios modelos buenos, se debe optar por aquel modelo que contenga el menor número de variables independientes (Brooks, 2014; Gujarati, 2015; Hanke y Wichern, 2014; Keith, 2019; Montgomery et al., 2012; Suárez et al., 2017; Young, 2017).

**2.1.10.4 Mínimos Cuadrados Ordinarios (MCO).** El método mínimos cuadrados ordinarios (MCO), inventado por el matemático alemán Carl Friedrich Gauss, usa información muestral para hallar el modelo de regresión ( $\hat{Y}$ ).

En el caso del análisis de regresión simple, el método MCO estima los parámetros ( $b_0$  y  $b_1$ ) minimizando la sumatoria de los cuadrados de las distancias verticales (residuos) entre los datos observados de  $Y$  y los pronosticados por la recta de regresión (Figura 12a). En el caso del análisis de regresión múltiple, y en particular cuando se usa dos variables independientes, el método MCO estima los parámetros ( $b_0$ ,  $b_1$  y  $b_2$ ) minimizando la sumatoria de los cuadrados de las distancias verticales (residuos) entre los datos observados de  $Y$  y los pronosticados por el plano de regresión (Figura 12b).

### Figura 12

*Errores en (a) una Recta de Regresión y (b) un Plano de Regresión*



**2.1.10.5 Tipos de Datos.** El método MCO se puede aplicar a series temporales (series de tiempo), a datos de corte transversal (datos estáticos, series transversales, datos de sección cruzada) y a datos de panel (paneles de datos, datos longitudinales). Una serie temporal se define como un grupo de datos de una variable, recopilados con cierta periodicidad (Rosales et al., 2013). La periodicidad de una serie temporal es el intervalo de tiempo para el cual han sido recopilado los datos, los intervalos más comunes son anual, trimestral y mensual (Doane y Seward, 2016).

Por otro lado, los datos de corte transversal consisten en información de una o más variables acopiadas en un mismo momento (Gujarati, 2015).

Finalmente, los datos de panel se componen de un conjunto de unidades de datos de corte transversal (personas, familias, empresas, regiones, países, etc.), que han sido observados a lo largo del tiempo (Hill et al., 2018).

**2.1.10.6 Supuestos del Análisis de Regresión Lineal con Series Temporales.** El método MCO y sus contrastes de hipótesis se fundamentan en los supuestos clásicos del análisis de regresión. Si los supuestos se cumplen, el método MCO suministra un conjunto potente de herramientas analíticas, sin embargo, si algún supuesto es violado, los parámetros estimados pueden dejar de ser eficientes y los contrastes de hipótesis pueden no ser fiables (Newbold et al., 2013). Por lo cual, es posible que la violación de los supuestos complique el empleo del modelo para pronosticar (Hoshmand, 2010).

**2.1.10.6.1 El Modelo Es Lineal.** El supuesto estipula que los parámetros del modelo son lineales, y en un modelo de regresión simple, una línea recta grafica la relación entre ambas variables (Brooks, 2014). Los parámetros son lineales cuando solo están elevados a la unidad y no se multiplican ni se dividen entre ellos (Gujarati, 2015) y cuando cada variable independiente es multiplicada por un parámetro (Brooks, 2014; Rodríguez y González, 2017). El supuesto no se extiende a las variables, éstas pueden ser o no lineales. Es decir, las variables pueden expresarse en términos logarítmicos, cuadráticos o cúbicos, e inclusive pueden interactuar. El modelo siguiente es lineal en los parámetros y en las variables:

$$Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$$

Seguidamente se presentan algunos ejemplos de modelos tanto lineales como de modelos no lineales (Tabla 7).

**Tabla 7***Ejemplos de Modelos Lineales y No Lineales*

Modelos lineales	Modelos no lineales
$Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}^2 + \mu_t$	$Y = \beta_0 + \beta_1^2 X_{1t} + \beta_2^2 X_{2t} + \mu_t$
$Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{1t} X_{2t} + \mu_t$	$Y = \beta_0 + \text{Ln}(\beta_1 X_{1t}) + \text{Ln}(\beta_2 X_{2t}) + \mu_t$
$Y = \beta_0 + \beta_1 \text{Ln} X_{1t} + \beta_2 \text{Ln} X_{2t} + \mu_t$	$Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} \mu_t$
$1/Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	$Y = \beta_0 + \beta_1 X_{1t}^{\beta_2} + \mu_t$

El supuesto de linealidad es el más trascendental, su violación puede implicar el insesgamiento de los parámetros, de los errores estándar, del coeficiente de determinación ( $R^2$ ) y de las pruebas de significancia (Keith, 2019). Su incumplimiento también origina problemas de especificación, tales como variables independientes irrelevantes y cambio estructural (Asteriou y Hall, 2011). Además, el método MCO solo se aplica a relaciones lineales, cuando una variable independiente viola el supuesto, la variable resulta en no significativa, es decir no aporta poder explicativo al modelo (Richardson, 2015).

**2.1.10.6.2 No Existe Colinealidad o Multicolinealidad Exacta.** El problema de colinealidad o multicolinealidad está normalmente presente en las series temporales, debido a que las series frecuentemente exhiben la misma tendencia (Pesaran, 2015), por lo que en realidad no es una cuestión de presencia o de ausencia, sino de grado (Quintana y Mendoza, 2016). Es decir, la multicolinealidad es un problema que está presente, en algún grado, en casi todos los conjuntos de datos (Richardson, 2015). Sin embargo, según se vaya incrementando el tamaño muestral, los modelos de regresión lineal serán menos afectados por la multicolinealidad (Hoffmann, 2010).

Muchas variables independientes presentan cierto grado de relación entre si y por eso son incluidas en un modelo de regresión. La colinealidad es una situación en la cual dos variables independientes se relacionan linealmente; cuando la situación involucra a más de dos variables independientes se denomina multicolinealidad. Sin embargo, es usual utilizar

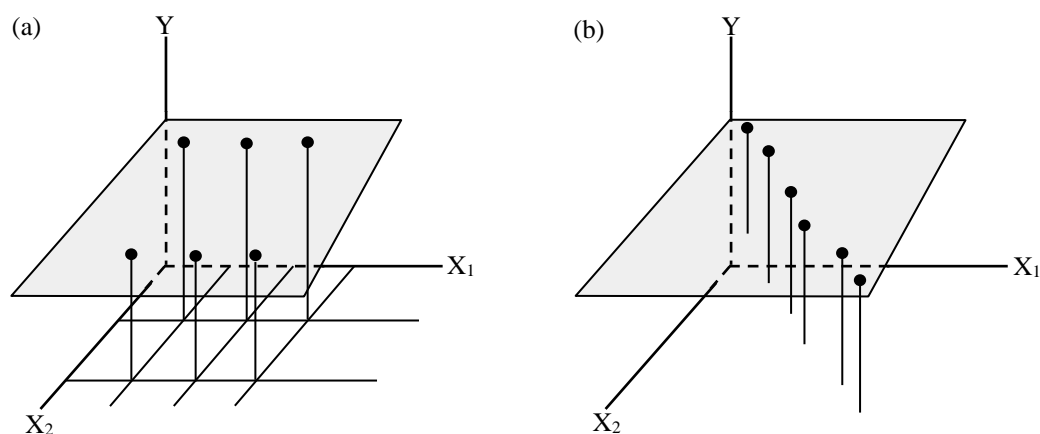
ambos términos indistintamente (Keith, 2019; Keller, 2018; Weisberg, 2014) y también es habitual emplear el vocablo multicolinealidad para referirse a cualquiera de las dos situaciones (Studenmund, 2017).

Un extremo del problema es la multicolinealidad exacta o perfecta, la cual ocurre cuando la relación lineal entre las variables independientes es exacta o perfecta, por ejemplo cuando  $X_1 = 3X_2$ . En este caso, por lo menos una variable independiente es redundante y podría omitirse, sin ninguna pérdida de información (Albright y Winston, 2017). El supuesto implica que las variables independientes no deben presentar multicolinealidad exacta o perfecta, por lo cual no se viola este supuesto cuando la relación entre las variables independientes es no lineal exacta.

Cuando el supuesto se cumple, el plano se ajusta a los puntos logrando estabilidad (Figura 13a), caso contrario, el plano es inestable y es muy sensible ante mínimos cambios en las variables (Figura 13b) (Montgomery et al., 2012). A la distribución de puntos (Y observados) en la figura 13b se le llama cerca de madera (picket fence) y representa a variables con multicolinealidad fuerte (Bowerman et al., 2015).

### Figura 13

*Variables (a) No Colineales y (b) Colineales*



*Nota.* Adaptado de (a) *A Data Set with Multicollinearity*. (b) *Orthogonal Regressors* (p. 119), por Montgomery et al., 2016, *Introduction to linear regression analysis* (5th ed.), Wiley.

Las causas de la multicolinealidad pueden ser de diversa índole. Konasani y Kadre (2015) afirmaron que la multicolinealidad puede ser originada por:

- Incluir variables independientes relacionadas. Por ejemplo, el índice de precios al consumidor (IPC) y la tasa de inflación son indicadores del nivel general de precios. La tasa de inflación es la variación del IPC.
- Tener muchas variables independientes para un tamaño muestral muy pequeño.

Según Court y Rengifo (2011), la ocurrencia de multicolinealidad exacta es poco probable y lo que ocurre con cierta frecuencia es la multicolinealidad cercana o imperfecta, y ante su presencia se tendrá:

- Coeficientes estimados que cambian fácilmente al omitir o incluir variables independientes, por tanto, no se podrá asegurar la estabilidad de los coeficientes.
- Coeficientes de determinación muy altos, conjuntamente con coeficientes estimados no significativos.
- Intervalos de confianza excesivamente amplios, por tanto, dificultad al practicar pruebas de hipótesis de los parámetros.

Además, existe una relación inversa cercana entre multicolinealidad y potencia de la prueba t de significancia individual. La multicolinealidad alta disminuye la potencia de la prueba t, originando el no rechazo de la hipótesis nula que  $\beta_1 = 0$ , aun siendo falsa (Pesaran, 2015), es decir, concluiríamos en que la variable asociada a  $\beta_1$  no aporta significativamente al modelo y debe ser excluida, cuando en realidad si lo hace y por tanto no debe omitirse. También puede originar que los parámetros estimados presenten signos diferentes a los esperados (Mendenhall y Sincich, 2012).

**2.1.10.6.3 El Promedio de los Errores Es Cero.** Este supuesto establece que las variables independientes son exógenas en sentido estricto, es decir, las variables independientes y el error son independientes (Wooldridge, 2020); y, además, no se emplean

como variables independientes a retardos de la variable dependiente (Hill et al., 2018). El supuesto se cumple cuando la sumatoria de los errores es cero. Tal como se observa en las figuras 12a y 12b, al estimar los parámetros de un modelo mediante MCO, se obtienen errores positivos y negativos, la sumatoria de dichos errores tiene que ser cero, de tal manera que su promedio también lo sea.

La violación de este supuesto puede deberse a la elección de una forma funcional incorrecta, o por la violación del supuesto de homocedasticidad o el de no autocorrelación (Caridad y Ocerin, 2012). Otra causa es la no inclusión de variables independientes relevantes (Davidson, 2018).

La violación del supuesto puede originar un intercepto sesgado (Asteriou y Hall, 2011; Kennedy, 2008). Y cuando se viola el supuesto y se omite al intercepto, el coeficiente de determinación ( $R^2$ ) puede ser negativo, es decir, la media de la variable dependiente explica más el comportamiento de la variable dependiente, que las propias variables independientes (Brooks, 2014).

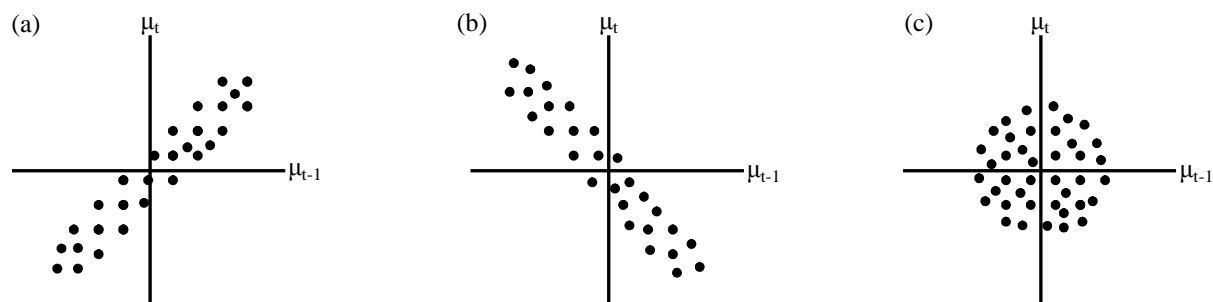
**2.1.10.6.4 Los Errores No Están Autocorrelacionados.** Al problema de autocorrelación también se le conoce como autocorrelación serial o correlación serial. Frecuentemente los errores de las series temporales presentan autocorrelación (Greene, 2018). El supuesto exige que cada error sea independiente de errores anteriores y posteriores a él, es decir, los errores no están correlacionados entre sí a lo largo del tiempo. Esto es, conocer el signo y el valor de uno o más errores no servirá para pronosticar el signo y el valor de cualquier otro error (Lee et al., 2013).

La autocorrelación se visualiza graficando el error ( $\mu_t$ ) versus su rezago ( $\mu_{t-1}$ ). Con autocorrelación positiva la mayoría de los puntos se encuentran en los cuadrantes uno y tres (Figura 14a). Con autocorrelación negativa la mayor parte de los puntos se concentran en los

cuadrantes dos y cuatro (Figura 14b). Por último, no existe autocorrelación cuando los puntos están distribuidos aleatoriamente en todos los cuadrantes (Figura 14c).

### Figura 14

*Errores (a) Con Autocorrelación Positiva, (b) Con Autocorrelación Negativa y (c) Sin Autocorrelación (Errores esféricos)*



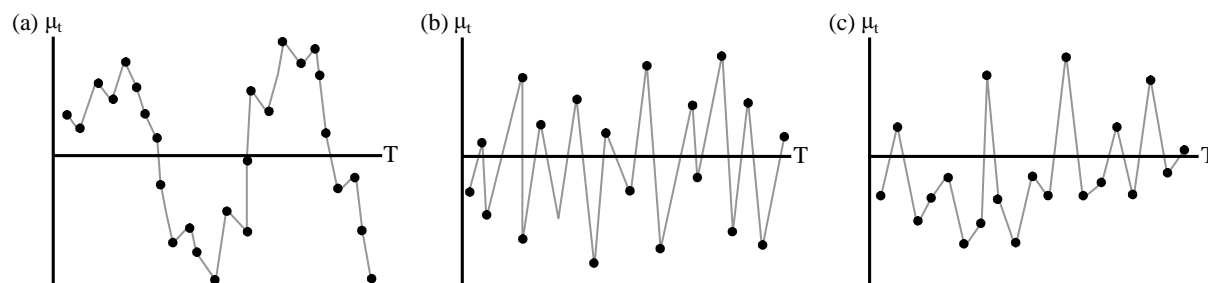
*Nota.* Adaptado de *Plot of  $u_t$  Against  $u_{t-1}$ , Showing Positive Autocorrelation* (p. 191), *Plot of  $u_t$  Against  $u_{t-1}$ , Showing Negative Autocorrelation* (p. 192), *Plot of  $u_t$  Against  $u_{t-1}$ , Showing No Autocorrelation* (p. 193), por Brooks, 2014, *Introductory econometrics for finance* (3rd ed.), Cambridge University Press.

La autocorrelación también se puede visualizar graficando el error ( $\mu_t$ ) versus el tiempo (T). En presencia de autocorrelación positiva, puntos positivos son seguidos de puntos negativos y estos a su vez por puntos positivos (Figura 15a). Cuando se alterna el signo de los puntos se configura la autocorrelación negativa (Figura 15b). Finalmente, no existe autocorrelación cuando los puntos no exhiben un patrón definido (Figura 15c).

La autocorrelación puede ser ocasionada por la presencia de estacionalidad, por una mala especificación de la forma funcional y por la no inclusión de variables independientes relevantes (Court y Rengifo, 2011). La autocorrelación también puede ser causada por un cambio estructural (Welc y Rodriguez, 2018).

**Figura 15**

*Errores (a) Con Autocorrelación Positiva, (b) Con Autocorrelación Negativa y (c) Sin Autocorrelación*



*Nota.* Adaptado de *Plot of  $u_t$  Over Time, Showing Positive Autocorrelation* (p. 191), *Plot of  $u_t$  Over Time, Showing Negative Autocorrelation* (p. 192), *Plot of  $u_t$  Over Time, Showing No Autocorrelation* (p. 193), por Brooks, 2014, *Introductory econometrics for finance* (3rd ed.), Cambridge University Press.

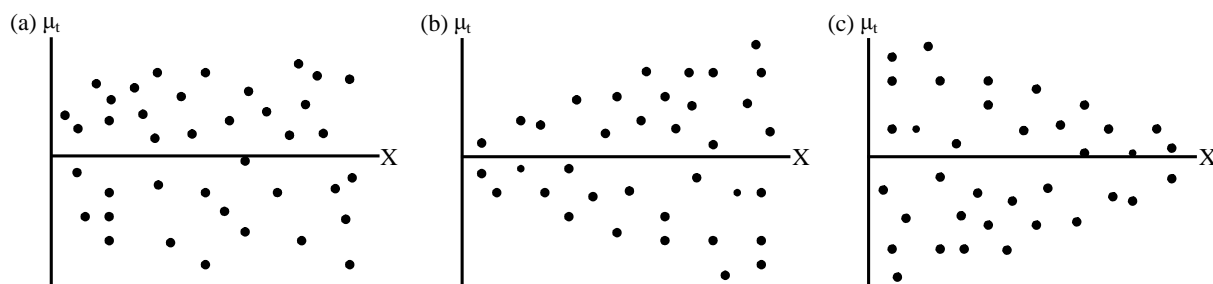
Además, la autocorrelación puede ser impura o pura. Es impura si es originada por una forma funcional errónea o por la exclusión de una variable relevante. Y es pura si se origina por la distribución subyacente del error de la verdadera especificación del modelo (Studenmund, 2017).

La presencia de autocorrelación trae como consecuencia estimadores no eficientes e invalidez de las pruebas de significancia de los parámetros (Caridad y Ocerin, 2012). Además puede causar pronósticos sesgados (Welc y Rodriguez, 2018).

**2.1.10.6.5 Los Errores Son Homocedásticos.** El supuesto implica que para cualquier valor de  $X$ , los errores tengan igual varianza (Figura 16a). Es decir, la variabilidad de los valores de  $Y$  es la misma sin importar que tan grande o que tan diminuto sea el valor de  $X$ . Si este supuesto se viola, entonces los errores son heterocedásticos (Figuras 16b y 16c). La heterocedasticidad es usual en regresiones con datos de corte transversal y es inusual en regresiones con series temporales (Pesaran, 2015; Wilson y Keating, 2009).

**Figura 16**

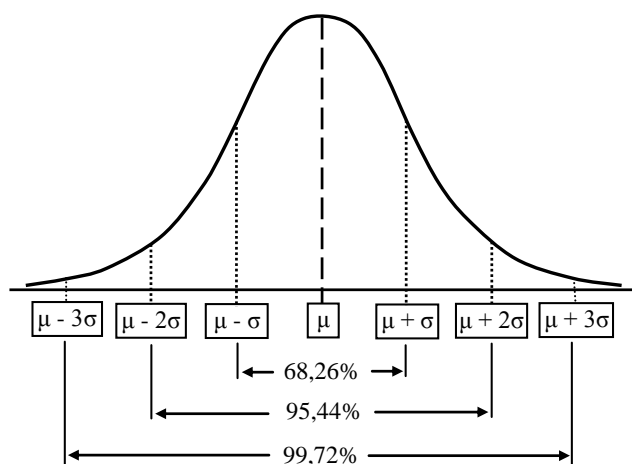
*Errores (a) Homocedásticos, (b) Heterocedásticos Crecientes y (c) Heterocedásticos Decrecientes*



La heterocedasticidad puede ser originada por la adopción de una forma funcional incorrecta, por la no inclusión de variables independientes relevantes y por cambio estructural (Welc y Rodriguez, 2018).

Ante errores heterocedásticos, los coeficientes estimados siguen siendo insesgados, pero son ineficientes; por tanto, las pruebas t y F serán imprecisas (Court y Rengifo, 2011). La heterocedasticidad también puede alterar el rango de los errores de los pronósticos y puede provocar que los errores presenten una distribución desconocida (Welc y Rodriguez, 2018). Sin embargo, sus efectos son mucho menores que los causados por la multicolinealidad o por la autocorrelación (Richardson, 2015).

**2.1.10.6.6 Los Errores Se Distribuyen Normalmente.** La distribución normal es la distribución de probabilidad continua más frecuentemente empleada en la economía y en los negocios (Newbold et al., 2013). La distribución tiene forma de campana y es simétrica con respecto a su media, sus colas se extienden indefinidamente en ambas direcciones sin llegar a tocar la horizontal y la mayor parte del área está concentrada alrededor de la media (Figura 17) (Anderson et al., 2017).

**Figura 17***La Distribución Normal*

*Nota.* Adaptado de *Areas Under the Curve for Any Normal Distribution* (p. 278), por Anderson et al., 2017, *Statistics for business and economics* (13th ed.), Cengage.

El supuesto requiere que los errores se distribuyan de forma normal con media cero y varianza constante. El cumplimiento del supuesto asegura que los coeficientes del modelo sean también normales (Pedace, 2013) y, que en promedio, los datos caigan en la línea verdadera (Baltagi, 2011). Además, el supuesto tiene gran relevancia para la inferencia estadística. Si los errores son no normales, los coeficientes no se distribuyen normalmente y los estadísticos t y F no tienen distribuciones t y F, respectivamente, lo que dificulta la obtención de los valores p (Wooldridge, 2020).

Sin embargo, si la distribución de los errores se aleja poco de la normalidad, el modelo no será gravemente afectado (Aljandali, 2017) y el efecto en el error Tipo I será muy pequeño (Mendenhall y Sincich). Además, cuando el tamaño muestral es relativamente grande, no se requiere de este supuesto para realizar inferencia estadística (Hill et al., 2018). En el extremo, la violación del supuesto prácticamente no acarrea consecuencias, siempre que el tamaño muestral sea lo suficientemente grande (Brooks, 2014; Caridad y Ocerin, 2012). Por lo cual, este supuesto es el menos importante (Darlington y Hayes, 2017; Matloff, 2017).

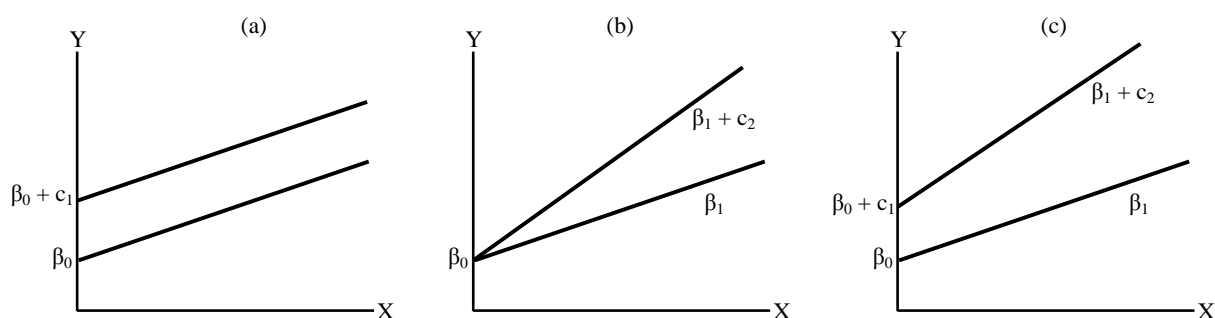
La no normalidad puede deberse a los valores atípicos (Court y Rengifo, 2011; Marcellino, 2016), los cuales son datos extremadamente grandes o extremadamente pequeños (McClave y Sincich, 2018) y se ubican a más-menos tres desviaciones estándar de la media (Figura 17, p. 58) (Groebner et al., 2018). Otras causas de la no normalidad son las muestras pequeñas, una transformación inadecuada (Quintana y Mendoza, 2016) y la estacionalidad (Court y Rengifo, 2011).

**2.1.10.6.7 Ausencia de Cambio Estructural.** Este supuesto también es conocido como ausencia de quiebre estructural o de estabilidad paramétrica. Un cambio estructural significa que los valores de los parámetros no son estables en el tiempo, es decir la relación entre la variable dependiente y las independientes cambia (Gujarati y Porter, 2010). No se pretende que los parámetros sean idénticos al volverse a estimar luego de algún tiempo, lo que se desea es que no sean diferentes en sentido estadístico (Court y Rengifo, 2011). Un cambio estructural es una forma de no linealidad que puede ser grande o pequeño y se puede dar de manera progresiva o brusca (Diebold, 2017), además puede ser eventual o permanente (Marcellino, 2016).

Los cambios estructurales pueden apreciarse en la variación del término constante, por el cambio en la pendiente o por los dos (Gujarati y Porter, 2010). Cuando el cambio estructural es originado por la variación de la constante ( $\beta_0$ ) aparece las regresiones paralelas (Figura 18a), cuando el cambio es por variación en la pendiente ( $\beta_1$ ) aparece las regresiones concurrentes (Figura 18b) y cuando el cambio es en ambos se tiene regresiones inconexas (Figura 18c) (Rosales et al., 2013).

**Figura 18**

*Regresiones (a) Paralelas, (b) Concurrentes y (c) Inconexas*



Los cambios estructurales pueden deberse a una data muy extensa (Karlsson, 2013), por cambios institucionales, variaciones en la política económica, productos nuevos, nuevos procesos de manufactura, globalización, conflictos bélicos, terrorismo, derrumbe del mercado financiero y valores atípicos (Ghysels y Marcellino, 2018).

Tomando en consideración la localización y la magnitud del cambio estructural, éste podría originar malos pronósticos (Stock y Watson, 2020). Es decir, un cambio estructural produce pronósticos sesgados, además el error del pronóstico puede violar los supuestos de homocedasticidad y no autocorrelación (Marcellino, 2016).

**2.1.10.6.8 Especificación Correcta.** Especificar correctamente un modelo implica la incorporación de variables independientes relevantes y la elección de una forma funcional correcta. Los errores de especificación se pueden dar por elegir una forma funcional incorrecta, por prescindir de una variable relevante, por incluir una variable irrelevante o por errores de medición (Gujarati y Porter, 2010). Otra causa de la mala especificación es la violación de los supuestos del modelo regresión múltiple (Hill et al., 2018). El promedio de los errores no es cero, cuando la especificación incorrecta es ocasionada por omisión de variables o por una inadecuada forma funcional (Court y Rengifo, 2011).

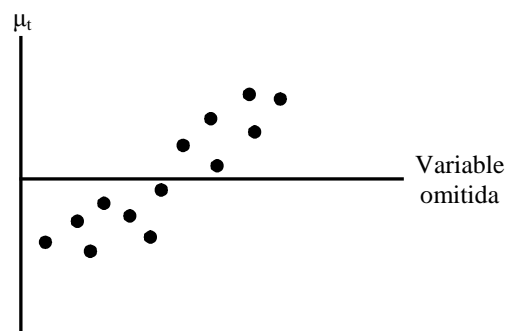
La forma funcional no está correctamente especificada, cuando la relación entre las variables del modelo, no está representada de forma apropiada (Wooldridge, 2020). La forma

funcional incorrecta se produce por inestabilidad paramétrica (Quintana y Mendoza, 2016). Una forma funcional no correcta puede ocasionar la violación de los supuestos de errores no autocorrelacionados y homocedásticos, además de originar pronósticos sesgados (Welc y Rodríguez, 2018). De igual manera, los parámetros estimados mediante MCO carecerán de insesgamiento y consistencia y no será posible realizar ninguna prueba de hipótesis confiable (Court y Rengifo, 2011).

En algunas circunstancias, la no inclusión de variables relevantes puede generar la violación de los supuestos de no autocorrelación, homocedasticidad y normalidad; originando desconfianza en las pruebas t y F (Marcellino, 2016). En otras circunstancias, origina especificación errónea o multicolinealidad, en este caso los parámetros estimados mediante MCO carecen de insesgamiento y eficiencia, y cualquier prueba de hipótesis estará errada, debido al agrandamiento de las desviaciones estándar (Court y Rengifo, 2011). Una variable es relevante cuando la media de la variable dependiente es influenciada por dicha variable (Adkins, 2018). El error de especificación por variable omitida se puede apreciar al graficar los errores versus la variable omitida (Figura 19).

### Figura 19

#### *Patrón de Errores por una Variable Omitida*



*Nota.* Adaptado de *Plots that Indicate Abnormality in Data: (f) Variable Omitted* (p. 546), por Devore, 2016, *Probability and statistics for engineering and the sciences* (9th ed.), Cengage.

Las repercusiones de la no inclusión de variables relevantes, podría significar la introducción desmedida de variables, complicando innecesariamente el modelo y agrandando las varianzas de los parámetros estimados, debido a la inclusión de variables irrelevantes (Hill et al., 2018). Si en el modelo se incluyen variables irrelevantes, es casi seguro que dichas variables sean no significativas y tengan que ser removidas del modelo; caso contrario el modelo será inadecuado debido a que MCO generará estimadores insesgados, pero su varianza no será la menor posible (Caridad y Ocerin, 2012).

#### **2.1.10.7 Propiedades de los Parámetros Estimados Mediante MCO**

**Insesgamiento.** Los parámetros estimados mediante MCO son insesgados. Cuando el valor esperado de un parámetro estimado es igual al valor poblacional, entonces dicho parámetro estimado es insesgado (Court y Rengifo, 2011). Los parámetros del modelo de regresión son insesgados cuando el modelo es lineal, cuando no existe colinealidad o multicolinealidad exacta y cuando el promedio de los errores es cero.

**Eficiencia.** Los parámetros estimados mediante MCO son eficientes. Un parámetro estimado es eficiente cuando presenta la menor varianza entre todos los posibles parámetros estimados (Court y Rengifo, 2011). Los parámetros estimados son eficientes cuando son insesgados y además los errores no están autocorrelacionados y son homocedásticos.

**Consistencia.** Los parámetros estimados mediante MCO son consistentes. Si un parámetro estimado es eficiente entonces también es consistente, debido a que según se incrementa el tamaño muestral, el valor del parámetro cada vez se acerca más al valor poblacional y su varianza tiende a cero (Court y Rengifo, 2011).

Según el teorema de Gauss-Markov, si los parámetros estimados mediante MCO son insesgados, eficientes y consistentes, entonces son los Mejores Estimadores Lineales Insesgados (MELI) (Wooldridge, 2020). Es decir, los coeficientes estimados mediante MCO serán mejores que los obtenidos por cualquier otro método (Wooldridge, 2020).

**2.1.10.8 Errores Tipos I y II, Nivel de Significancia y Potencia.** Las pruebas de hipótesis siempre se realizan para rechazar o no rechazar una hipótesis nula ( $H_0$ ). Al realizar una prueba de hipótesis se pueden cometer dos tipos de errores. Por un lado, el error Tipo I ocurre cuando se rechaza una  $H_0$  verdadera y su probabilidad de ocurrencia es el nivel de significancia de la prueba de hipótesis ( $\alpha$ ). De otro lado, el error Tipo II sucede cuando no se rechaza una  $H_0$  falsa y su probabilidad de ocurrencia es  $\beta$ . La potencia de una prueba de hipótesis ( $1 - \beta$ ) es la probabilidad de rechazar acertadamente una  $H_0$  falsa.

Es muy común fijar el valor de  $\alpha$  en 0,01, 0,05 o 0,10 (Hill et al., 2018; Paoletta, 2019), siendo el  $\alpha$  de 0,05 el más ampliamente utilizado (Agresti et al., 2018; Darlington y Hayes, 2017; Triola, 2018; Wooldridge, 2020). Un  $\alpha$  de 0,05 quiere decir que  $H_0$  puede ser rechazada con un nivel de confianza de 95%. Tomar en cuenta que elegir un  $\alpha$  de 0,01 es muy poco común, debido a que  $\beta$  podría ser demasiado grande; es decir la potencia de la prueba ( $1 - \beta$ ) podría ser muy pequeña (Bowerman et al., 2015).

**2.1.10.9 El Valor p.** Existen tres enfoques o criterios para decidir rechazar o no rechazar  $H_0$ : El método tradicional, los intervalos de confianza y el método del valor p. Sin embargo, a diferencia de los otros dos métodos, el valor p permite rechazar o no rechazar  $H_0$  sin requerir de valores críticos para diferentes  $\alpha$  (McEvoy, 2018), esto lo convierte en un método rápido para llegar a una conclusión con respecto a  $H_0$  (Andren, 2007). Además, los valores p son calculados por los programas estadísticos, por lo cual es utilizado por la mayoría de analistas e investigadores (McClave y Sincich, 2018).

El valor p, llamado también probabilidad de significación, es el valor de probabilidad exacto de incurrir en un error Tipo I, es decir, es el  $\alpha$  más bajo con el cual es posible rechazar  $H_0$ . El valor p nos dice que tan fuerte o débil es la evidencia empírica en contra de  $H_0$ , así un valor p bajo respalda el rechazo de  $H_0$ , mientras que un valor p elevado abona al no rechazo

de  $H_0$  (Wooldridge, 2020). Sobre el particular, Lind et al. (2018) interpretaron el peso de la evidencia en contra  $H_0$  según el valor p:

- Si el valor  $p < 0,10$ , entonces existe evidencia débil contra  $H_0$ .
- Si el valor  $p < 0,05$ , entonces existe evidencia fuerte contra  $H_0$ .
- Si el valor  $p < 0,01$ , entonces existe evidencia muy fuerte contra  $H_0$ .
- Si el valor  $p < 0,001$ , entonces existe evidencia extremadamente fuerte contra  $H_0$ .

Si el valor p es menor e igual a  $\alpha$ , entonces se rechaza  $H_0$  y no se rechaza  $H_0$  si el valor p excede a  $\alpha$  (Triola, 2018; Vu, 2015).

#### 2.1.10.10 Detección de la Violación de los Supuestos

**2.1.10.10.1 Linealidad: El Multiplicador de Lagrange.** La verificación del supuesto de linealidad se puede realizar con el Multiplicador de Lagrange (LM) (Box et al., 2016). La prueba LM, también conocida como prueba score o prueba de Rao, no necesita la estimación del modelo no lineal, característica que lo hace muy ventajoso (Amendola y Francq, 2009; Enders, 2015).

El procedimiento de la prueba LM para la verificación el supuesto de linealidad, para la inclusión de términos cuadráticos (logarítmicos) es el siguiente:

1. Correr el modelo y obtener sus errores.
2. Construir un modelo auxiliar con el error estimado en el paso 1 como variable dependiente, las variables independientes son las originales y sus cuadrados (logaritmos). Para regresiones con dos variables independientes el modelo auxiliar es, según corresponda:

$$\mu = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1^2 + \alpha_4 X_2^2 + v$$

o

$$\mu = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 \text{Ln}X_1 + \alpha_4 \text{Ln}X_2 + v$$

Y para regresiones con tres variables independientes el modelo auxiliar es:

$$\mu = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + \alpha_6 X_3^2 + v$$

o

$$\mu = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 \text{Ln}X_1 + \alpha_5 \text{Ln}X_2 + \alpha_6 \text{Ln}X_3 + v$$

3. Plantear las hipótesis siguientes:

H<sub>0</sub>: La relación es lineal

H<sub>1</sub>: La relación no es lineal

4. De acuerdo a H<sub>0</sub> de relación lineal, el producto del tamaño muestral (n) y el coeficiente de determinación (R<sup>2</sup>) de la regresión auxiliar, se distribuye como chi cuadrado (χ<sup>2</sup>) con grados de libertad (gl) igual al número de variables independientes del modelo auxiliar.

5. Si el valor p de la prueba LM es mayor al 5% (0,05), no se rechaza H<sub>0</sub>, por tanto la relación es lineal. Caso contrario se rechaza H<sub>0</sub> y se colige que la relación no es lineal.

**2.1.10.10.2 Colinealidad: El Factor Inflacionario de Varianza (FIV).** La colinealidad no es una cuestión poblacional sino muestral, por tanto no se requiere realizar pruebas de hipótesis, pero si se puede establecer su fuerza (Gujarati y Porter, 2010). El FIV (VIF, Variance inflation factor) mide la gravedad de la colinealidad en términos de cuantas veces el problema aumenta la varianza de un parámetro estimado. Cada variable independiente tiene un FIV (Studenmund, 2017). Así, un FIV de 5,23 indica que la varianza es 5,23 veces más grande que la obtenida sin colinealidad. La fórmula del FIV es:

$$\text{FIV}_j = \frac{1}{1 - R_j^2}$$

En la fórmula, R<sub>j</sub><sup>2</sup> es el coeficiente de determinación entre la j-ésima variable independiente y las demás variables independientes. Cuando no existe relación entre la j-ésima variable independiente y las demás, R<sub>j</sub><sup>2</sup> es cero y FIV es uno. Cuando hay relación, el FIV > 1, sin embargo, cuando el FIV es cercano a uno la colinealidad no es un problema (Hanke y Wichern, 2014). Cuando el FIV es uno, las variables son ortogonales (Montgomery et al., 2012), es decir, son no correlacionadas o independientes. Y, cuando se presenta una

relación lineal exacta,  $R_j^2$  es uno y FIV es infinito, y no existe un estimado único de los parámetros del modelo (Freund et al., 2010).

Acerca del valor del FIV, para considerar la colinealidad un problema, no existe consenso. Para algunos con un  $FIV \geq 5$  la colinealidad es moderadamente fuerte, demasiado extrema, severa o estadísticamente significativa. Para otros es severa cuando el FIV es 100. Aunque, para la mayoría un  $FIV > 10$  ya es un problema (Tabla 8).

**Tabla 8**

*Valores del FIV y su Interpretación Según Autor*

FIV	La colinealidad es	Autor
>1, <5	Moderada	Young (2017, p. 111)
2	Leve	Doane y Seward (2016, p. 574)
$\geq 5$	Moderadamente fuerte	Bowerman et al. (2015, p. 162)
	Motivo de corrección	Chase (2013, p. 186)
	Demasiado extrema	Groebner et al. (2018, p. 610)
	Severa	Studenmund (2017, p. 234)
	Estadísticamente significativa	Welc y Rodriguez (2018, p. 124)
	Un problema	Young (2017, p. 111)
> 5	Fuerte	Konasani y Kadre (2015, p. 390)
6 - 7	Excesiva	Keith (2019, p. 215)
> 10	Un problema	Adkins (2014, p. 133), Asteriou y Hall (2011, p. 100)
	Un problema	Chatterjee y Hadi (2012, p. 250), Wooldridge (2020, p. 92)
	Un problema potencial	Yan y Su (2009, p. 86)
	Un problema serio	Montgomery et al. (2012, p. 118)
	Preocupante	Aljandali (2017, p. 5), Matloff (2017, p. 315)
	Preocupante	Tattar et al. (2016, p. 442)
	Potencialmente dañina	Best y Wolf (2015, p. 107)
	Seria	Albright y Winston (2017, p. 500)
	Fuerte	Doane y Seward (2016, p. 574)
	Severa	Bowerman et al. (2015, p. 162)
	Severa	Mendenhall y Sincich (2012, p. 364), Pesaran (2015, p. 70)
100	Severa	Doane y Seward (2016, p. 574)

**2.1.10.10.3 El Promedio de los Errores Es Cero.** Este supuesto es intrínseco al método MCO, por lo que no requiere ser verificado (Rodríguez y González, 2017). Es decir, el proceso MCO origina que la sumatoria de los errores, y su promedio, sea cero (Young,

2017). Además, a medida que el tamaño muestral crece indefinidamente, el promedio se acerca a cero (Studenmund, 2017).

Sin embargo, este supuesto se dará por cumplido siempre que en el modelo se incluya al intercepto, por lo que se sugiere incluirlo y probar su significancia (Court y Rengifo, 2011). El intercepto es la parte fija de la variable dependiente que no es explicada por las variables independientes, y el término de error es la parte aleatoria de la variable dependiente que no es explicada por las variables independientes (Studenmund, 2017).

Suponga que en un modelo de regresión múltiple, el promedio de los errores no es cero sino cierta cantidad  $Z$ , entonces el promedio esperado es  $\mu_t - Z = 0$ . Ahora si en un modelo de regresión múltiple,  $Z$  es sumada al intercepto y restada al error se tiene:

$$Y = (\beta_0 + Z) + \beta_1 X_{1t} + \beta_2 X_{2t} + (\mu_t - Z)$$

Finalmente, la ecuación es re-expresada para que sea cero el promedio del error  $\mu_t^*$ , de tal manera que los parámetros  $\beta_1$  y  $\beta_2$  no son alterados:

$$Y = \beta_0^* + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t^*$$

**2.1.10.10.4 Autocorrelación: La Prueba d de Durbin-Watson (DW).** No es infrecuente que los modelos que involucren series temporales macroeconómicas presenten errores con autocorrelación de primer orden (Verbeek, 2017). Aunque es posible la existencia de autocorrelación de órdenes superiores, de darse el caso, es muy probable que igualmente se presente la de primer orden (Caridad y Ocerin, 2012). La autocorrelación de primer orden implica que un error está relacionado con el error previo:

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t$$

La expresión matemática es conocida como proceso Markov de primer orden o también como modelo autorregresivo de primer orden o simplemente AR(1). En la ecuación  $\rho$  (Rho) es el coeficiente de autocorrelación,  $\mu_{t-1}$  es el error rezagado un periodo y  $\varepsilon_t$  representa al término de error normalmente distribuido con media cero y varianza constante.

La fuerza de la autocorrelación se mide por medio de  $\rho$ , mientras más se acerque a uno en valor absoluto, más grande es la fuerza de la autocorrelación (Studenmund, 2017). En presencia de autocorrelación fuerte dos variables pueden aparentar estar relacionadas, cuando en realidad no lo están (Hanke y Wichern, 2014).

La prueba d de Durbin-Watson (DW) es la más conocida para detectar la existencia de autocorrelación de primer orden (Quintana y Mendoza, 2016). El estadístico d de DW está dado por la expresión matemática siguiente:

$$d = \frac{\sum_{t=1}^n (\mu_t - \mu_{t-1})^2}{\sum_{t=1}^n \mu_t^2}$$

Dado que  $\rho$  y el d de DW se relacionan de manera aproximada mediante la expresión  $d \approx 2(1 - \rho)$ , y debido a que  $\rho$  es una correlación y por ello solo puede tomar valores de  $\pm 1$ ; es que el valor del d de DW se ubica entre cero y cuatro. Cuando no existe autocorrelación el valor de  $\rho$  es cero, por lo que la prueba d de DW plantea las hipótesis siguientes:

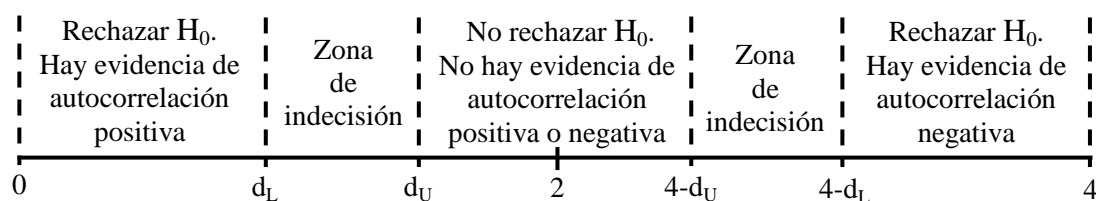
$H_0$ : Los errores no están autocorrelacionados o  $\rho = 0$

$H_1$ : Los errores están autocorrelacionados o  $\rho \neq 0$

Como el valor de d siempre estará entre cero y cuatro, se configura cinco zonas: dos de rechazo de  $H_0$ , dos de indecisión y una de no rechazo de  $H_0$ ; los límites de las zonas están dados por el d inferior ( $d_L$ ) y por el d superior ( $d_U$ ) (Figura 20). Si d es cuatro ( $\rho = -1$ ), o cercano, se rechaza  $H_0$ , por tanto los errores están autocorrelacionados negativamente. Si d es cero ( $\rho = 1$ ), o próximo, también se rechaza  $H_0$ , los errores están autocorrelacionados positivamente. Si d es dos ( $\rho = 0$ ), o cercano, no se rechaza  $H_0$ , los errores no están autocorrelacionados. Si el d cae en una zona de indecisión, lo aconsejable es aceptar la presencia de autocorrelación y corregir el problema (Caridad y Ocerin, 2012).

**Figura 20**

*Zonas de No Rechazo y Rechazo de  $H_0$  para Autocorrelación*



Ahora bien, para que ésta prueba sea válida es necesaria la presencia del intercepto ( $\beta_0$ ), que las variables independientes sean estocásticas, que rezagos de la variable dependiente ( $Y_{t-1}$ ,  $Y_{t-2}$ , etc.) no sean variables independientes y que el error se distribuya normalmente (Gujarati, 2015).

**2.1.10.10.5 Heterocedasticidad: La Prueba de White.** La prueba de heterocedasticidad de White verifica si los errores del modelo adoptan cierto patrón pre-establecido, según los cambios de las variables independientes, el cuadrado de las variables independientes y los productos cruzados de las variables independientes (Court y Rengifo, 2011). Para la verificación del supuesto de homocedasticidad se plantea:

$H_0$ : Los errores son homocedásticos

$H_1$ : Los errores son heterocedásticos

Gujarati y Porter (2010), precisaron que el procedimiento de la prueba de heterocedasticidad de White consta de los pasos siguientes:

1. Construir el modelo y obtener sus errores.
2. Construir un modelo auxiliar con el cuadrado del error estimado en el paso 1 como variable dependiente, las variables independientes son las originales, sus cuadrados y sus productos cruzados. Para regresiones con dos variables independientes el modelo es:

$$\mu^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1^2 + \alpha_4 X_2^2 + \alpha_5 X_1 X_2 + v$$

Y para regresiones con tres variables independientes el modelo es:

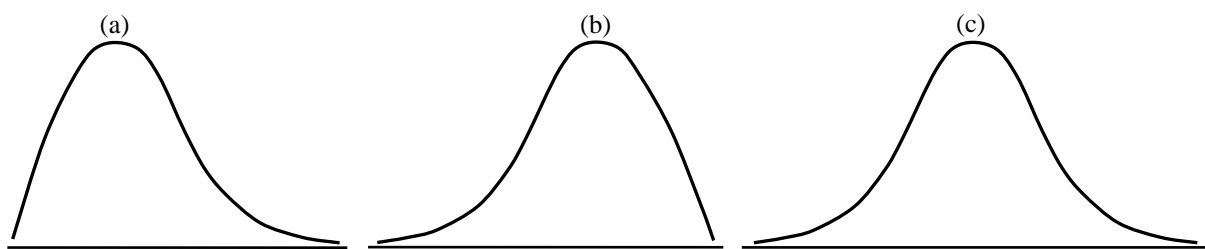
$$\mu^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + \alpha_6 X_3^2 + \alpha_7 X_1 X_2 + \alpha_8 X_1 X_3 + \alpha_9 X_2 X_3 + v$$

3. De acuerdo a  $H_0$  de errores homocedásticos, el producto del tamaño muestral ( $n$ ) y el coeficiente de determinación ( $R^2$ ) del modelo auxiliar, se distribuye como chi cuadrado ( $\chi^2$ ) con grados de libertad ( $gl$ ) igual al número de variables independientes (sin contar la constante) del modelo auxiliar.
4. Si el valor  $p$  de la prueba de heterocedasticidad de White es mayor al 5% (0,05), entonces no se rechaza  $H_0$  y en consecuencia los errores son homocedásticos. Caso contrario se rechaza  $H_0$  y se concluye que los errores son heterocedásticos.

**2.1.10.10.6 Normalidad: La Prueba de Doornik-Hansen (DH).** La prueba de normalidad de Doornik-Hansen (DH) se fundamenta en transformaciones de asimetría y curtosis, y puede ser aplicada a los errores de una regresión múltiple (Rokunuzzaman, 2013). Asimetría y curtosis son medidas que establecen la forma de una distribución. Asimetría es la diferencia en la longitud de las colas (Figuras 21a y 21b); si las colas son iguales, entonces la distribución es simétrica (Figura 21c).

### Figura 21

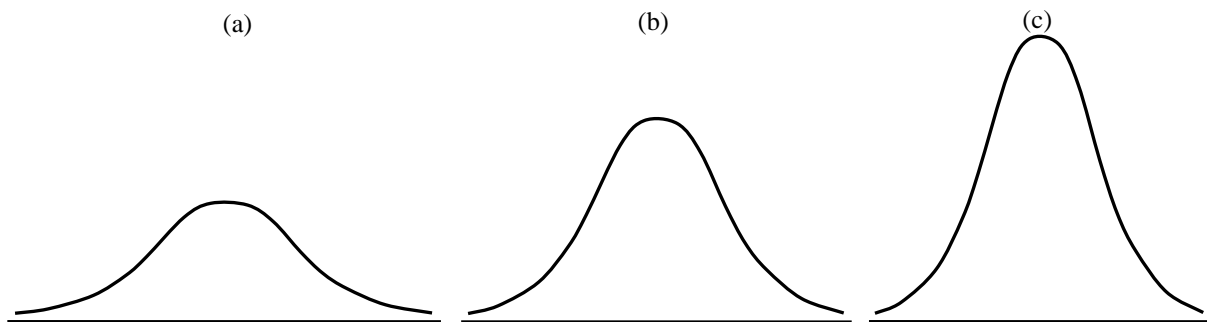
*Distribución (a) Asimétrica Derecha o Positiva, (b) Asimétrica Izquierda o Negativa y (c) Simétrica*



La curtosis clasifica la altura de una distribución en platicúrtica (Figura 22a), mesocúrtica (Figura 22b) y leptocúrtica (Figura 22c). Cuando los errores son normales, su asimetría es cero (son simétricos con respecto a su media) y su curtosis es tres (mesocúrtica).

**Figura 22**

*Distribución (a) Platicúrtica, (b) Mesocúrtica y (c) Leptocúrtica*



Para la verificación del supuesto de normalidad se plantean las hipótesis siguientes:

$H_0$ : La distribución de los errores es normal

$H_1$ : La distribución de los errores no es normal

Si el valor  $p$  de la prueba de normalidad de DH es mayor al 5% (0,05), no se rechaza  $H_0$  y entonces los errores se distribuyen normalmente. De no ser así, se rechaza  $H_0$  y se deduce que los errores no se distribuyen normalmente. Otros test de normalidad son Shapiro-Wilk, Lilliefors y Jarque-Bera. Sin embargo, las pruebas de DH y Shapiro-Wilk son mejores cuando el tamaño muestral es pequeño (Cottrell y Lucchetti, 2022a).

**2.1.10.10.7 Cambio Estructural: Las Gráficas CUSUM y CUSUMQ.** Las gráficas CUSUM (Cumulative Sum of Residuals) y CUSUMQ (Cumulative Sum of Square Residuals) verifican el supuesto de ausencia de cambio estructural. Las gráficas se construyen con los residuales recursivos (Brooks y Tsolacos, 2010). Estos residuales se obtienen al correr una regresión en una sub-muestra, luego se agrega un dato y se vuelve a correr la regresión, hasta que se han adicionado todos los datos (Brooks, 2014). Además, las gráficas revelan más información que una prueba y no requieren del momento del cambio estructural (Baltagi, 2011). Asimismo, detectan cualquier tipo de inestabilidad, no solo la paramétrica (Heij et al., 2004). También, la gráfica CUSUM puede ser útil para descubrir no linealidades (Enders, 2015). Aunque, la gráfica CUSUMQ es superior a la CUSUM, pues el carácter

cuadrático de los errores, anula la distorsión procedente de los signos. Por otro lado, la gráfica CUSUMQ puede descubrir alteraciones en la varianza (Dhrymes, 2017; Enders, 2015; Heij et al., 2004). Para la verificación del supuesto se plantea:

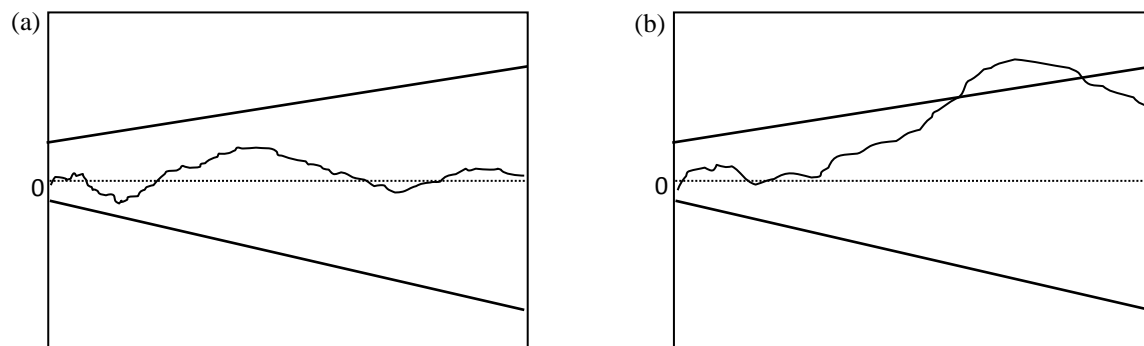
$H_0$ : No existe cambio estructural

$H_1$ : Existe cambio estructural

Bajo  $H_0$ , el valor de CUSUM es cero y el valor de CUSUMQ empieza en cero y finaliza en uno (Brooks, 2014). Si las gráficas no exceden las bandas de confianza, entonces no existe quiebre estructural (Figuras 23a y 24a), caso contrario si existe cambio estructural (Figuras 23b y 24b). Las bandas de confianza se construyen a más-menos dos desviaciones estándar, normalmente en torno a cero (Brooks y Tsolacos, 2010).

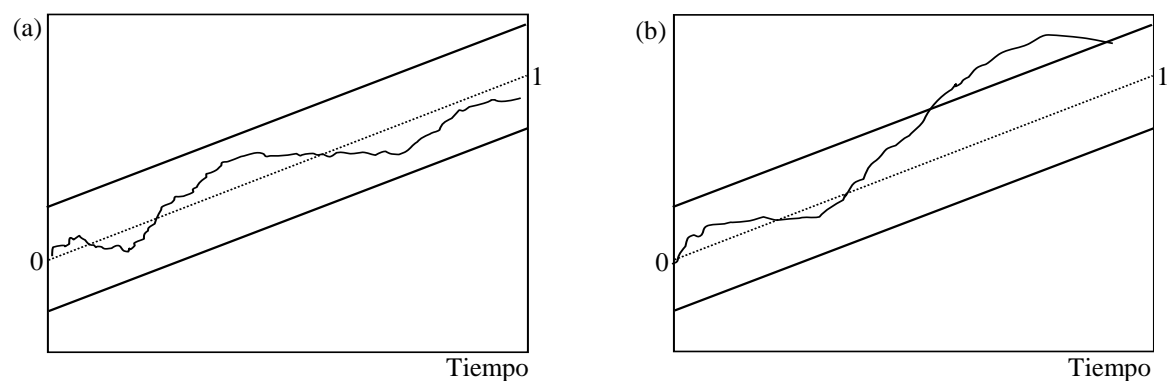
### Figura 23

Gráfica CUSUM: (a) Estabilidad Estructural y (b) Cambio Estructural



### Figura 24

Gráfica CUSUMQ: (a) Estabilidad Estructural y (b) Cambio Estructural



**2.1.10.10.8 Especificación: La Prueba RESET de Ramsey (RR).** La detección de los errores de especificación se hace con la finalidad de cambiar el modelo de regresión hasta lograr uno correctamente especificado (Caridad y Ocerin, 2012). La prueba RESET (Regression Specification Error Test) de Ramsey (RR) detecta errores de especificación debido a variables relevantes omitidas y a una forma funcional incorrecta (Hill et al., 2018). Para la verificación del supuesto de especificación correcta se plantea:

H<sub>0</sub>: El modelo está correctamente especificado

H<sub>1</sub>: El modelo no está correctamente especificado

Gujarati y Porter (2010), precisaron que el procedimiento de la prueba RR es:

1. Construir el modelo, estimar los valores de la variable dependiente ( $\hat{Y}$ ) y obtener el coeficiente de determinación del modelo original ( $R_o^2$ ).
2. En el modelo del paso 1, introducir  $\hat{Y}$  como variable independiente en forma de cuadrados ( $\hat{Y}^2$ ) y cubos ( $\hat{Y}^3$ ), aplicar el análisis de regresión y obtener el coeficiente de determinación del modelo aumentado ( $R_a^2$ ).
3. Calcular estadístico de contraste F:

$$F = \frac{(R_o^2 - R_a^2) / \text{Número de variables independientes nuevas}}{(1 - R_o^2) / (n - \text{Número de parámetros del modelo aumentado})}$$

4. Si el valor p de la prueba RR es mayor al 5% (0,05), entonces no se rechaza H<sub>0</sub>, el modelo está correctamente especificado. Caso contrario se rechaza H<sub>0</sub> y, por ende, el modelo no está correctamente especificado.

### **2.1.10.11 Violación de los Supuestos: Medidas Correctivas**

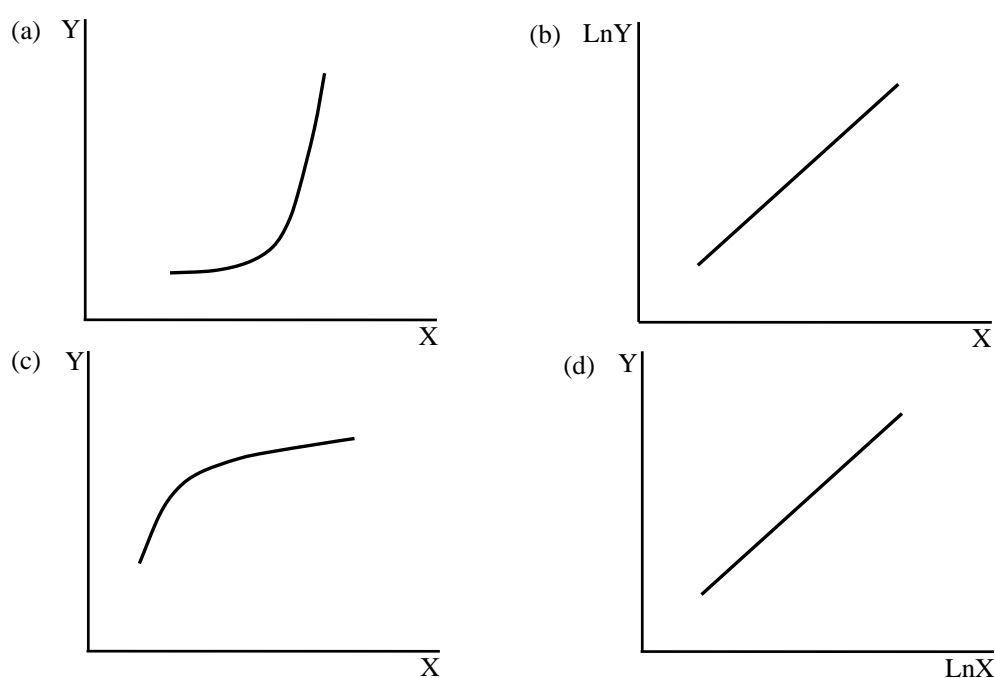
**2.1.10.11.1 El Modelo No Es Lineal.** Muchos modelos que aparentan no ser lineales pueden linealizarse aplicando logaritmos u otra transformación (Brooks, 2014). Las transformaciones también se realizan para corregir las violaciones a los supuestos del análisis de regresión (Levine et al., 2017).

La transformación no debe provocar la violación de ninguno de los supuestos relacionados con el error (Marcellino, 2016). Cuando hay heterocedasticidad, la linealidad se consigue transformando la variable dependiente ( $\text{Ln}Y$ ,  $\sqrt{Y}$  o  $Y^{-1}$ ) y cuando hay homocedasticidad, la linealidad se alcanza transformando las variables independientes ( $X^{-1}$ ,  $X^2$  o  $X^3$ ) (Young, 2017).

En cuanto a los logaritmos, los de base 10 y los naturales ( $\text{Ln}$ ) son los más utilizados (Darlington y Hayes, 2017). En el contexto del análisis de regresión simple, es fácil mostrar como modelos no lineales (Figuras 25a y 25c) se pueden linealizar con el  $\text{Ln}$  de la variable dependiente (Figura 25b) o con el  $\text{Ln}$  de la variable independiente (Figura 25d).

### Figura 25

*Modelos No Lineales (a y c) Linealizados por Logaritmos (b y d)*



*Nota.* Adaptado de *Log-linear:  $\ln(y_t) = \beta_1 + \beta_2 x_{2t} + u_t$ ; a 1-Unit Increase in  $x_{2t}$  Causes a  $100 \times \beta_2\%$  Increase in  $y_t$*  (p. 222), *Linear-log:  $y_t = \beta_1 + \beta_2 \ln(x_{2t}) + u_t$ ; a 1% Increase in  $x_{2t}$  Causes a  $0,01 \times \beta_2$ -Unit Increase in  $y_t$*  (p. 222), por Brooks, 2014, *Introductory econometrics for finance* (3rd ed.), Cambridge University Press.

Adicionalmente, la Tabla 9 presenta ejemplos de modelos no lineales transformados en lineales mediante logaritmos.

**Tabla 9**

*Modelos No Lineales Transformados en Lineales por Medio de Logaritmos*

Modelo no lineal	Transformación lineal
$Y = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu}$	$\text{Ln}Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$
$Y = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2} \mu$	$\text{Ln}Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{Ln}\mu \rightarrow \text{Ln}Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu^*$
$Y = \beta_0 e^{\beta_1 X_1 + \beta_2 X_2} \mu$	$\text{Ln}Y = \text{Ln}\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{Ln}\mu \rightarrow \text{Ln}Y = \beta_0^* + \beta_1 X_1 + \beta_2 X_2 + \mu^*$
$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} e^\mu$	$\text{Ln}Y = \text{Ln}\beta_0 + \beta_1 \text{Ln}X_1 + \beta_2 \text{Ln}X_2 + \mu \rightarrow \text{Ln}Y = \beta_0^* + \beta_1 \text{Ln}X_1 + \beta_2 \text{Ln}X_2 + \mu$
$Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \mu$	$\text{Ln}Y = \text{Ln}\beta_0 + \beta_1 \text{Ln}X_1 + \beta_2 \text{Ln}X_2 + \text{Ln}\mu \rightarrow \text{Ln}Y = \beta_0^* + \beta_1 \text{Ln}X_1 + \beta_2 \text{Ln}X_2 + \mu^*$
$e^Y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} \mu$	$Y = \text{Ln}\beta_0 + \beta_1 \text{Ln}X_1 + \beta_2 \text{Ln}X_2 + \text{Ln}\mu \rightarrow Y = \beta_0^* + \beta_1 \text{Ln}X_1 + \beta_2 \text{Ln}X_2 + \mu^*$

Otra solución es incluir potencias de la variable independiente (modelos polinomiales), los que mantienen el supuesto de linealidad paramétrica (Meuleman et al., 2015). Mientras más alto es el orden del polinomio, más flexible es el modelo y es posible el ajuste de más formas; sin embargo, esto implica un mayor número de variables independientes, lo que podría resultar en coeficientes menos precisos (Stock y Watson, 2020). Lo más común es la inclusión del término cuadrático y aún el cúbico, el término cuártico rara vez es empleado (Darlington y Hayes, 2017). Un modelo polinomial de segundo orden o cuadrático (Figura 26a) es:

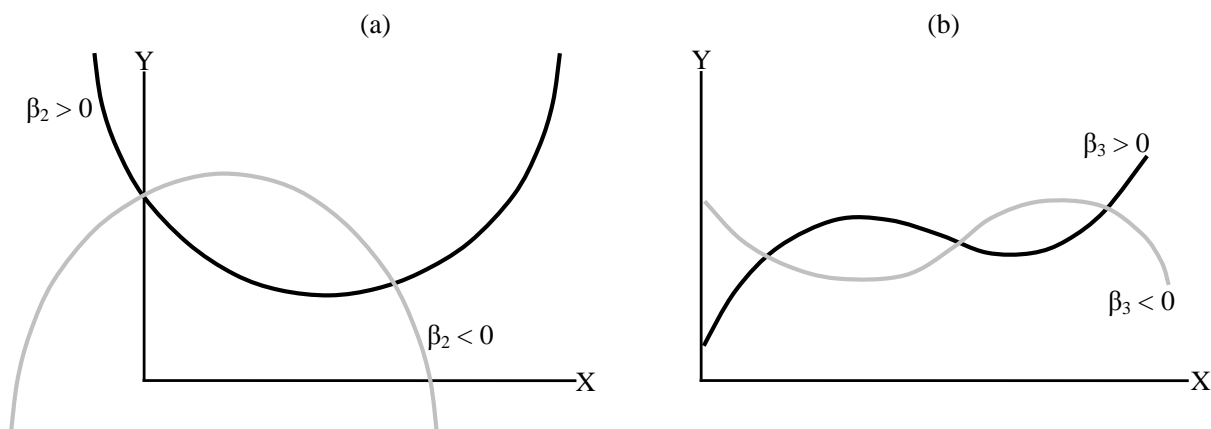
$$Y = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \mu_t$$

Y un modelo polinomial de tercer orden o cúbico (Figura 26b) es:

$$Y = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \beta_3 X_t^3 + \mu_t$$

**Figura 26**

Modelo Polinomial (a) Cuadrático y (b) Cúbico



Nota. Adaptado de *Second-Order Regression Models* (p. 622), *Third-Order Regression Models* (p. 623), por Groebner et al., 2018, *Business statistics: A decision-making approach* (10th ed.), Pearson.

**2.1.10.11.2 Colinealidad.** La solución del problema de colinealidad implica la adopción de algunas pautas. Las pautas comprenden excluir la variable altamente colineal, transformar los datos, mezclar series temporales con datos de corte transversal o incrementar el tamaño muestral; sin embargo, la utilidad de las pautas obedecerá a las características de los datos y a la severidad del problema (Gujarati y Porter, 2010). No obstante, cuando se decide la exclusión de una variable altamente colineal, pero que a la vez es relevante, se podría causar un error de especificación (Doane y Seward, 2016). Y cuando se utiliza la transformación polinomial, es necesario centralizar la variable independiente para mitigar la colinealidad entre las potencias de las variables (Suárez et al., 2017). Una variable se centraliza cuando a cada valor de la variable se le resta su promedio, obteniéndose otra variable con promedio igual a cero (Darlington y Hayes, 2017). Al final, solo se debe considerar remediar la colinealidad cuando origina estadísticos t no significativos o coeficientes estimados poco confiables (Studenmund, 2017), y únicamente después de verificar la no violación del supuesto de especificación correcta (Chatterjee y Hadi, 2012).

**2.1.10.11.3 El Promedio de los Errores No Es Cero.** Es suficiente que el modelo incluya al intercepto para que el promedio de los errores sea cero.

**2.1.10.11.4 Errores Autocorrelacionados.** Cuando se detecta autocorrelación en los errores, primero se debe descartar si se trata de autocorrelación impura, es decir verificar si la forma funcional del modelo es la correcta y que no se hayan omitido variables relevantes (Studenmund, 2017). Descartada la autocorrelación impura, se puede corregir la autocorrelación pura mediante la aplicación de mínimos cuadrados generalizados (MCG) (Quintana y Mendoza, 2016). El método MCG implica aplicar el método MCO a variables que han sido previamente transformadas, de tal manera que satisfagan los supuestos MCO (Gujarati y Porter, 2010; Stock y Watson, 2020).

La transformación empieza planteando, por ejemplo, un modelo de regresión múltiple con dos variables independientes, cuyo error presenta autocorrelación de primer orden:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$$

En la ecuación  $\mu_t = \rho\mu_{t-1} + \varepsilon_t$ , donde  $\varepsilon_t$  es el término de error no autocorrelacionado. Seguidamente, al re-expresar el modelo para el tiempo  $t - 1$  se tiene:

$$Y_{t-1} = \beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + \mu_{t-1}$$

Luego, al multiplicar la ecuación por  $\rho$  resulta:

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{1t-1} + \rho\beta_2 X_{2t-1} + \rho\mu_{t-1}$$

Después, al sustraer la ecuación anterior del modelo de regresión múltiple con dos variables independientes, se tiene la ecuación generalizada siguiente:

$$Y_t - \rho Y_{t-1} = \beta_0 (1 - \rho) + \beta_1 (X_{1t} - \rho X_{1t-1}) + \beta_2 (X_{2t} - \rho X_{2t-1}) + \mu_t - \rho\mu_{t-1}$$

En la ecuación, normalmente  $\rho$  es desconocido (Wooldridge, 2020). Sin embargo, es posible aplicar la técnica iterativa de Prais-Winsten (PW) para obtener un estimado de  $\rho$  corriendo la regresión  $\mu_t = \rho\mu_{t-1} + \varepsilon_t$  (Studenmund, 2017). A todos los métodos que utilizan un  $\rho$  estimado en lugar del  $\rho$  verdadero, se les conoce como MCG factibles (MCGF) o MCG

estimados (MCGE) (Gujarati y Porter, 2010). Finalmente, reemplazando se obtiene la ecuación transformada siguiente:

$$Y_t^* = \beta_0^* + \beta_1 X_{1t}^* + \beta_2 X_{2t}^* + \varepsilon_t$$

Al observar la ecuación generalizada, es evidente que en el proceso se pierde el primer dato, debido a que no tiene precedente. Para no perder el primer dato se puede utilizar la transformación de PW, la cual se obtiene al multiplicar el primer dato por la raíz cuadrada de  $(1 - \rho^2)$  (Young, 2017). El método MCGF transformado por PW es conocido con el nombre de MCGF completo (MCGFC) (Gujarati y Porter, 2010).

**2.1.10.11.5 Errores Heterocedásticos.** Ante errores heterocedásticos, la solución más ampliamente utilizada es la corrección por heterocedasticidad de las desviaciones estándar de los parámetros del modelo, y seguir utilizando el método MCO para calcular las pendientes (Studenmund, 2017). La corrección incrementa las desviaciones estándar de los parámetros, reduce los valores t y hace más difícil decidir acerca de la significancia de los parámetros (Brooks, 2014).

La corrección original o de White es conocida como HC0 (Heteroscedasticity Consistent), posteriormente se realizaron versiones mejoradas conocidas como HC1, HC2, HC3 y HC4 (Darlington y Hayes, 2017). La variante HC3 es la que parece ser la más ampliamente empleada (Weisberg, 2014) y la variante HC4 puede dar mejores resultados cuando el tamaño muestral es pequeño y existen observaciones influyentes (Fox, 2016). Una observación es influyente cuando al omitirla, los estimados MCO difieren significativamente (Wooldridge, 2020).

Otra solución frecuente es reducir la magnitud de cada dato mediante su logaritmo respectivo, siempre y cuando todos los valores de las variables sean positivos (Court y Rengifo, 2011); caso contrario es posible sumar cualquier número positivo a todos los datos, de tal manera que todos los negativos se tornen positivos (Bisgaard y Kulahci, 2011). La

transformación logarítmica estabiliza la varianza del error (McClave y Sincich, 2018) y también es útil porque convierte un modelo multiplicativo, como el modelo exponencial, en un modelo aditivo (Brooks, 2014). Otra transformación útil es la raíz cuadrada de la variable dependiente (McClave y Sincich, 2018).

Una tercera solución es cambiar la especificación del modelo mediante, por ejemplo, la inclusión de otras variables independientes; sin embargo, esta re-especificación puede resultar en la violación de otros supuestos MCO (Marcellino, 2016).

Por último, una cuarta opción es la utilización de MCG, sin embargo para poder aplicar esta opción es necesario conocer la forma de la heterocedasticidad (Court y Rengifo, 2011), por lo cual en la práctica esta opción no es normalmente viable (Brooks, 2014). El método MCG minimiza la suma ponderada de residuos al cuadrado, es decir, con MCG cada dato tiene un peso (ponderación) diferente, por lo que a MCG se le conoce también como mínimos cuadrados ponderados (MCP) (Gujarati y Porter, 2010).

**2.1.10.11.6 Errores No Normales.** Ante errores no distribuidos normalmente se pueden tomar dos medidas. Primero, si la muestra es pequeña, intentar incluir más datos y volver a correr la regresión. Segundo, se puede transformar los datos empleando el método Box-Cox, siendo la transformación logarítmica la más empleada (Quintana y Mendoza, 2016). Por medio de la transformación logarítmica es posible reducir la asimetría (Gujarati y Porter, 2010; Warner, 2013). En una transformación logarítmica, puede que no todos los datos sean positivos, en este caso sumar cualquier número positivo a los datos, de tal manera que todos los negativos se tornen positivos (Bisgaard y Kulahci, 2011). Otra opción es restarle a cada dato el menor valor de los datos originales (Yan y Su, 2009).

La familia de transformaciones Box-Cox, llamadas también transformaciones de potencia, son convenientes ante la inexistencia de referencias de cuál es la mejor transformación (Tattar et al., 2016). Además, estas transformaciones se pueden aplicar a

variables dependientes e independientes (Darlington y Hayes, 2017). La transformación de la variable dependiente es la idónea para solucionar no solo la presencia de errores no normales, sino también el problema de errores heterocedásticos y la ausencia de linealidad (Brooks, 2014; Chatterjee y Hadi, 2012; Kutner et al., 2005). Para explicar esta transformación, Gujarati y Porter (2010) formularon el siguiente modelo Box-Cox:

$$Y_t^\lambda = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$$

En el modelo,  $Y_t$  toma valores positivos, y dado que el parámetro de transformación  $\lambda$  (lambda) puede asumir cualquier valor, se podrán obtener tantas transformaciones (modelos) según los valores que tome  $\lambda$  (Tabla 10) (Gujarati y Porter, 2010). Todos los valores de  $Y_t$  deben ser positivos, debido a que ciertas transformaciones tales como raíz cuadrada, logarítmica y recíproca, son indefinidas para negativos y/o cero (Fox, 2016). En un extremo, si todos los datos son negativos, la transformación se realiza sobre los valores absolutos de los datos originales (Darlington y Hayes, 2017).

**Tabla 10**

*Transformaciones Box-Cox Según el Valor de  $\lambda$*

Valor de $\lambda$	Modelo resultante	Denominación del modelo
2	$Y_t^2 = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	Cuadrado
1	$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	Lineal
0,5	$\sqrt{Y_t} = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	Raíz cuadrada
0	$\ln Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	Exponencial o log-lineal
-0,5	$1 / \sqrt{Y_t} = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	Raíz cuadrada inversa
-1	$1 / Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$	Recíproco

*Nota.* Adaptado de *Según el Valor de  $\lambda$ , Tenemos los Sigüientes Modelos de Regresión, que se Muestran en Forma Tabular* (p. 187), por Gujarati y Porter, 2010, *Econometría* (5a ed.), McGraw-Hill.

El valor de  $\lambda$  permite que la distribución de los errores se aproxime, lo más posible, a la normalidad (Weisberg, 2014). Es usual probar varios  $\lambda$  de fácil interpretación tales como 2, 1,5, 1, 0,5, 0, -0,5, -1, -1,5, -2, y elegir el mejor; al conjunto de valores seleccionados se le conoce como escala de transformación (Chatterjee y Hadi, 2012).

**2.1.10.11.7 Cambio Estructural.** Una solución muy sencilla ante cambios estructurales es la inclusión de variables dummy, las cuales valen uno durante el lapso de cambio estructural y valen cero para el lapso restante (Ghysels y Marcellino, 2018). Además, su inclusión es también considerada como una prueba de la presencia o ausencia de cambios estructurales (Asteriou y Hall, 2011). A las variables dummy también se les conoce como variables indicadoras, variables categóricas, variables ficticias, variables discretas, variables binarias o variables cualitativas.

Las variables dummy incorporan el cambio estructural en el modelo de regresión (Welc y Rodriguez, 2018), es decir, su inclusión permite a los parámetros cambiar en el tiempo (Marcellino, 2016). Estas variables son como otras variables independientes, esto es, se les puede aplicar cualquier prueba estadística pertinente (Jaggia y Kelly, 2020) y pueden removerse si son no significativas (Richardson, 2015).

Según Welc y Rodriguez (2018), el procedimiento para la incorporación de variables dummy es el siguiente:

1. Construir  $n-1$  variables dummy, el primer periodo de la primera variable dummy ( $D_1$ ) vale uno y el resto de periodos (desde 2 hasta  $n$ ) valen cero. Luego, los dos primeros periodos de la segunda variable dummy ( $D_2$ ) valen uno y el resto de periodos (desde 3 hasta  $n$ ) valen cero. Finalmente, los  $n-1$  periodos de la  $n-1$  variable dummy ( $D_{n-1}$ ) valen uno y el resto de periodos ( $n$ ) valen cero (Tabla 11).

**Tabla 11***Construcción de Variables Dummy (D)*

Periodo (n)	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>	D <sub>6</sub>	D <sub>7</sub>	D <sub>8</sub>	...	D <sub>n-1</sub>
1	1	1	1	1	1	1	1	1	...	1
2	0	1	1	1	1	1	1	1	...	1
3	0	0	1	1	1	1	1	1	...	1
4	0	0	0	1	1	1	1	1	...	1
5	0	0	0	0	1	1	1	1	...	1
6	0	0	0	0	0	1	1	1	...	1
7	0	0	0	0	0	0	1	1	...	1
8	0	0	0	0	0	0	0	1	...	1
9	0	0	0	0	0	0	0	0	...	1
⋮	0	0	0	0	0	0	0	0	...	1
n - 1	0	0	0	0	0	0	0	0	...	1
n	0	0	0	0	0	0	0	0	...	0

*Nota:* Adaptado de *Auxiliary Dummy Explanatory Variables (Di)* (p. 161), por Welc y Rodriguez, 2018, *Applied regression analysis for business*, Springer.

2. Construir el modelo de regresión original, sin incluir ninguna variable dummy, y obtener su coeficiente de determinación ( $R^2$ ).
3. Construir n-1 regresiones auxiliares, cada regresión debe incluir las variables independientes originales y una variable dummy a la vez. Obtener el coeficiente de determinación de cada regresión auxiliar ( $R_a^2$ ).
4. Calcular el estadístico F de cada regresión auxiliar para n periodos y k variables independientes originales:

$$F = \frac{(R^2 - R_a^2)}{(1 - R^2)}(n - (k+1))$$

5. Obtener el valor más alto de F ( $F_{\max}$ ).
6. Calcular el valor crítico  $F_c$  para n periodos y k variables independientes originales. Los grados de libertad son  $g_{ln}$  y  $g_{ld}$ , donde  $g_{ln}$  es dos y  $g_{ld} = (n - (k+1))$ .
7. Si  $F_{\max} > F_c$ , entonces la variable dummy asociada a ese  $F_{\max}$  debe incluirse en el modelo.

La variable dummy (D) puede ser incluida en un modelo hasta en tres formas distintas (Ghysels y Marcellino, 2018):

1. Cuando el cambio estructural es en el intercepto se construye el modelo siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_0^* D + \mu_t$$

Si D vale uno se tiene:

$$Y = (\beta_0 + \beta_0^*) + \beta_1 X_1 + \beta_2 X_2 + \mu_t$$

2. Cuando el cambio estructural es en las pendientes se construye el modelo siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_1^* D X_1 + \beta_2^* D X_2 + \mu_t$$

Si D es igual a uno se tiene:

$$Y = \beta_0 + (\beta_1 + \beta_1^*) X_1 + (\beta_2 + \beta_2^*) X_2 + \mu_t$$

3. Cuando el cambio estructural es en el intercepto y en las pendientes, el modelo es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_0^* D + \beta_1^* D X_1 + \beta_2^* D X_2 + \mu_t$$

Si D es igual a uno se tiene:

$$Y = (\beta_0 + \beta_0^*) + (\beta_1 + \beta_1^*) X_1 + (\beta_2 + \beta_2^*) X_2 + \mu_t$$

Y, en cualquiera de los tres casos antes mencionados, si D es igual a cero se tiene:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu_t$$

En las ecuaciones,  $\beta_0^*$  (intercepto diferencial) cuantifica la diferencia del intercepto entre los datos que toman el valor de uno y los datos que toman el valor de cero (Gujarati y Porter, 2010). Mientras que  $\beta_1^*$  y  $\beta_2^*$  (pendientes diferenciales) cuantifican la diferencia de las pendientes entre ambos periodos (Gujarati y Porter, 2010). Otros remedios son la inclusión de variables independientes que expliquen el cambio estructural o la adición de términos no lineales (Heij et al., 2004).

**2.1.10.11.8 Especificación Incorrecta.** Cuando la mala especificación es causada por la inclusión de una variable irrelevante, la solución es la eliminación de esa variable. Y si la omisión de una variable relevante origina mala especificación, el remedio es la inclusión de dicha variable. Sin embargo, cuando la variable relevante no es observable o medible; la medida correctiva es utilizar una variable proxy (Wooldridge, 2020). Una variable proxy es la

que se relaciona, pero no perfectamente, con una variable no observable ni medible (Zax, 2011). Una buena variable proxy es aquella en la que sus fluctuaciones, y las de la variable no observable ni medible, son similares (Studenmund, 2017). La variable proxy puede reemplazar a la variable no observable ni medible, atrapar su influencia y eliminar el problema de especificación (Rosales et al., 2013). Por tanto, incluir una variable proxy posibilita la obtención de coeficientes insesgados y consistentes (Asteriou y Hall, 2011).

Suponga que en el modelo de regresión múltiple  $Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_t$ , la variable independiente  $X_{1t}$  no es observable o medible, pero se relaciona linealmente con cierta variable proxy  $W$ , de tal manera que  $X_{1t} = \delta + \pi W_t$ , entonces por sustitución se tiene:

$$Y = \beta_0 + \beta_1 (\delta + \pi W_t) + \beta_2 X_{2t} + \mu_t$$

$$Y = \beta_0 + \beta_1 \delta + \beta_1 \pi W_t + \beta_2 X_{2t} + \mu_t$$

En el modelo resultante, la variable proxy reemplaza a la variable independiente relevante, por tanto ahora ya no se viola el supuesto de especificación correcta por omisión de variable. Sin embargo, en algunas aplicaciones, es imposible obtener una variable proxy; en estas circunstancias es posible utilizar la variable dependiente rezagada un periodo ( $Y_{t-1}$ ) como proxy (Wooldridge, 2020). Es decir, se plantea el modelo siguiente:

$$Y = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 Y_{t-1} + \mu_t$$

Por último, cuando la especificación incorrecta es causada por la elección de una forma funcional lineal, cuando en realidad la forma funcional correcta es una no lineal, la medida correctiva es la transformación de las variables (ver numeral 2.1.10.11.1, p. 73).

**2.1.10.12 Significancia de los Coeficientes.** Cuando no se violan los supuestos MCO, o caso contrario, se aplican las medidas correctivas pertinentes; es posible verificar la significancia global e individual de los coeficientes.

**2.1.10.12.1 Significancia Global: La Prueba F.** La prueba F tiene múltiples aplicaciones, una de ellas es probar si al menos un coeficiente del modelo de regresión, es

significativamente diferente de cero. Es decir, mediante la prueba F se determina si por lo menos una variable independiente se relaciona linealmente y de manera significativa con la variable dependiente. Las hipótesis de la prueba F son:

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

$$H_1: \text{al menos un } \beta_k (k = 1, 2, \dots, k) \neq 0$$

Cuando el valor p de la prueba F es menor e igual a 5% (0,05),  $H_0$  es rechazada y se determina que, como mínimo, una de las variables independientes se relaciona linealmente y de manera significativa con la variable dependiente. Y si el valor p de la prueba F es mayor a 5% (0,05), no se rechaza  $H_0$  y se concluye que ninguna de las variables independientes se relaciona linealmente con la variable dependiente.

**2.1.10.12.2 Significancia Individual: La Prueba t.** Si al aplicar la prueba F se rechaza  $H_0$ , entonces al menos una de las variables independientes es significativa. La prueba t establece que coeficientes del modelo son significativamente diferentes de cero. Es decir, mediante la prueba t se determina si cada variable independiente, de manera individual, se relaciona significativamente con la variable dependiente. Se plantean tantas hipótesis según variables independientes existan en el modelo:

$$H_0: \beta_1 = 0$$

$$H_0: \beta_2 = 0$$

$$H_0: \beta_k = 0$$

$$H_1: \beta_1 \neq 0$$

$$H_1: \beta_2 \neq 0$$

$$H_1: \beta_k \neq 0$$

Cuando el valor p de la prueba t es menor e igual a 0,05, se rechaza  $H_0$  y se concluye que la variable independiente se relaciona de manera significativa con la variable dependiente y por tanto es útil al modelo. Y si el valor p de la prueba t es mayor a 0,05, no se rechaza  $H_0$  y se concluye que la variable independiente no se relaciona de manera significativa con la variable dependiente y debería ser eliminada del modelo, la variable no es útil al modelo. Además, la prueba t también sirve para verificar la significancia del intercepto ( $\beta_0$ ).

**2.1.10.13 El Coeficiente de Determinación ( $r^2$  o  $R^2$ ).** El coeficiente de determinación, simple ( $r^2$ ) o múltiple ( $R^2$ ), mide el poder explicativo del modelo. Es decir, que tanto las variables independientes explican el comportamiento de la variable dependiente. El rango de valores posibles del  $R^2$  está entre cero y uno, y usualmente es multiplicado por cien. Si  $R^2$  es cero, las variables independientes no explican en absoluto el comportamiento de la variable dependiente. Y cuando  $R^2$  es uno (100%), las variables independientes explican en su totalidad el comportamiento de la variable dependiente.

Sobre el particular, Waters (2011) especificó que un  $R^2$  por encima de 0,50 es calificado como bueno. A su turno, Waller (2008) precisó que un  $R^2$  mayor a 0,80 se considera muy bueno. Y Ryan (2013) manifestó que, para la mayoría de usos, se recomienda un  $R^2$  mayor e igual a 0,90. Por su parte, Rodríguez y González (2017) afirmaron, que un  $R^2$  por debajo de 0,6 indica un ajuste regular o malo, uno entre 0,6 y 0,7 se considera un buen ajuste, uno entre 0,7 y 0,96 significa muy buen ajuste y uno mayor a 0,96 es sospechoso. Sin embargo, cuando la regresión carece de intercepto,  $R^2$  puede ser inferior a cero o superior a uno; si esto ocurre es probable que el modelo haya violado algunos supuestos MCO (Rodríguez y González, 2017). El  $R^2$  se expresa como:

$$R^2 = \frac{\sum(\hat{Y}_t - \bar{Y})^2}{\sum(Y_t - \bar{Y})^2}$$

Martínez (2005), manifestó que  $R^2$  y el tamaño muestral se relacionan inversamente, con un tamaño muestral pequeño se obtiene un  $R^2$  casi perfecto. Además, en el caso de regresión múltiple, McClave y Sincich (2018) señalaron que para usar  $R^2$  como medida de utilidad del modelo para pronosticar, el tamaño muestral debe ser sustancialmente mayor que el número de parámetros. Así mismo Barrow (2017), sostuvo que las regresiones con series temporales presentan  $R^2$  mucho más altos que las regresiones con información de corte

trasversal. Y Montgomery et al. (2015), precisaron que un  $R^2$  grande no es garantía para la obtención de un error de pronóstico pequeño.

Sin embargo Dougherty (2011), manifestó que es posible obtener un  $R^2$  alto, aun con un modelo de regresión con grandes errores de especificación, por lo que  $R^2$  debe ser considerado solo como uno de los criterios para evaluar la utilidad de un modelo.

**2.1.10.14 El Coeficiente de Determinación Corregido ( $r_c^2$  o  $R_c^2$ ).** La utilización del  $R^2$  presenta un problema, a medida que se agregan variables independientes, el  $R^2$  aumenta, aun cuando la variable adicionada no sea estadísticamente significativa y, por tanto, no aporte poder explicativo al modelo. Por tanto, la comparación de diferentes modelos de regresión múltiple es mejor con el coeficiente de determinación corregido, simple ( $r_c^2$ ) o múltiple ( $R_c^2$ ), ajustado para el tamaño muestral ( $n$ ) y el número de variables ( $k$ ) (Jaggia y Kelly, 2020; Fabozzi et al., 2014);  $R_c^2$  mide de forma más precisa el ajuste de un modelo (Chase, 2013). A diferencia del  $R^2$ ,  $R_c^2$  puede reducirse si la inclusión de una variable independiente aporta muy poco poder explicativo (Kennedy, 2008). Inclusive  $R_c^2$  puede ser menor a cero si  $R^2$  es pequeño y son muchas las variables independientes (Anderson et al., 2017). Un  $R_c^2$  negativo es equivalente a cero (Martínez, 2005). Matemáticamente el  $R_c^2$  se expresa como:

$$R_c^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

Para utilizar  $R^2$  y  $R_c^2$ , el tamaño muestral y la variable dependiente deben ser los mismos (Gujarati y Porter, 2010). Al restar  $R_c^2$  de  $R^2$  se obtiene la merma (Hahs-Vaughn, 2017). Mientras más grande sea la merma, más pobre será el ajuste del modelo (Chase, 2013). Además, las mermas grandes indican sobreajuste del modelo o micronumerosidad (McCormick y Salcedo, 2015). Y, cuando el tamaño muestral es muy grande, en relación al número de variables independientes, entonces  $R^2$  y  $R_c^2$  no serán muy diferentes (Fabozzi et al., 2014).

Sin embargo, a pesar de sus ventajas, el uso de  $R_c^2$  es limitado, pues la inclusión de una variable eleva el  $R_c^2$ , siempre que el valor absoluto del valor t de la variable supere la unidad, sin que necesariamente, esa variable sea significativa (Dougherty, 2011).

**2.1.10.15 Pronóstico Fuera de la Muestra.** Realizar un pronóstico fuera de la muestra es de suma importancia para estimar la precisión del modelo (Chase, 2016). El que un modelo reproduzca la información histórica de la mejor manera no implica, necesariamente, la obtención de pronósticos precisos (Hanke y Wichern, 2014). Por lo que la prueba final de un modelo radica en su capacidad para realizar pronósticos fuera de la muestra (Stock y Watson, 2020). Un pronóstico fuera de la muestra pronostica los valores futuros de la variable dependiente, conociendo los valores de las variables independientes (Gujarati y Porter, 2010). El método consiste en construir un modelo tomando una parte de los datos y reservar el resto para establecer la capacidad de pronóstico del modelo (Wooldridge, 2020). La información reservada es normalmente alrededor del 20%, aunque depende del tamaño muestral y del número de periodos a pronosticar (Hyndman y Athanasopoulos, 2018). En los proyectos de inversión, el horizonte de pronóstico de demanda debe ser de al menos cinco años (Baca, 2013). Aunque, realizar pronósticos para los próximos cinco o 10 años involucra riesgo y gran dificultad (Kumar y Suresh, 2009).

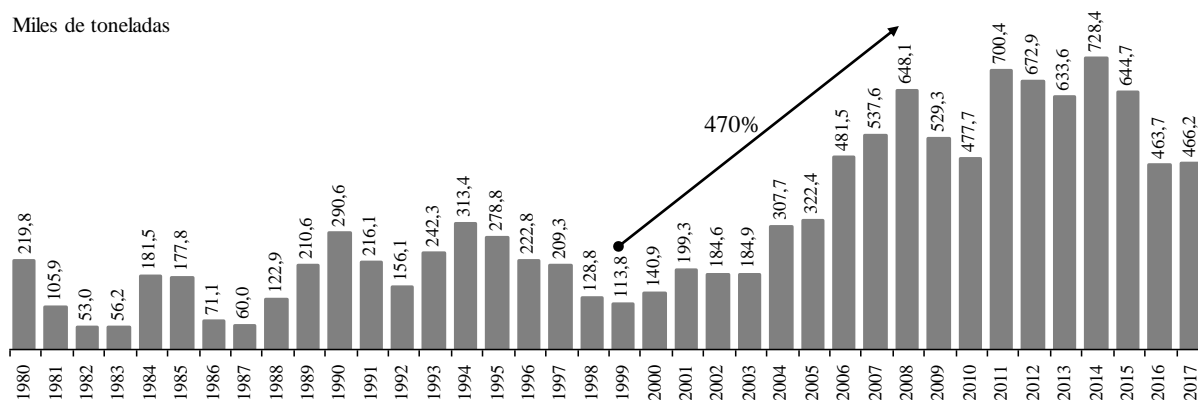
### ***2.1.11 La Industria de Productos Hidrobiológicos Congelados en el Perú***

En el Perú, la industria de productos hidrobiológicos congelados tuvo su origen en 1936, ese año inició sus operaciones la Compañía Nacional de Pesca, la cual se dedicó a la exportación de pescado congelado a los Estados Unidos de Norteamérica. Desde 1936, y hasta 1955, la industria pesquera se dedicó casi exclusivamente a la producción de congelados y enlatados. En la década de 1950, capitalistas peruanos y la empresa norteamericana Wilbur Ellis, se asociaron para producir congelados para el mercado estadounidense (Maldonado y Puertas, 2011).

**2.1.11.1 Desembarque para la Producción de Congelados (1980-2017).** Según cifras del INEI (1996) y del PRODUCE (2005, 2007, 2010 y 2019), entre los años 1980-2017, el desembarque para la producción de congelados se expandió en 112%, de 219,8 miles de t a 466,2 miles de t, lo cual representó un aumento promedio anual de 2,05%. Sin embargo, el crecimiento no fue constante. Entre los años 1980-1998, el desembarque mostró un comportamiento irregular, el desembarque máximo se alcanzó en 1994 (313,4 miles de t) y el desembarque mínimo se registró en 1982 (53,0 miles de t). Luego, durante el periodo 1999-2008, se presentó un ciclo expansivo, durante el cual el desembarque creció en 470%, de 113,8 miles de t (1999) a 648,1 miles de t (2008), lo que significó un crecimiento promedio de 21,32% anual. Después, el desembarque decreció hasta 2010 (477,7 miles de t), revirtiendo la caída en 2011 (700,4 miles de t) y en 2014 se logró el desembarque record histórico (728,4 miles de t), luego volvió a decrecer hasta 2016 (463,7 miles de t) y se recuperó ligeramente en 2017 (466,2 miles de t) (Figura 27).

**Figura 27**

*Perú: Desembarque para la Producción de Congelados (1980-2017)*



*Nota.* Fuente INEI, PRODUCE.

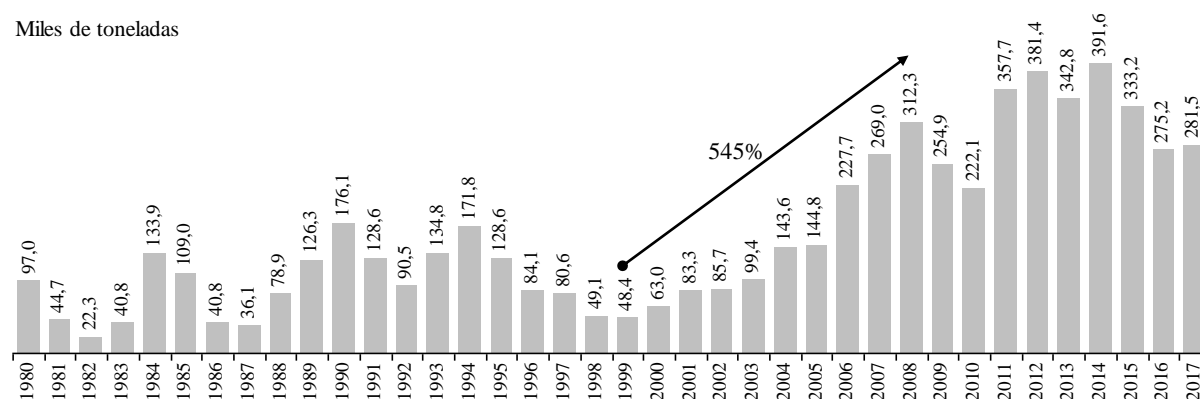
**2.1.11.2 Producción de Congelados (1980-2017).** De acuerdo a los anuarios estadísticos del INEI (1996) y del PRODUCE (2005, 2010 y 2019) entre los años 1980-2017, la producción de congelados se expandió en 190%, de 97,0 miles de t a 281,5 miles de t, lo cual representó un incremento medio anual de 2,92%. Sin embargo, la expansión no fue

continua. Entre los años 1980-1998, la producción de congelados mostró un comportamiento irregular, el record del periodo se alcanzó en 1990 (176,1 miles de t) y el mínimo se registró en 1982 (22,3 miles de t). Luego, entre 1999-2008 ocurrió un ciclo expansivo, en este periodo el record de producción sucedió en 2008 (312,3 miles de t). Durante el periodo de referencia la producción creció en 545%, lo que significó una expansión promedio anual de 23,02%. Después, la producción decreció hasta 2010 (222,1 miles de t), revirtiendo la caída en 2011. En 2014 se logró el record histórico (391,6 miles de t), luego decreció hasta 2016 (275,2 miles de t) y se recuperó ligeramente en 2017 (281,5 miles de t) (Figura 28).

### Figura 28

*Perú: Producción de Productos Hidrobiológicos Congelados (1980-2017)*

Miles de toneladas

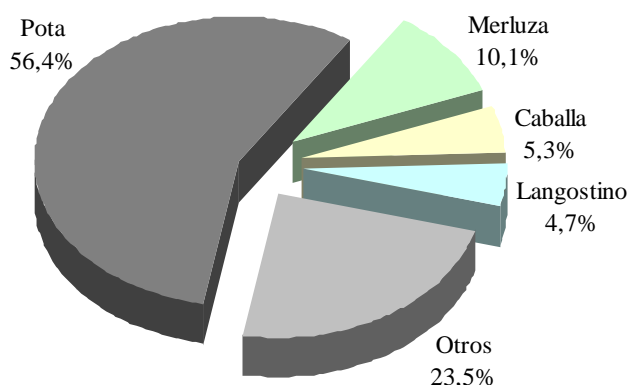


Nota. Fuente INEI, PRODUCE.

**2.1.11.3 Producción de Congelados Según Especie (1988-2017).** Según las estadísticas oficiales del INEI (1996 y 1998) y del PRODUCE (2007, 2010 y 2019), entre los años 1988 y 2017 la producción de congelados totalizó 5 567 093 t. Durante este periodo la especie dominante fue pota (56,4%), seguido por merluza (10,1%), caballa (5,3%) y langostino (4,7%) (Figura 29).

**Figura 29**

*Perú: Producción de Productos Hidrobiológicos Congelados Según Especie (1988-2017)*

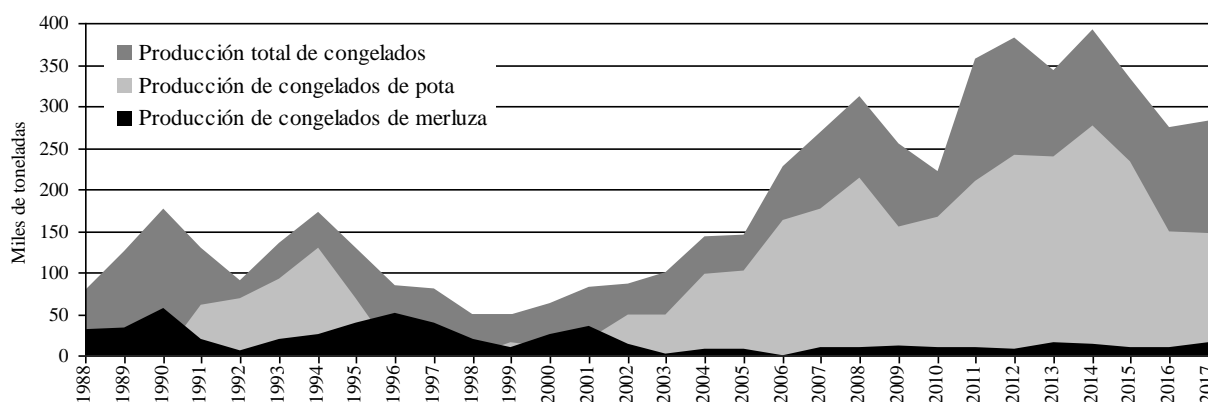


*Nota.* Fuente INEI, PRODUCE.

Entre 1988 y 1990, el recurso merluza fue la especie principal de la industria de congelados. Luego, desde 1991 (60,6 miles de t, 47,1%) y hasta 1995 (68,9 miles de t, 53,6%), el recurso pota dominó el rubro. Sin embargo, durante el periodo 1996-2001 su importancia decreció, siendo su participación máxima de 31,2% con 15,1 miles de t (1999). Y desde 2002 (49,3 miles de t, 57,5%) en adelante, el recurso pota volvió a dominar el rubro. La producción record de congelados de pota se registró en 2014 (276,5 miles de t) y su mayor participación se logró en 2010 (75,1%) (Figuras 30 y 31).

**Figura 30**

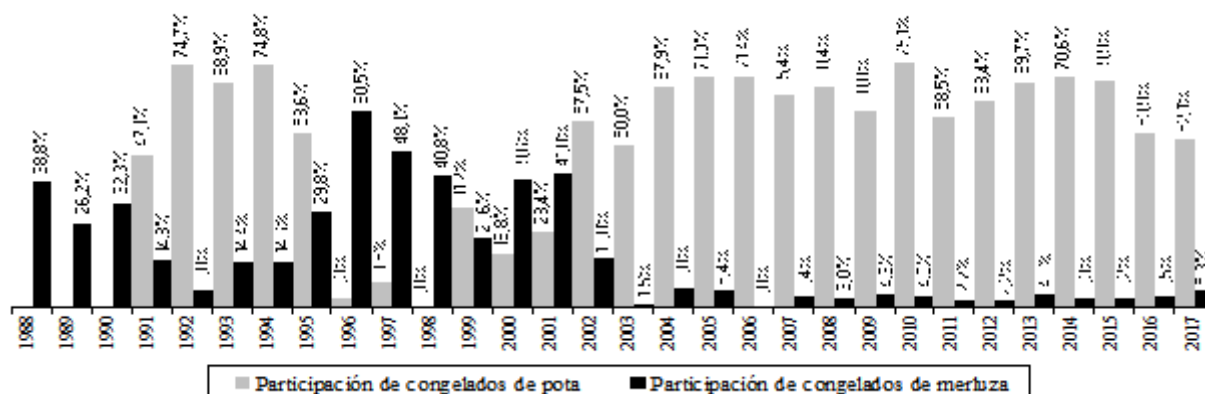
*Perú: Producción de Congelados Según Especies Principales (1988-2017)*



*Nota.* Fuente INEI, PRODUCE.

**Figura 31**

*Perú: Participación de Congelados de Pota y Merluza en la Producción Total (1988-2017)*

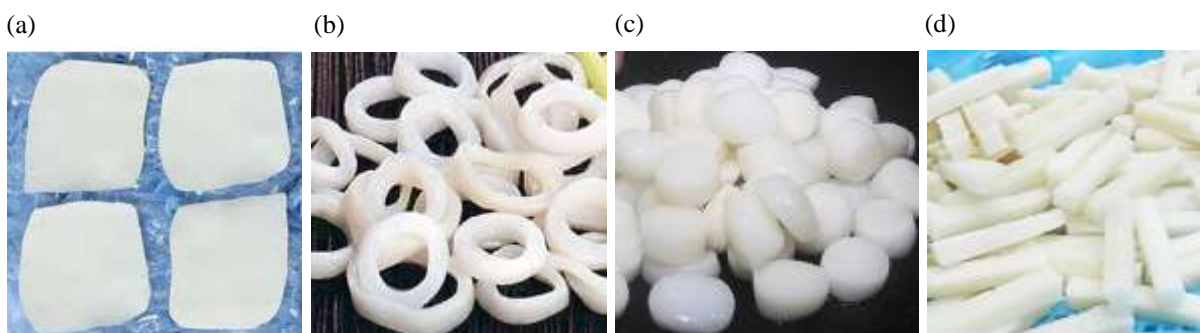


Nota. Fuente INEI, PRODUCE.

Como ha quedado evidenciado la especie principal del rubro congelados es la pota, por lo cual se muestra a continuación algunas de sus presentaciones, tales como filetes precocidos (Figura 32a), anillas (Figura 32b), botones (Figura 32c) y rabas (Figura 32d).

**Figura 32**

*Pota Congelada Presentada en (a) Filetes Precocidos, (b) Anillas, (c) Botones y (d) Rabas*



Nota. Tomado de (a) <https://pe.all.biz/filete-de-pota-precocidas-congelada-g36184>

(b) [http://burruezocongelados.es/tienda-online/2-anillas\\_de\\_calamar-110005](http://burruezocongelados.es/tienda-online/2-anillas_de_calamar-110005)

(c) [https://www.produmar.com/es/productos\\_inter\\_categoria.php?id3=33&id=5&idd=26](https://www.produmar.com/es/productos_inter_categoria.php?id3=33&id=5&idd=26)

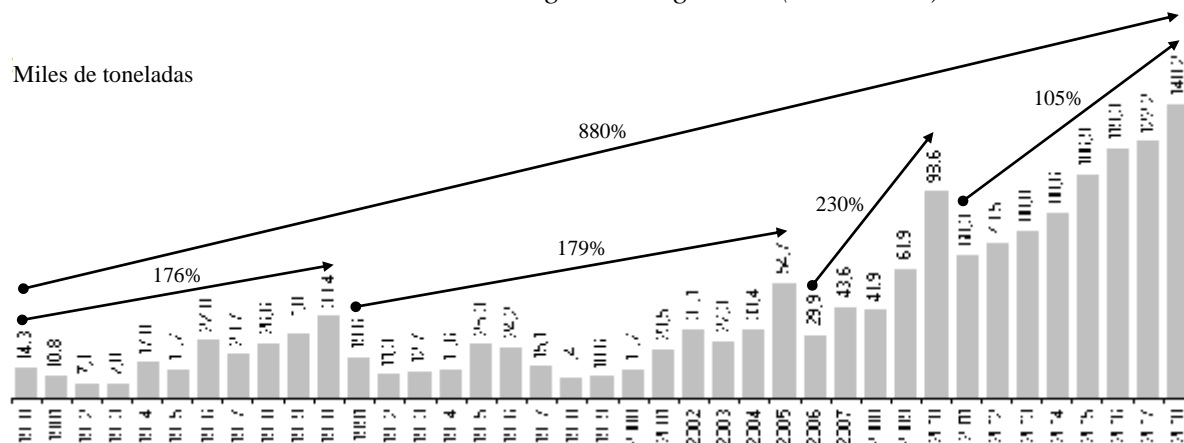
(d) <https://www.perupez.com/web/pota.php>

**2.1.11.4 Demanda de Congelados (1980-2018).** De acuerdo a los compendios estadísticos de Webb y Fernández (1992), del INEI (1998) y del PRODUCE (2007, 2010, 2018a, 2018b y 2019), durante el periodo 1980-2018 la demanda de productos hidrobiológicos congelados en el Perú tuvo una tendencia expansiva. En 1980 se demandaron 14,3 miles de t y en 2018 la demanda llegó al histórico de 140,2 miles de t, lo cual se tradujo en una expansión de 880% y en un crecimiento promedio de 6,19% anual (Figura 33).

Al analizar la serie se puede apreciar cuatro momentos expansivos. Durante el primer momento (1980-1990) se alcanzó en 1990 el primer record (39,4 miles de t), en este periodo la demanda creció en 176%. En el segundo momento (1991-2005), en general se apreció una tendencia creciente, aunque en 1991 se consumió la mitad de 1990 y algo similar ocurrió en 1992 con respecto a 1991, y en los años posteriores el comportamiento fue irregular, llegando a alcanzar en 2005 otra demanda hito (54,7 miles de t); durante este periodo la demanda se expandió en 179%. Durante el tercer momento (2006-2010) se retoma el crecimiento y en 2010 se llega a otro record (98,6 miles de t), en este lustro la demanda aumentó en 230%. Finalmente, en el cuarto momento (2011-2018) se observó, otra vez, un comportamiento creciente, aunque a diferencia de los otros ciclos expansivos, la demanda creció en todos los años; en este periodo la demanda aumentó en 105% (Figura 33).

**Figura 33**

*Perú: Demanda de Productos Hidrobiológicos Congelados (1980-2018)*

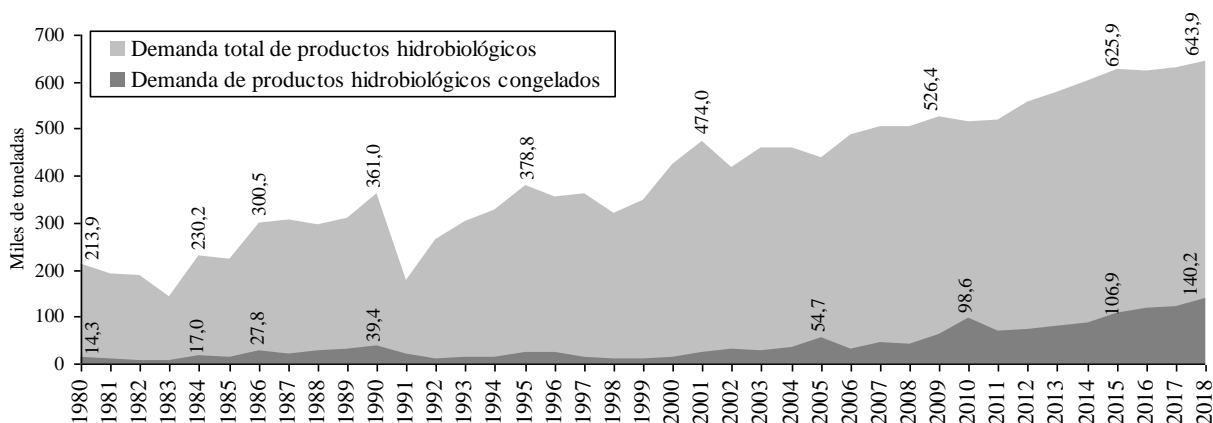


Nota. Fuente INEI, PRODUCE.

Así mismo, la contribución de productos hidrobiológicos congelados en la demanda total de productos hidrobiológicos, ha sido cada vez más relevante. En 1980 la demanda de congelados significó el 6,7% de la demanda total, mientras que para 2018 se registró la participación histórica de 21,8%. Es decir, entre 1980 y 2018 la contribución de productos hidrobiológicos congelados creció en 2,25 veces. Cabe mencionar que la demanda total de productos hidrobiológicos se expandió en 200%, de 213,9 miles de t (1980) a 643,9 miles de t (2018), mientras que la demanda de congelados lo hizo en 880% (Figuras 34 y 35).

**Figura 34**

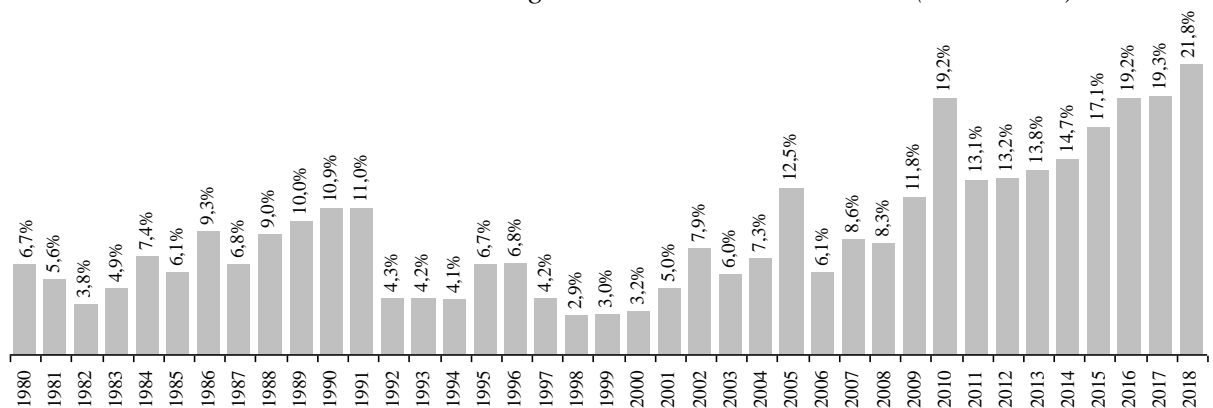
*Perú: Demanda de Productos Hidrobiológicos y de Congelados (1980-2018)*



Nota. Fuente INEI, PRODUCE.

**Figura 35**

*Perú: Contribución de la Demanda Congelados en la Demanda Total (1980-2018)*

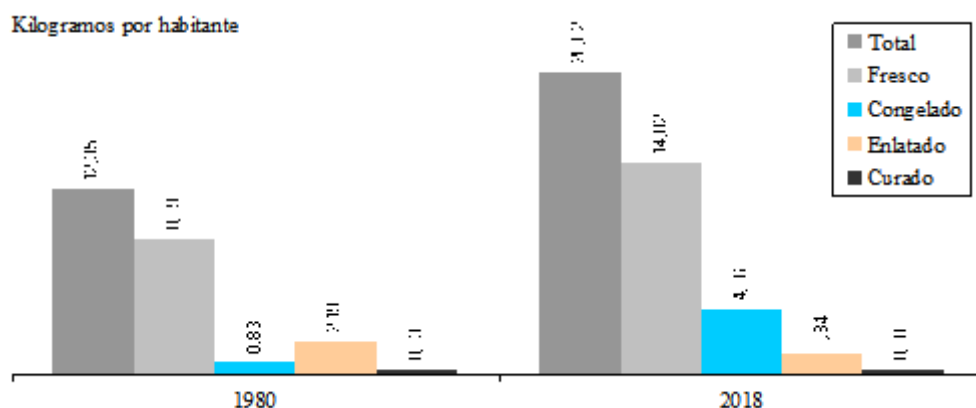


Nota. Fuente INEI, PRODUCE.

**2.1.11.5 Demanda Per Cápita de Productos Hidrobiológicos (1980 y 2018).** Los productos hidrobiológicos se componen de cuatro rubros: Fresco, congelado, enlatado y curado. Al comparar los años 1980 y 2018, se aprecia que la demanda per cápita de productos hidrobiológicos se expandió en 62,1%, de 12,35 kg por habitante (1980) a 20,02 kg por habitante (2018). Sin embargo, no todos los rubros mostraron el mismo comportamiento. Por un lado, el per cápita de los rubros fresco y congelado creció en 56,0% y 425,3%, respectivamente. Por otro lado, el per cápita de los rubros enlatado y curado decreció en 38,8% y 9,1%, respectivamente (Figura 36).

**Figura 36**

*Perú: Demanda Per Cápita de Productos Hidrobiológicos Según Rubros (1980 y 2018)*



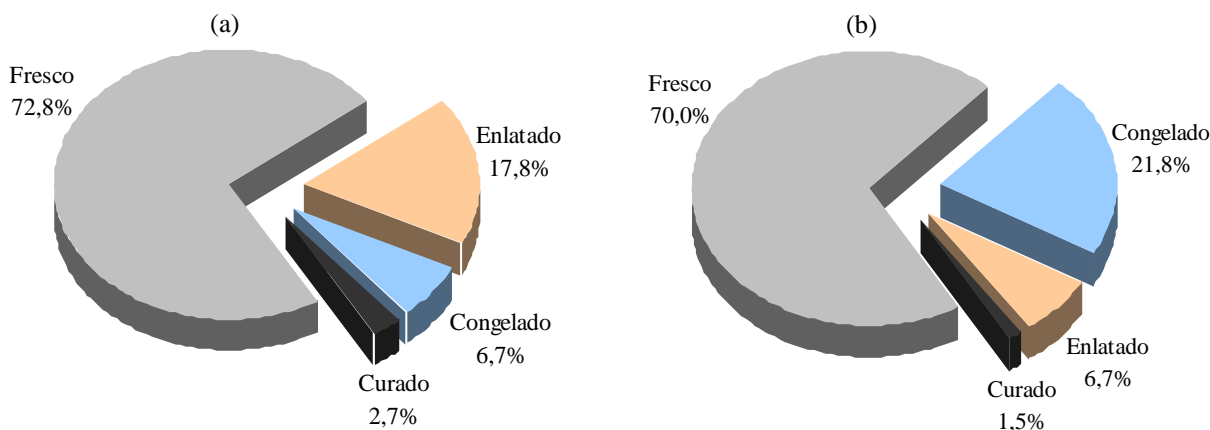
*Nota.* Fuente INEI, PRODUCE.

En cuanto a participación en la demanda per cápita de productos hidrobiológicos, solo el rubro congelados experimentó un incremento, de 6,7% (1980) a 21,8% (2018). En tanto que la participación del rubro fresco decreció de 72,8% (1980) a 70,0% (2018). Lo propio ocurrió con el rubro enlatado al contraerse de 17,8% (1980) a 6,7% (2018). Finalmente, la participación del rubro curado se redujo de 2,7% (1980) a 1,5% (2018) (Figura 37).

**Figura 37**

*Perú: Contribución en la Demanda Per Cápita de Productos Hidrobiológicos Según Rubros*

*(a) en 1980 y (b) en 2018*

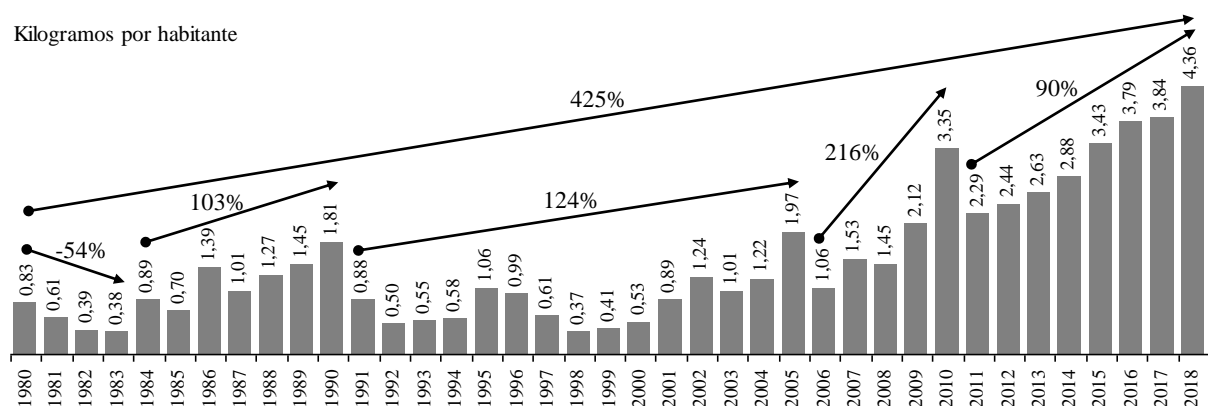


*Nota.* Fuente INEI, PRODUCE.

**2.1.11.6 Demanda Per Cápita de Congelados (1980-2018).** Entre los años 1980 y 2018, la demanda per cápita de congelados se expandió en 425%, de 0,83 kg por habitante (1980) a 4,36 kg por habitante (2018). Sin embargo, al inicio de la serie el per cápita se contrajo en 54%, al pasar de 0,83 kg por habitante (1980) a 0,38 kg por habitante (1983). Luego sucedieron cuatro periodos expansivos. En el primer periodo, el per cápita creció en 103%, de 0,89 kg por habitante (1984) a 1,81 kg por habitante (1990). Durante el segundo periodo, el per cápita aumentó en 124%, de 0,88 kg por habitante (1991) a 1,97 kg por habitante (2005). Mientras que el tercer periodo fue el más expansivo, el per cápita aumentó en 216%, de 1,06 kg por habitante (2006) a 3,35 kg por habitante (2010). Finalmente, en el cuarto periodo el per cápita experimentó una expansión de 90%, al crecer de 2,29 kg por habitante (2011) a 4,36 kg por habitante (2018) (Figura 38).

**Figura 38**

*Perú: Demanda Per Cápita de Congelados (1980-2018)*

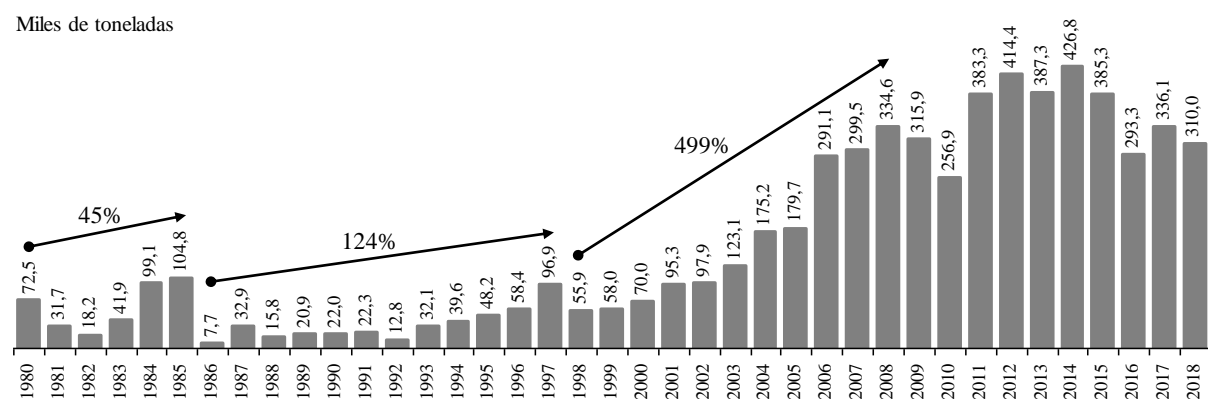


Nota. Fuente INEI, PRODUCE.

**2.1.11.7 Exportaciones de Congelados en Volumen (1980-2018).** De acuerdo a las cifras del INEI (1996), PRODUCE (2005, 2007, 2010, 2019) y PROMPERÚ (2019), entre 1980 y 2018 el volumen de las exportaciones peruanas de congelados crecieron un 328%, de 72,5 miles de t (1980) a 310,0 miles de t (2018), lo cual significó una expansión media anual de 3,90%. El ciclo expansivo más importante ocurrió durante el periodo 1998-2008, en el cual las exportaciones pasaron de 55,9 miles de t (1998) a 334,6 miles de t (2008), lo cual significó un crecimiento de 499% y un promedio de 19,59% por año. El record histórico sucedió en 2014, año en el cual las exportaciones alcanzaron las 426,8 miles de t (Figura 39).

**Figura 39**

*Perú: Exportación de Congelados en Volumen (1980-2018)*

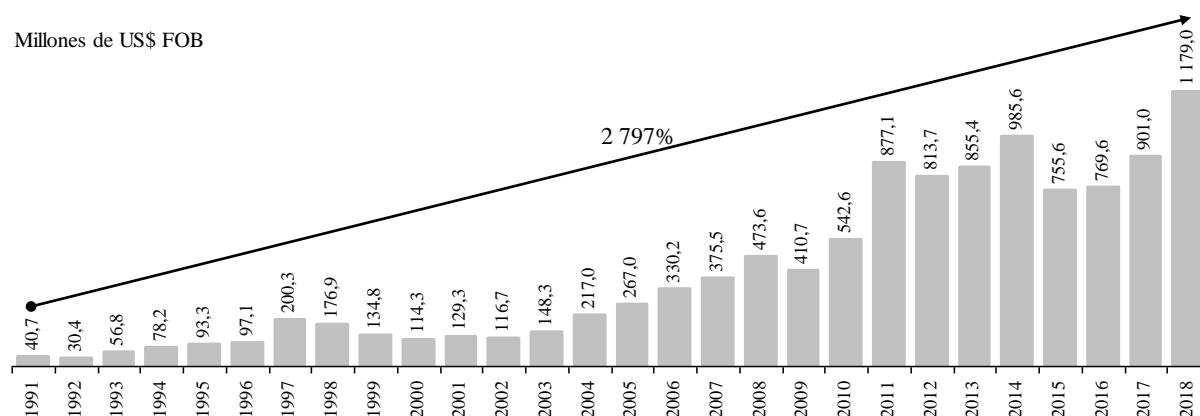


Nota. Fuente INEI, PRODUCE, PROMPERÚ.

**2.1.11.8 Exportaciones de Congelados en Valor FOB (1991-2018).** Según cifras oficiales del PRODUCE (2005, 2007, 2010 y 2019) y de PROMPERÚ (2019), durante el periodo 1991-2018 el valor FOB, de las exportaciones de productos hidrobiológicos congelados, experimentaron una fuerte tendencia creciente. En 1991 se exportaron congelados por un valor de US\$ 40,7 millones FOB y en 2018 el valor de las exportaciones alcanzó el histórico de US\$ 1 179,0 millones FOB, lo cual se tradujo en una expansión de 2 797% y en un crecimiento promedio de 13,3% anual (Figura 40).

**Figura 40**

*Perú: Valor FOB de las Exportaciones de Congelados (1991-2018)*

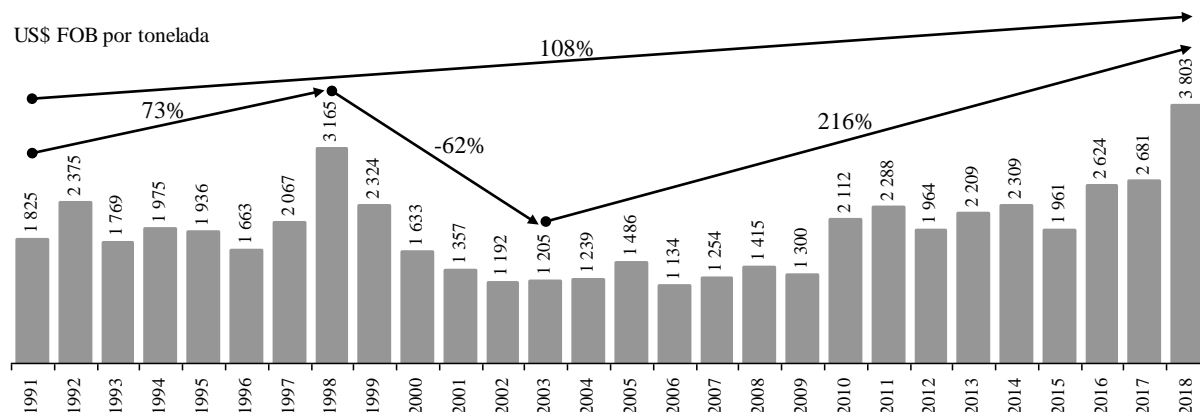


*Nota.* Fuente PRODUCE, PROMPERÚ.

**2.1.11.9 Exportaciones de Congelados en Valor FOB Promedio (1991-2018).** De acuerdo a la información estadística del PRODUCE (2005, 2007, 2010 y 2019) y de PROMPERÚ (2019), en el lapso 1991-2018, el valor FOB promedio de las exportaciones de congelados tuvo, en general, un comportamiento creciente. En 1991 el valor FOB promedio fue de US\$ 1 825/t y en 2018 fue de US\$ 3 803/t, lo que significó una expansión de 108% y un crecimiento medio anual de 2,76%. Al interior de la serie ocurrió un periodo expansivo (73%), seguido por uno contractivo (-62%) y continuado por otro expansivo (216%) (Figura 41).

**Figura 41**

*Perú: Valor FOB Promedio de las Exportaciones de Congelados (1991-2018)*

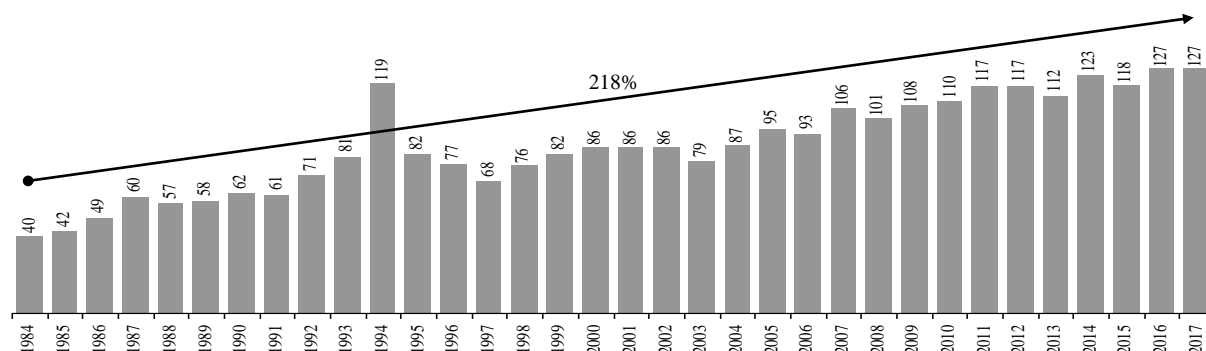


Nota. Fuente PRODUCE, PROMPERÚ.

**2.1.11.10 Capacidad Instalada (1984-2017).** Según los datos de Webb y Fernández (1992), el INEI (1996, 2002, 2009a y 2017) y el PRODUCE (2019), entre 1984 y 2017, la capacidad instalada de la industria de productos hidrobiológicos congelados tuvo una tendencia notoriamente creciente, aunque con algunos altibajos. En 1984 la industria de congelados estaba compuesta por 40 plantas que sumaban una capacidad instalada de 982 t/día y tenía una media de 24,6 t/día. Mientras que en 2017 el número de plantas ascendió a 127, alcanzando una capacidad instalada de 9 696 t/día y un promedio de 76,4 t/día. Es decir, el número de plantas se expandió en 218% a un ritmo anual de 3,6%. De igual manera, la capacidad instalada total creció en 887% a un promedio anual de 7,2%. Finalmente, la capacidad instalada promedio aumentó en 211% con una media de 3,5% por año (Figuras 42, 43 y 44).

**Figura 42**

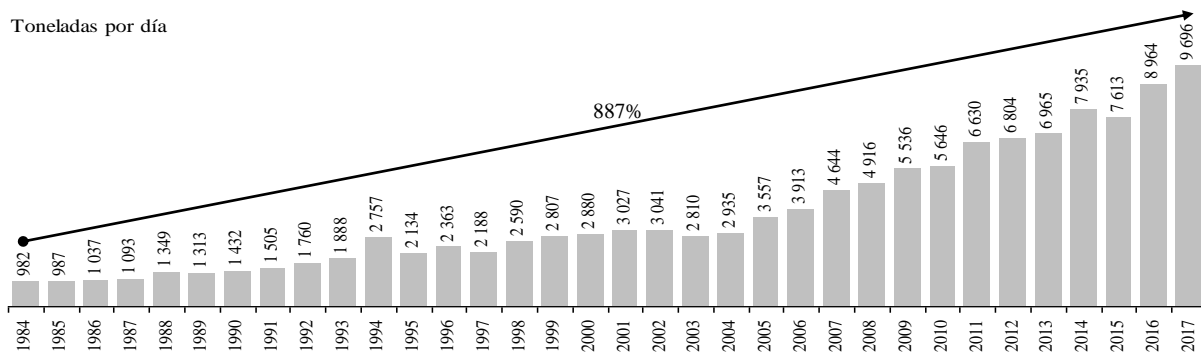
*Perú: Número de Plantas de Productos Hidrobiológicos Congelados (1984-2017)*



Nota. Fuente Webb y Fernández, INEI, PRODUCE.

**Figura 43**

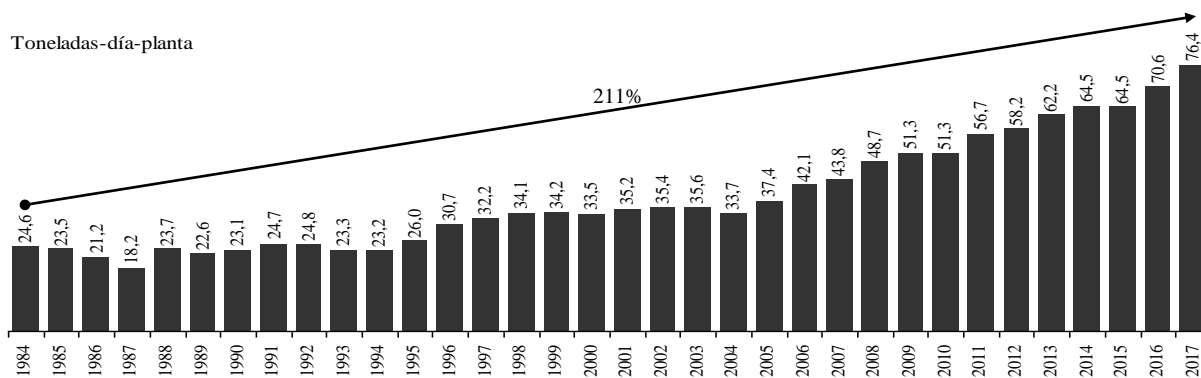
*Perú: Capacidad Instalada de la Industria de Productos Hidrobiológicos Congelados (1984-2017)*



Nota. Fuente Webb y Fernández, INEI, PRODUCE.

**Figura 44**

*Perú: Capacidad Instalada Promedio de la Industria de Productos Hidrobiológicos Congelados (1984-2017)*

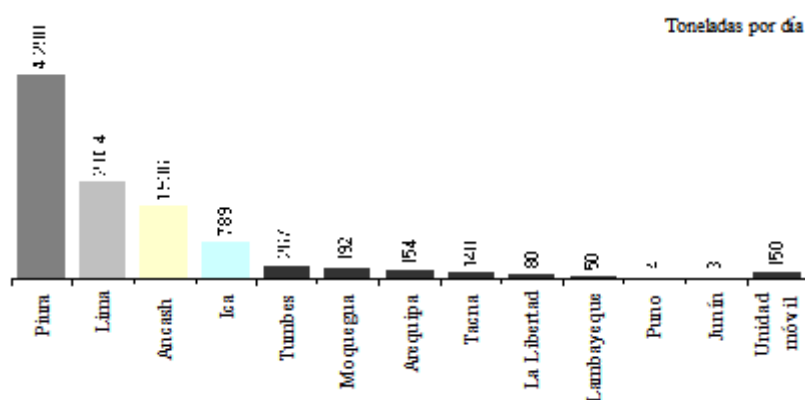


Nota. Fuente Webb y Fernández, INEI, PRODUCE.

**2.1.11.11 Capacidad Instalada Según Región (2017).** Según el PRODUCE (2019), en 2017 la región con mayor capacidad instalada para la producción de congelados fue Piura (4 298 t/día, 44,3%), seguido de Lima (2 034 t/día, 21,0%), Ancash (1 536 t/día, 15,8%) e Ica (789 t/día, 8,1%) (Figuras 45 y 46).

**Figura 45**

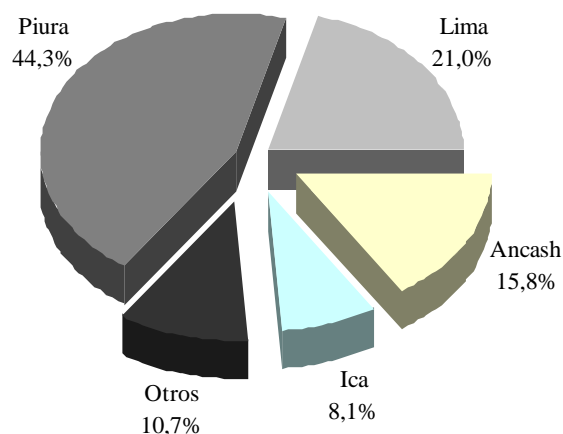
*Perú: Capacidad Instalada de la Industria de Productos Hidrobiológicos Congelados Según Región (2017)*



*Nota.* Fuente PRODUCE.

**Figura 46**

*Perú: Participación de la Capacidad Instalada de la Industria de Productos Hidrobiológicos Congelados Según Región (2017)*

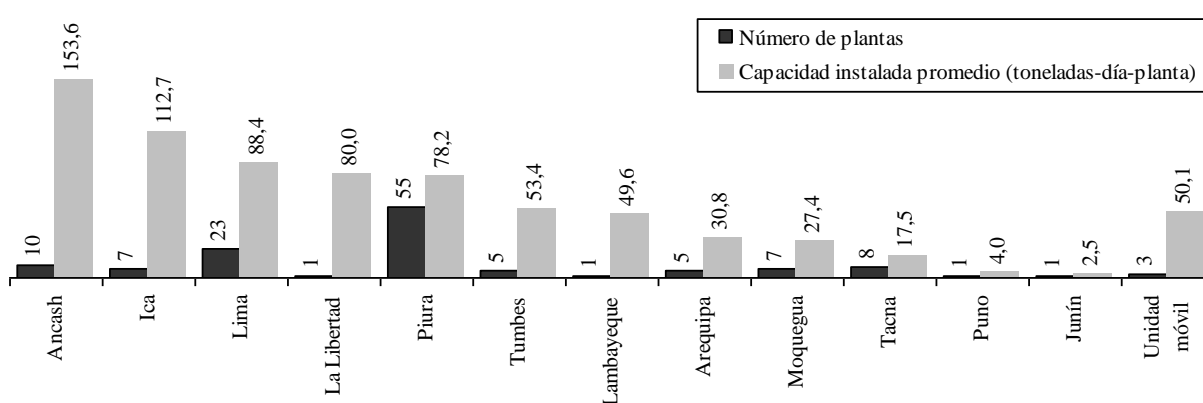


*Nota.* Fuente PRODUCE.

De igual forma, la región Piura concentró el mayor número de plantas (55), seguido de Lima (23) y Ancash (10). Sin embargo, la región Ancash tuvo la más alta capacidad instalada promedio (153,6 t/día) seguido de Ica (112,7 t/día), Lima (88,4 t/día), La Libertad (80,0 t/día) y Piura (78,2 t/día) (Figura 47).

**Figura 47**

*Perú: Número de Plantas y Capacidad Instalada Promedio de la Industria de Productos Hidrobiológicos Congelados Según Región (2017)*



*Nota.* Fuente PRODUCE.

### 2.1.12 El Software Gretl

Gretl (Gnu Regression, Econometrics and Time series Library) es un programa econométrico gratuito, fácil de usar, potente, amigable, expandible, flexible, preciso, modificable y reparable (Adkins, 2018). Entre sus múltiples opciones dispone de técnicas econométricas de vanguardia y puede instalarse en sistemas Linux, MS Windows y Mac OSX (Ducoing, 2015). Además, está disponible en diversos idiomas que incluyen el inglés, francés, italiano, español, polaco, portugués, alemán, vasco, turco, ruso, albanés y griego (Cottrell y Lucchetti, 2022b).

### **III Método**

#### **3.1 Tipo de Investigación**

La investigación es de tipo cuantitativa y aplicada. La investigación cuantitativa recopila información numérica de tipo cuantificable y estandarizada, aplica métodos estadísticos para analizarlos y realizar inferencia, con la finalidad de generalizar los resultados. La investigación aplicada emplea los conocimientos de la investigación básica, con la finalidad de solucionar inconvenientes de tipo práctico, empírico y tecnológico que afectan al aparato productivo de la sociedad (Muñoz, 2011).

El nivel de la investigación es el correlacional. La investigación correlacional pretende revelar y establecer el tipo de relación (positiva, negativa o nula) existente entre dos o más conceptos, personas, acontecimientos, categorías o variables (Domínguez et al., 2009). Además, presenta valor explicativo parcial al establecer que dos variables se relacionan de alguna manera (Hernández Sampieri y Mendoza, 2018).

El diseño de la investigación es no experimental, correlacional-longitudinal. En la investigación no experimental, las variables independientes no son premeditadamente alteradas para establecer cómo afectan otras variables. En los estudios longitudinales se recopila información en cierto horizonte temporal, para realizar pruebas de hipótesis sobre cómo ha evolucionado el problema, sus causas y sus efectos (Hernández Sampieri y Mendoza, 2018).

#### **3.2 Población y Muestra**

##### **3.2.1 Población**

Las variables independientes, que por lo general se utilizan para pronosticar la demanda, son indicadores macroeconómicos como PBI, PBI per cápita, inflación, tipo de cambio, desempleo o desempleo abierto. Otras variables tales como PBI sectorial, balanza comercial, población económicamente activa (PEA), etc. pueden también utilizarse (Baca,

2013). En particular, cuando se desea pronosticar la demanda de alimentos, las variables independientes incluyen población, ingreso personal y PBI (Carbonel, 2016).

Para establecer, tanto la población como la muestra, se seguirá el criterio que empleó Barreto (2018). Según el criterio, la población está compuesta por las siguientes series temporales:

- Demanda de productos hidrobiológicos congelados en miles de toneladas (1980-2018)
- Población (1950-2018)
- Producto Bruto Interno (PBI) en miles soles constantes de 2007 (1950-2018)
- PBI pesquero en miles de soles constantes de 2007 (1950-2018)
- PBI manufacturero en miles de soles constantes de 2007 (1950-2018)
- Demanda interna en miles de soles constantes de 2007 (1950-2018)
- Inversión privada en miles de soles constantes de 2007 (1950-2018)
- Inflación (1950-2018)

### **3.2.2 Muestra**

La muestra está compuesta por las series temporales listadas en el numeral 3.2.1, pero para el periodo 1980-2018 (Anexo 2), es decir, las series tendrán una extensión de 39 años:

- Demanda de productos hidrobiológicos congelados en miles de toneladas (CONG)
- Población (POB)
- Producto Bruto Interno en miles soles constantes de 2007 (PBI)
- PBI pesquero en miles de soles constantes de 2007 (PBIpes)
- PBI manufacturero en miles de soles constantes de 2007 (PBIman)
- Demanda interna en miles de soles constantes de 2007 (DI)
- Inversión privada en miles de soles constantes de 2007 (IP)
- Inflación (INF)

A la serie temporal INF se le ha sumado una unidad, debido a que en el año 2001 la inflación fue negativa (-0,13%), por lo cual la abreviatura de la nueva variables es INF1.

### 3.3 Operacionalización de Variables

En la investigación, se ha identificado como variable independiente (X) al análisis de regresión lineal y como variable dependiente (Y) a la precisión del pronóstico de demanda. La Tabla 12 presenta las definiciones operacionales de tales variables, así como sus dimensiones e indicadores.

**Tabla 12**

#### *Operacionalización de Variables*

X: Análisis de regresión lineal		
Definición operacional	Dimensión	Indicadores
Método estadístico que utiliza series temporales para construir modelos que pronostican la demanda en función al valor futuro de una o más variables independientes.	Número de variables independientes.	Número de variables independientes usadas para construir los modelos de regresión lineal.
	Extensión de las series temporales.	Número de años que componen las series temporales usadas para construir los modelos de regresión lineal.
	Modelo validado.	Modelo que no viola los supuestos del análisis de regresión lineal con series temporales y que es significativo.
Y: Precisión del pronóstico de demanda		
Definición operacional	Dimensión	Indicadores
Porcentaje en que la demanda pronosticada se desvía de la demanda real. Se mide por el Error Medio Absoluto Porcentual (EMAP).	1) Precisión muy buena	$EMAP \leq 10\%$
	2) Precisión buena	$10\% < EMAP \leq 20\%$
	3) Precisión moderada	$20\% < EMAP \leq 30\%$
	4) Precisión pobre	$EMAP > 30\%$

### 3.4 Instrumentos

Los instrumentos son los medios que se utilizan para recopilar información acerca de las variables de estudio (Hernández Sampieri y Mendoza, 2018). Esta investigación no utilizó ningún instrumento, debido a que la información requerida se recopiló directamente de las páginas web de fuentes secundarias oficiales, tales como el Banco Central de Reserva del Perú (BCRP), el Instituto Nacional de Estadística e Informática (INEI), el Ministerio de la

Producción (PRODUCE), el Ministerio de Economía y Finanzas (MEF) y la Comisión de Promoción del Perú para la Exportación y el Turismo (PROMPERÚ).

### **3.5 Procedimientos**

El procedimiento consta de un número limitado y secuencial de actividades que, de forma eficaz y relativamente eficiente, hace posible la realización de tareas iterativas (Caballero, 2014). El procedimiento practicado en la investigación fue el siguiente:

#### ***3.5.1 Construcción, Verificación, Corrección y Significancia de los Modelos Iniciales***

1. Inicialmente se construyeron modelos de regresión simple utilizando toda la información disponible (1980-2018). Es decir, se elaboraron modelos teniendo como variable dependiente a la demanda de congelados (CONG) y como variables independientes a POB, PBI, PBIpes, PBIman, DI, IP, INF1 y Time; una a la vez. Se incluyó la variable de tendencia Time (1, 2, 3, 4, ...n), debido a que es práctica común utilizarla como variable independiente.
2. Los modelos se ordenaron de mayor a menor, según el  $r^2$ .
3. Al modelo con mayor  $r^2$ , se le aplicó las pruebas para verificar la no violación de los supuestos MCO. Si fue necesario, se practicó la medida remedial respectiva para intentar corregir el modelo.
4. Se repitió el paso 3 con el modelo que obtuvo el segundo mayor  $r^2$ , luego con el siguiente hasta agotar los modelos simples.
5. A cada modelo que no haya violado ningún supuesto o que haya podido ser corregido, se le aplicaron las pruebas de significancia, tanto global como individual. A cada modelo que resultó significativo se le denominó modelo validado.
6. Cada modelo validado fue replicado, pero con la data del periodo 1980-2013. A su vez, estos modelos también fueron validados. La información del periodo 2014-2018 se reservó para calcular la precisión (EMAP) de los modelos.

### ***3.5.2 Extensión de las Series Temporales y Precisión del Pronóstico***

1. Se volvieron a construir los modelos 1980-2013 validados en el paso 6 del numeral 3.6.1, pero para el periodo 1981-2013, después para los años 1982-2013, y así sucesivamente. Es decir, cada vez se disminuyó en un año la extensión de las series temporales.
2. Luego, en cada modelo, se verificó la no violación de los supuestos MCO. Seguidamente, se aplicó la medida correctiva correspondiente a todo modelo que haya violado algún supuesto.
3. A cada modelo que no haya violado ningún supuesto o que haya podido ser corregido, se le aplicaron las pruebas de significancia, tanto global como individual, para validarlos.
4. La parte reservada de las series temporales (2014-2018), se utilizó para realizar el pronóstico fuera de la muestra y calcular el EMAP y el sesgo, únicamente en los modelos validados.
5. Se calculó el coeficiente de correlación ( $r$ ) y se verificó su significancia, con la finalidad de establecer la relación entre la extensión de las series temporales y la precisión del pronóstico de demanda.

### ***3.5.3 Número de Variables Independientes y Precisión del Pronóstico***

1. De todos los modelos validados en el paso tres del numeral 3.6.2, se hizo una selección bajo dos criterios. Primero, se eligió a los modelos simples y a sus versiones aumentadas. Segundo, ambos modelos, simples y aumentados, tenían que haberse construido con series temporales de la misma extensión. No fue necesario calcular el EMAP de cada modelo, debido a que se ya había calculado en el paso 4 del numeral 3.6.2.
2. Se calculó el coeficiente de correlación ( $r$ ) y se verificó su significancia, con la finalidad de establecer la relación entre el número de variables independientes y la precisión del pronóstico de demanda.

### ***3.5.4 Modelo de Regresión Lineal Validado y Precisión del Pronóstico***

1. Con la información del periodo 1980-2013, se construyeron las versiones no validadas MCO de los modelos validados en el paso 6 del numeral 3.6.1.
2. Con la segunda parte de las series temporales (2014-2018) y, para los modelos no validados (1980-2013), se realizó el pronóstico fuera de la muestra y se calculó el EMAP y el sesgo. El EMAP de los modelos validados ya había calculado en el paso 4 del numeral 3.6.2.
3. Se llevó acabo la prueba t de diferencia de medias para datos pareados, con la finalidad de establecer si la precisión del pronóstico de demanda, de los modelos de regresión lineal validados, es mayor a la precisión del pronóstico de demanda de los modelos de regresión lineal no validados.

### **3.6 Análisis de Datos**

La construcción de los modelos de regresión lineal, así como su verificación, validación y, en caso de ser necesario, corrección; así como también los respectivos pronósticos, se realizaron con el software econométrico Gretl versión 2019a. El cálculo del EMAP y del sesgo se hizo en la hoja de cálculo Excel.

## IV Resultados

### 4.1 Contrastación de Hipótesis

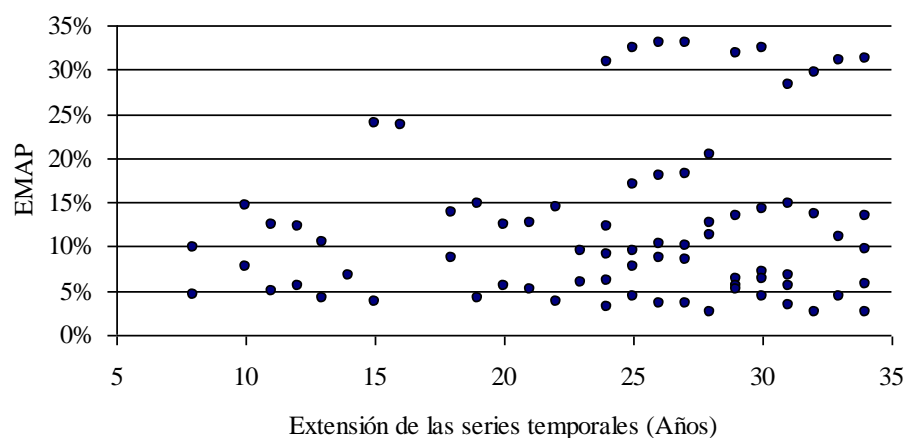
#### 4.1.1 Hipótesis Específica 1 ( $H_1$ )

Existe una relación positiva entre la extensión de las series temporales y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

La Figura 48 muestra el diagrama de dispersión entre la extensión de las series temporales y la precisión del pronóstico (EMAP). El diagrama consta de 76 pares de datos obtenidos de los modelos de regresión lineal (Anexo 3).

#### Figura 48

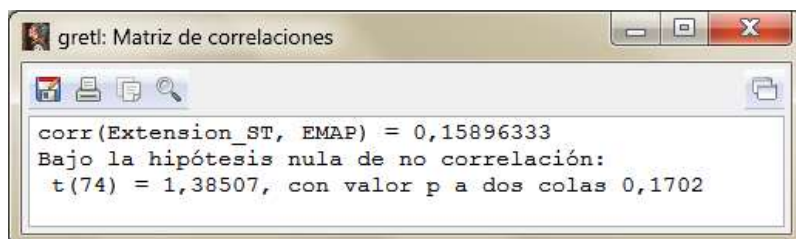
*Diagrama de Dispersión Entre la Extensión de las Series Temporales y la Precisión del Pronóstico de Demanda (EMAP)*



A continuación, la Figura 49 muestra que el coeficiente de correlación ( $r$ ) es aproximadamente 0,16 y su valor  $p$  es 0,1702, el cual es mayor al  $\alpha$  de 0,05, por tanto no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de que la correlación entre ambas variables es cero. Es decir, la extensión de las series temporales no se relaciona con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

### Figura 49

*Prueba t de Correlación Entre la Extensión de las Series Temporales y la Precisión del Pronóstico de Demanda*



Nota. Fuente software Gretl 2019a.

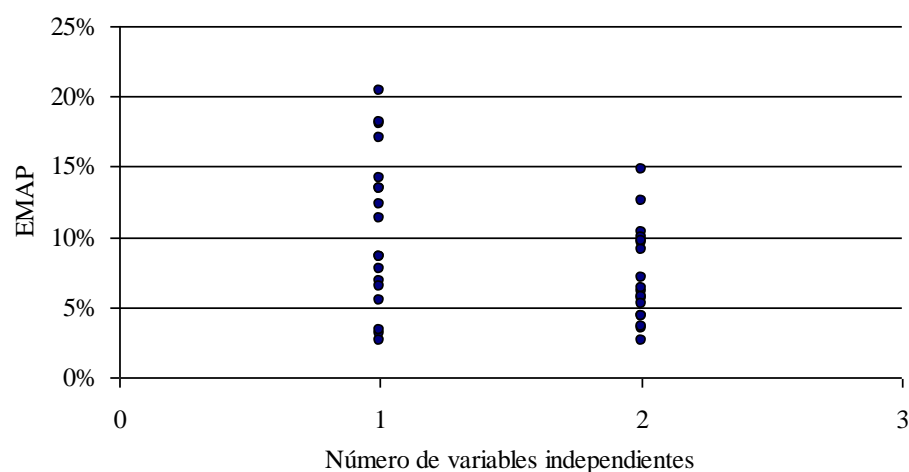
#### 4.1.2 Hipótesis Específica 2 ( $H_2$ )

El número de variables independientes se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

La Figura 50 muestra el diagrama de dispersión entre el número de variables independientes y la precisión del pronóstico (EMAP). El diagrama consta de 36 pares de datos obtenidos de los modelos de regresión lineal (Anexo 4).

### Figura 50

*Diagrama de Dispersión Entre el Número de Variables Independientes y la Precisión del Pronóstico de Demanda (EMAP)*

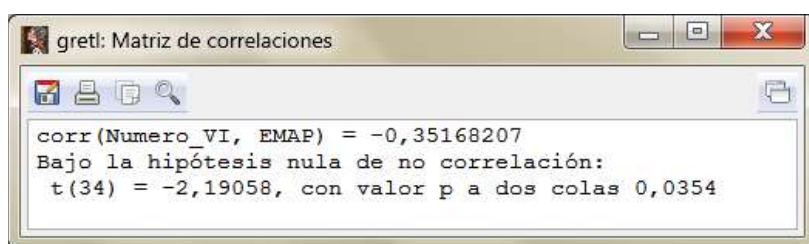


Seguidamente, la Figura 51 muestra que el  $r$  es aproximadamente  $-0,35$  y su valor  $p$  es  $0,0354$ , el cual es menor al  $\alpha$  de  $0,05$ , por tanto existe suficiente evidencia que justifica el

rechazo de la  $H_0$  de que la correlación entre ambas variables es cero. Es decir, el número de variables independientes se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

### Figura 51

*Prueba t de Correlación Entre el Número de Variables Independientes y la Precisión del Pronóstico de Demanda*



Nota. Fuente software Gretl 2019a.

#### 4.1.3 Hipótesis Específica 3 ( $H_3$ )

Un modelo de regresión lineal validado presenta mayor precisión en el pronóstico de demanda de productos hidrobiológicos congelados en el Perú, frente a un modelo de regresión lineal no validado.

La Tabla 13 presenta el EMAP (2014-2018) de los modelos no validados, de los modelos validados y el diferencial. El EMAP diferencial es el resultado de sustraer el EMAP de un modelo validado al EMAP de un modelo no validado.

**Tabla 13**

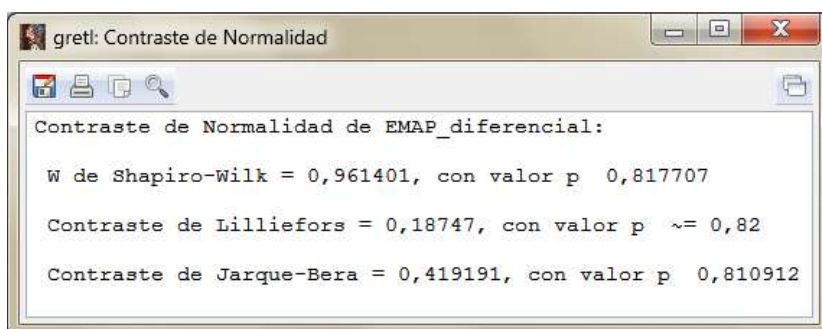
*EMAP (2014-2018) de Modelos No Validados, de Modelos Validados y Diferencial*

Modelo no validado	EMAP	Modelo validado	EMAP	EMAP diferencial
MCO: CONG vs. DI	19,32%	PW: L-CONG vs. DI	13,44%	5,88%
MCO: CONG vs. DI e INF1	16,98%	PW: L-CONG vs. DI e INF1	5,78%	11,20%
MCO: CONG vs. PBI	20,32%	PW: L-CONG vs. PBI	2,66%	17,66%
MCO: CONG vs. PBI e INF1	17,54%	PW: L-CONG vs. PBI e INF1	9,70%	7,84%
MCO: CONG vs. PBIman e INF1	34,22%	PW: L-CONG vs. PBIman e INF1	31,34%	2,88%

Dado que el tamaño muestral de la variable EMAP diferencial no es mayor a 30, es necesario probar la normalidad de los datos. En la Figura 52 se observa que el valor p de los tres contrastes superan al  $\alpha$  de 0,05, por tanto no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de que los datos son normales.

## Figura 52

### *Pruebas de Normalidad para el EMAP Diferencial*



*Nota.* Fuente software Gretl 2019a.

Seguidamente, se practica la prueba t de diferencia de medias para datos pareados. El estadístico de contraste es:

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

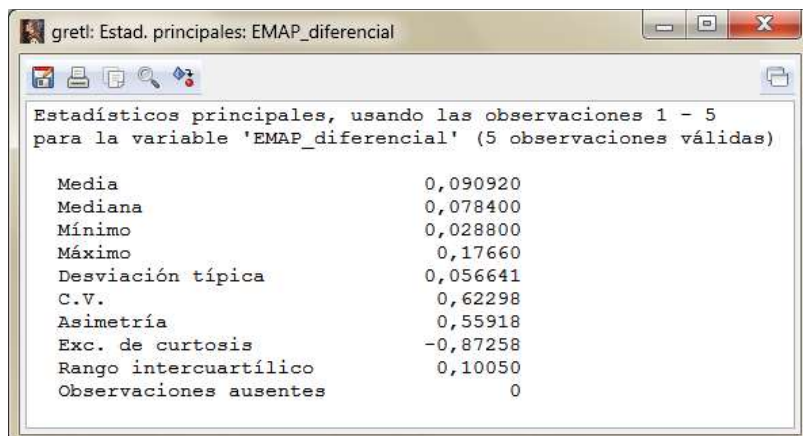
Donde:  $\bar{d}$  = Media diferencial

$s_d$  = Desviación típica del diferencial

$n$  = Número de pares

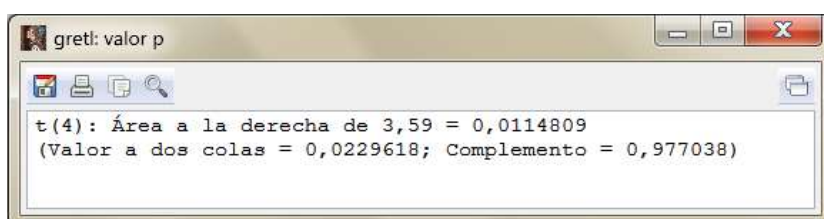
Según la fórmula, y de acuerdo a los estadísticos del EMAP diferencial (Figura 53), el valor del estadístico de contraste es:

$$t = \frac{0,09092}{\frac{0,056641}{\sqrt{5}}} = 3,59$$

**Figura 53***Estadísticos Principales del EMAP Diferencial*

*Nota.* Fuente software Gretl 2019a.

En la Figura 54 se aprecia que el valor p del estadístico de contraste ( $t = 3,59$ ) es 0,02, el cual es menor al  $\alpha$  de 0,05, por tanto existe suficiente evidencia que justifica el rechazo de la  $H_0$  de medias diferenciales iguales. Es decir, un modelo validado presenta mayor precisión en el pronóstico de demanda de productos hidrobiológicos congelados en el Perú, frente a un modelo no validado.

**Figura 54***Prueba t de Diferencia de EMAP Medio Entre Modelos Validados y Modelos No Validados*

*Nota.* Fuente software Gretl 2019a.

**4.1.4 Hipótesis General**

El análisis de regresión lineal se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

Se ha sometido a contrastación tres aspectos del análisis de regresión lineal, que pueden relacionarse con la precisión del pronóstico de demanda de productos hidrobiológicos

congelados en el Perú. Los aspectos estudiados (hipótesis específicas) fueron la extensión de las series temporales ( $H_1$ ), el número de variables independientes ( $H_2$ ) y la validación del modelo ( $H_3$ ).

De acuerdo a los resultados de la contrastación de hipótesis, la extensión de las series temporales ( $H_1$ ) no se relaciona con la precisión del pronóstico de demanda. En cambio, el número de variables independientes ( $H_2$ ) se relaciona positivamente con la precisión del pronóstico de demanda y los modelos validados ( $H_3$ ) presentan mayor precisión en el pronóstico de demanda que los modelos no validados.

Por tanto, se colige que el análisis de regresión lineal se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

## **4.2 Análisis e Interpretación**

### ***4.2.1 Construcción, Verificación, Corrección y Significancia de los Modelos Iniciales***

**4.2.1.1 Construcción de los Modelos 1980-2018.** De acuerdo al procedimiento descrito en el numeral 3.5.1 (p. 106) se construyeron ocho modelos de regresión simple, la variable dependiente fue CONG y las variables independientes fueron POB, PBI, PBIpes, PBIman, DI, IP, INF1 y Time; respectivamente. El significado de los acrónimos se presentó en el numeral 3.2.2 (p. 104).

La Tabla 14 presenta los ocho modelos, ordenados de mayor a menor según el  $r^2$ . En la tabla se aprecia que el modelo con variable independiente PBI presentó el más alto poder explicativo, es decir, la variable independiente PBI explicó en un 87,85% el comportamiento de la variable dependiente CONG.

**Tabla 14**

*Modelos de Regresión Lineal Simple Ordenados Según  $r^2$  (1980-2018)*

N°	Método	Variable		$r^2$
		dependiente	independiente	
1	MCO	CONG	PBI	87,85%
2	MCO	CONG	DI	87,04%
3	MCO	CONG	IP	79,56%
4	MCO	CONG	PBIman	78,16%
5	MCO	CONG	Time	67,51%
6	MCO	CONG	POB	62,36%
7	MCO	CONG	PBIpes	46,07%
8	MCO	CONG	INF1	0,29%

La Figura 55 exhibe el modelo con más alto  $r^2$  (Tabla 14), es decir, el modelo MCO: CONG vs. PBI (1980-2018), tal y como lo presenta el software Gretl. La figura no solo muestra los coeficientes del modelo, sino también información relevante como  $r_c^2$  y el  $d$  de DW. Adicionalmente, también se presentan los valores  $p$  de las pruebas de significancia  $t$  y  $F$ .

**Figura 55**

*Modelo MCO: CONG vs. PBI (1980-2018)*

	Coefficiente	Desv. típica	Estadístico t	valor p
const	-34,4134	5,07375	-6,783	5,53e-08 ***
PBI	0,000281000	1,71780e-05	16,36	1,60e-018 ***

Media de la vble. dep.	41,48359	D.T. de la vble. dep.	36,30359
Suma de cuad. residuos	6083,748	D.T. de la regresión	12,82286
R-cuadrado	0,878525	R-cuadrado corregido	0,875241
F(1, 37)	267,5883	Valor p (de F)	1,60e-18
Log-verosimilitud	-153,8100	Criterio de Akaike	311,6200
Criterio de Schwarz	314,9471	Crit. de Hannan-Quinn	312,8137
rho	0,500027	Durbin-Watson	0,999521

Nota. Fuente Software Gretl 2019a.

#### 4.2.1.2 Verificación del Modelo MCO: CONG vs. PBI (1980-2018)

**Verificación del Supuesto 1: El Modelo Es Lineal.** Los valores  $p$  de las pruebas LM de linealidad son menores al nivel de significancia ( $\alpha$ ) de 0,05, por lo cual existe suficiente evidencia que justifica el rechazo de la  $H_0$  de relación lineal, por tanto el modelo viola el

supuesto de linealidad. Por un lado, en la Figura 56 se muestra que el valor p, de la versión cuadrados de la prueba LM, es 0,00052 ( $\leq 0,05$ ); lo cual implica que la relación no es lineal. Por otro lado, la Figura 57 muestra que el valor p de la versión logaritmos de la prueba LM es 0,00049 ( $\leq 0,05$ ), en consecuencia la relación no es lineal.

**Figura 56**

*Modelo MCO: CONG vs. PBI (1980-2018). Prueba LM de Linealidad (Cuadrados)*

	Coeficiente	Desv. típica	Estadístico t	valor p
const	56,6264	14,7592	3,837	0,0005 ***
PBI	-0,000416723	0,000104959	-3,970	0,0003 ***
sq_PBI	6,41096e-010	1,59927e-010	4,009	0,0003 ***

R-cuadrado = 0,308617

Estadístico de contraste:  $TR^2 = 12,0361$ ,  
con valor p =  $P(\text{Chi-cuadrado}(1) > 12,0361) = 0,000521807$

Nota. Fuente Software Gretl 2019a.

**Figura 57**

*Modelo MCO: CONG vs. PBI (1980-2018). Prueba LM de Linealidad (Logaritmos)*

	Coeficiente	Desv. típica	Estadístico t	valor p
const	1312,81	324,951	4,040	0,0003 ***
PBI	0,000382737	9,58234e-05	3,994	0,0003 ***
l_PBI	-114,031	28,2229	-4,040	0,0003 ***

R-cuadrado = 0,311986

Estadístico de contraste:  $TR^2 = 12,1674$ ,  
con valor p =  $P(\text{Chi-cuadrado}(1) > 12,1674) = 0,000486308$

Nota. Fuente Software Gretl 2019a.

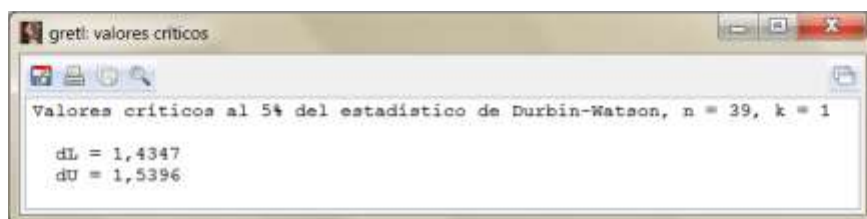
**Verificación del Supuesto 2: No Existe Colinealidad Exacta.** Por definición, la verificación de este supuesto solo es posible en modelos de regresión lineal múltiple, es decir, en aquellos modelos que cuentan con más de una variable independiente.

**Verificación del Supuesto 3: El Promedio de los Errores Es Cero.** Para verificar el cumplimiento de este supuesto basta la inclusión del intercepto en el modelo de regresión.

**Verificación del Supuesto 4: Los Errores No Están Autocorrelacionados.** El estadístico  $d$  de DW del modelo MCO es 0,9995 (Figura 55, p. 115) y los valores  $d_L$  y  $d_U$  son 1,4347 y 1,5396, respectivamente (Figura 58).

### Figura 58

Valores de  $d_L$  y  $d_U$  para la Prueba  $d$  de DW de Autocorrelación ( $n=39$ ,  $k=1$ )

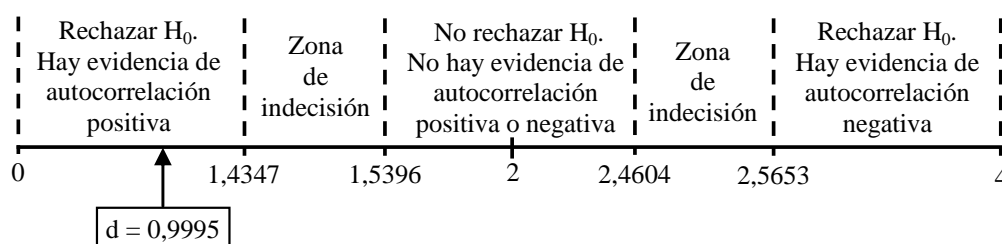


Nota. Fuente Software Gretl 2019a.

Con lo cual, al establecer los límites de las zonas de la prueba DW, el estadístico  $d$  cae en la zona delimitada por cero y 1,4347; por lo cual existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores no autocorrelacionados (Figura 59), por tanto el modelo viola el supuesto de no autocorrelación.

### Figura 59

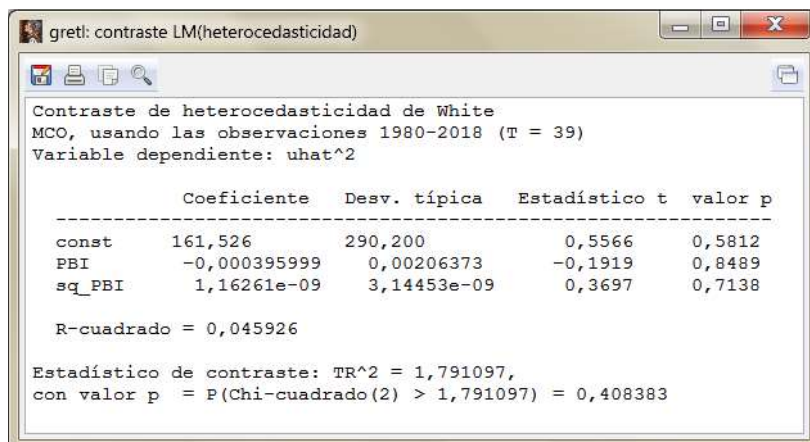
Modelo MCO: CONG vs. PBI (1980-2018). Prueba  $d$  de DW de Autocorrelación



**Verificación del Supuesto 5: Los Errores Son Homocedásticos.** El valor  $p$  de la prueba de heterocedasticidad de White es 0,4084 (Figura 60), mayor al  $\alpha$  de 0,05, por lo cual no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores homocedásticos, por tanto el modelo no viola el supuesto de homocedasticidad.

**Figura 60**

*Modelo MCO: CONG vs. PBI (1980-2018). Prueba de Heterocedasticidad de White*

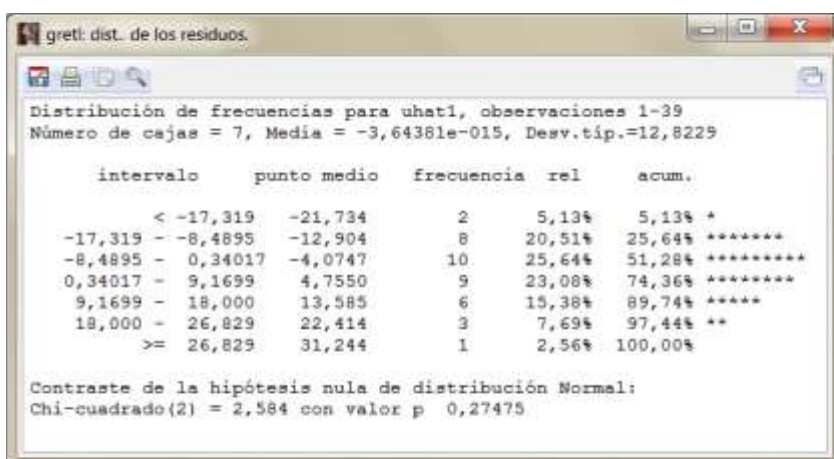


Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 6: Los Errores Se Distribuyen Normalmente.** El valor p de la prueba de normalidad de DH es 0,27475 (Figura 61), es superior al  $\alpha$  de 0,05 y, en consecuencia, no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores normalmente distribuidos, por tanto el modelo no viola el supuesto de normalidad.

**Figura 61**

*Modelo MCO: CONG vs. PBI (1980-2018). Prueba de Normalidad de DH.*



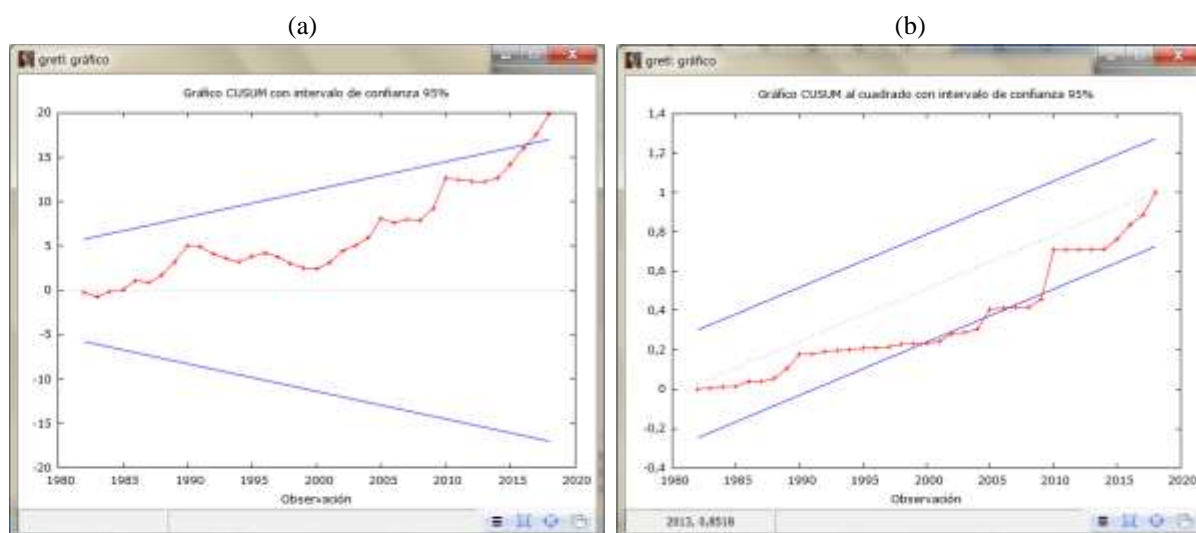
Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 7: Ausencia de Cambio Estructural.** Las gráficas de las pruebas CUSUM (Figura 62a) y CUSUMQ (Figura 62b) exceden las bandas de confianza,

por lo cual existe suficiente evidencia que justifica el rechazo de la  $H_0$  de ausencia de cambio estructural y, por tanto, el modelo viola el supuesto de estabilidad estructural.

### Figura 62

*Modelo MCO: CONG vs. PBI (1980-2018). Pruebas Gráficas de Cambio Estructural (a) CUSUM y (b) CUSUMQ*



Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 8: Especificación Correcta.** El valor p de la prueba de especificación RR es 0,00156 (Figura 63), es inferior al  $\alpha$  de 0,05, por lo tanto existe suficiente evidencia que justifica el rechazo de la  $H_0$  de especificación correcta, por tanto el modelo viola el supuesto de especificación correcta.

### Figura 63

*Modelo MCO: CONG vs. PBI (1980-2018). Prueba de Especificación RR*

	Coefficiente	Desv. típica	Estadístico t	valor p
const	14,2051	29,0621	0,4888	0,6280
PBI	1,12925e-05	0,000176294	0,06405	0,9493
yhat^2	0,00888125	0,0125722	0,7064	0,4846
yhat^3	-4,35779e-06	7,09226e-05	-0,06144	0,9514

Estadístico de contraste:  $F = 7,814328$ ,  
con valor p =  $P(F(2,35) > 7,81433) = 0,00156$

Nota. Fuente Software Gretl 2019a.

**4.2.1.3 Corrección del Modelo MCO: CONG vs. PBI (1980-2018).** Luego de la verificación del modelo MCO: CONG vs. PBI (Figura 55, p. 115), se concluye que éste viola los supuestos de linealidad, de no autocorrelación, de ausencia de cambio estructural y de especificación correcta. La medida remedial fue la inclusión de la variable INF1 para intentar mejorar la especificación del modelo. Es decir, el modelo de regresión lineal simple original fue ampliado a un modelo de regresión lineal múltiple (Figura 64).

#### Figura 64

*Modelo MCO: CONG vs. PBI e INF1 (1980-2018)*

	Coefficiente	Desv. típica	Estadístico t	valor p	
const	-39,8477	4,65217	-8,565	3,28e-010	***
PBI	0,000294597	1,53744e-05	19,16	1,80e-020	***
INF1	0,00516454	0,00142589	3,622	0,0009	***

Media de la vble. dep.	41,48359	D.T. de la vble. dep.	36,30359
Suma de cuad. residuos	4458,892	D.T. de la regresión	11,12916
R-cuadrado	0,910968	R-cuadrado corregido	0,906022
F(2, 36)	184,1754	Valor p (de F)	1,24e-19
Log-verosimilitud	-147,7509	Criterio de Akaike	301,3019
Criterio de Schwarz	306,4926	Crit. de Hannan-Quinn	303,2925
rho	0,316944	Durbin-Watson	1,320567

Nota. Fuente Software Gretl 2019a.

#### 4.2.1.4 Verificación del Modelo MCO: CONG vs. PBI e INF1 (1980-2018)

**Verificación del Supuesto 1: El Modelo Es Lineal.** Los valores p de ambas pruebas LM de linealidad son menores al  $\alpha$  de 0,05, por lo cual existe suficiente evidencia que justifica el rechazo de la  $H_0$  de relación lineal, por tanto el modelo viola el supuesto de linealidad. Por un lado, en la Figura 65 se aprecia que el valor p de la versión cuadrados de la prueba LM, es 0,0047 ( $\leq 0,05$ ); por tanto la relación no es lineal. Por otro lado, la Figura 66 muestra que el valor p de la versión logaritmos de la prueba LM, es 0,0104 ( $\leq 0,05$ ), por tanto la relación no es lineal.

**Figura 65**

*Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Cuadrados)*

	Coeficiente	Desv. típica	Estadístico t	valor p	
const	44,5057	14,4891	3,072	0,0042	***
PBI	-0,000326698	0,000100382	-3,255	0,0026	***
INF1	0,00293595	0,00483436	0,6073	0,5477	
sq_PBI	5,01610e-010	1,51011e-010	3,322	0,0021	***
sq_INF1	-5,78207e-07	6,53875e-07	-0,8843	0,3828	

R-cuadrado = 0,275027

Estadístico de contraste:  $TR^2 = 10,7261$ ,  
con valor p =  $P(\text{Chi-cuadrado}(2) > 10,7261) = 0,00468667$

Nota. Fuente Software Gretl 2019a.

**Figura 66**

*Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Logaritmos)*

	Coeficiente	Desv. típica	Estadístico t	valor p	
const	1081,95	443,820	2,438	0,0202	**
PBI	0,000308097	0,000117179	2,629	0,0128	**
INF1	-0,00120816	0,00180523	-0,6693	0,5079	
l_PBI	-93,7166	37,9920	-2,467	0,0188	**
l_INF1	-0,305491	1,76707	-0,1725	0,8638	

R-cuadrado = 0,233988

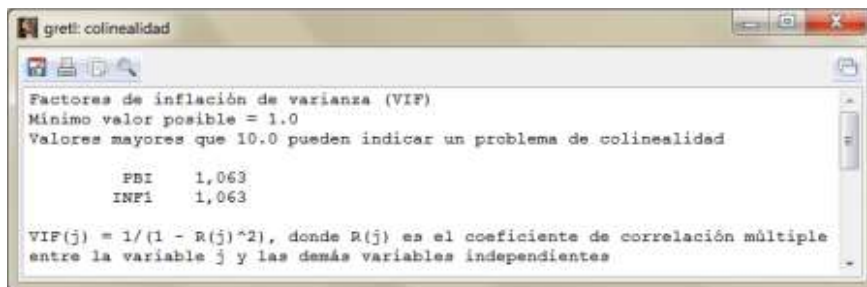
Estadístico de contraste:  $TR^2 = 9,12554$ ,  
con valor p =  $P(\text{Chi-cuadrado}(2) > 9,12554) = 0,0104331$

Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 2: No Existe Colinealidad Exacta.** Los FIVs de las variables independientes son menores a 10. El FIV, tanto de PBI como de INF1, es 1,063 (Figura 67); en consecuencia el modelo no viola el supuesto de inexistencia de colinealidad exacta.

### Figura 67

Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Colinealidad: FIV



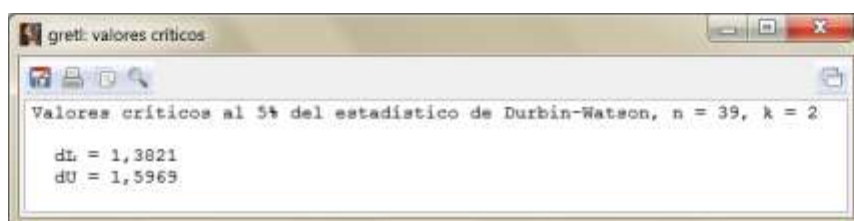
Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 3: El Promedio de los Errores Es Cero.** Para verificar el cumplimiento de este supuesto basta la inclusión del intercepto en el modelo de regresión.

**Verificación del Supuesto 4: Los Errores No Están Autocorrelacionados.** El estadístico  $d$  de DW del modelo MCO es 1,3206 (Figura 64, p. 120) y los valores  $d_L$  y  $d_U$  son 1,3821 y 1,5969, respectivamente (Figura 68).

### Figura 68

Valores de  $d_L$  y  $d_U$  para la Prueba  $d$  de DW de Autocorrelación ( $n=39$ ,  $k=2$ )

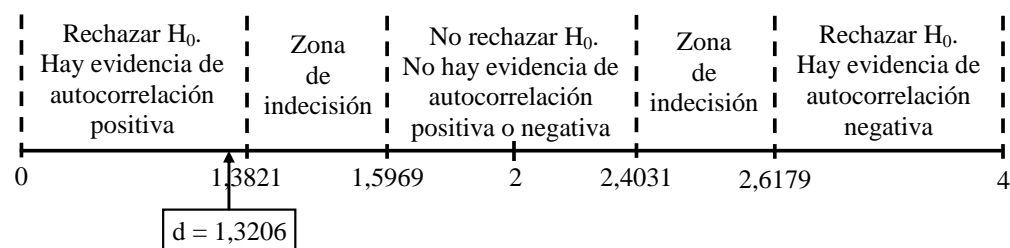


Nota. Fuente Software Gretl 2019a.

Y, al establecer las zonas de la prueba DW, el  $d$  cae entre cero y 1,3821, por lo cual hay suficiente evidencia que justifica el rechazo de la  $H_0$  de errores no autocorrelacionados (Figura 69), por tanto el modelo viola el supuesto de no autocorrelación.

**Figura 69**

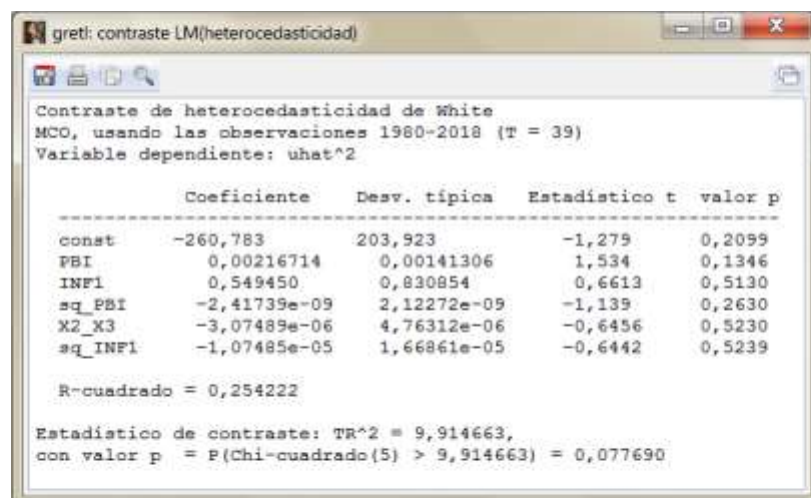
Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba d de DW de Autocorrelación



**Verificación del Supuesto 5: Los Errores Son Homocedásticos.** El valor p de la prueba de heterocedasticidad de White es 0,0777 (Figura 70), es mayor al  $\alpha$  de 0,05, por lo cual no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores homocedásticos, por tanto el modelo no viola el supuesto de homocedasticidad.

**Figura 70**

Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba de Heterocedasticidad de White

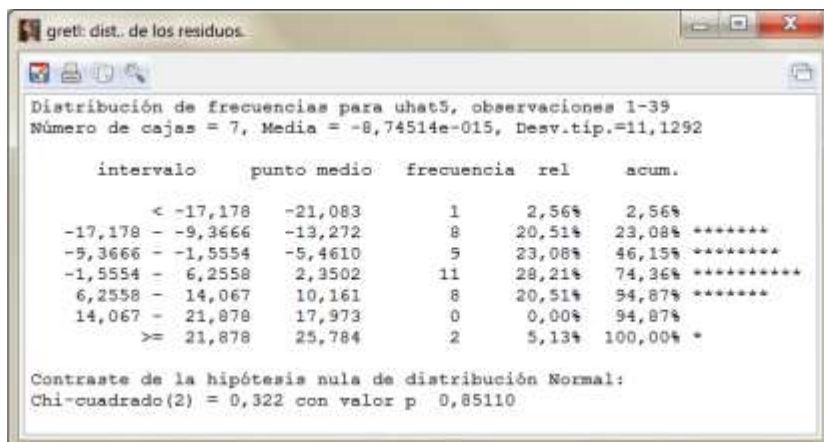


Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 6: Los Errores Se Distribuyen Normalmente.** El valor p de la prueba de normalidad de DH es 0,8511 (Figura 71), éste es superior al  $\alpha$  de 0,05, por lo tanto no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores normalmente distribuidos, es decir, el modelo no viola el supuesto de normalidad.

**Figura 71**

*Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba de Normalidad de DH.*



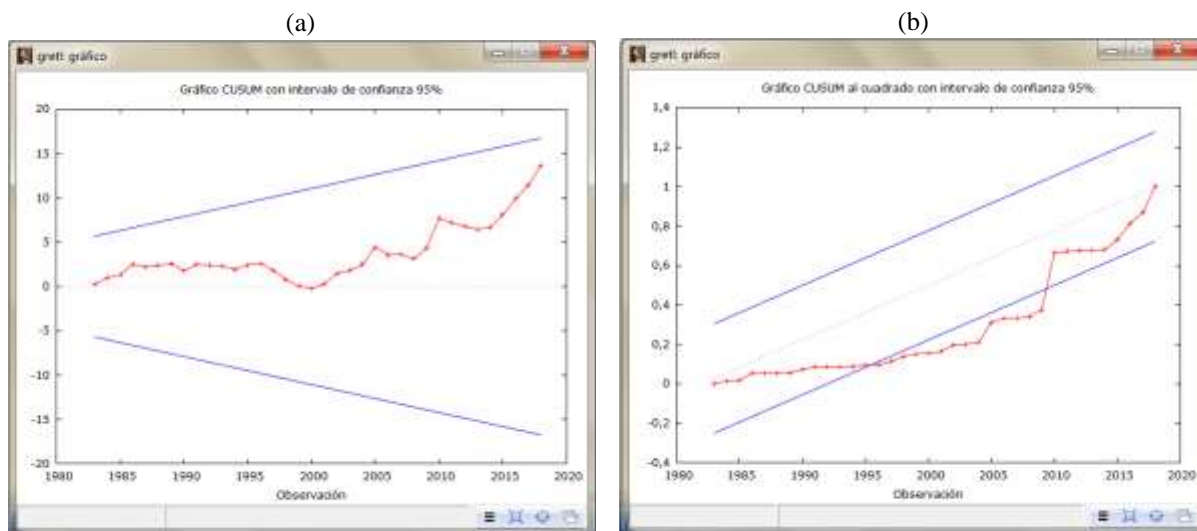
Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 7: Ausencia de Cambio Estructural.** La gráfica CUSUM (Figura 72a) indica que el modelo no viola el supuesto, pero la gráfica CUSUMQ (Figura 72b) sugiere lo opuesto. Sin embargo, la gráfica CUSUMQ es más potente, por tanto existe suficiente evidencia que justifica el rechazo de la  $H_0$  de ausencia de cambio estructural, en consecuencia, el modelo viola el supuesto de estabilidad estructural.

**Figura 72**

*Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Pruebas Gráficas de Cambio Estructural*

(a) CUSUM y (b) CUSUMQ



Nota. Fuente Software Gretl versión 2019a.

**Verificación del Supuesto 8: Especificación Correcta.** El valor p de la prueba de especificación RR es 0,0049 (Figura 73), es inferior al  $\alpha$  de 0,05, por tanto existe suficiente evidencia que justifica el rechazo de la  $H_0$  de especificación correcta, en consecuencia, el modelo viola el supuesto de especificación correcta.

**Figura 73**

*Modelo MCO: CONG vs. PBI e INF1 (1980-2018). Prueba de Especificación RR*

	Coeficiente	Desv. típica	Estadístico t	valor p
const	-23,1397	25,8798	-0,8941	0,3775
PBI	0,000218421	0,000148847	1,467	0,1515
INF1	0,00453594	0,00264023	1,718	0,0949 *
yhat^2	-0,00314785	0,0101895	-0,3089	0,7593
yhat^3	5,20006e-05	5,73635e-05	0,9065	0,3710

Estadístico de contraste: F = 6,248551,  
con valor p = P(F(2,34) > 6,24855) = 0,00489

*Nota.* Fuente Software Gretl 2019a.

**4.2.1.5 Corrección del Modelo MCO: CONG vs. PBI e INF1 (1980-2018).** Luego de verificar el modelo ampliado MCO: CONG vs. PBI e INF1 (Figura 64, p. 120), se colige que éste viola los supuestos de linealidad, de no autocorrelación, de ausencia de cambio estructural y el de especificación correcta. Dado que la ampliación del modelo no corrigió ninguna de las violaciones, se decidió cambiar la forma funcional a un modelo log-lineal. Es decir, la nueva especificación fue  $L\text{-CONG} = \beta_0 + \beta_1\text{PBI} + \beta_2\text{INF1} + \mu_t$  (Figura 74).

Figura 74

Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018)

	Coefficiente	Desv. típica	Estadístico t	valor p
const	1,54262	0,146841	10,51	1,64e-012 ***
PBI	6,57746e-06	4,85279e-07	13,55	1,04e-015 ***
INF1	0,000171951	4,50069e-05	3,821	0,0005 ***
Media de la vble. dep.	3,377823	D.T. de la vble. dep.	0,845207	
Suma de cuad. residuos	4,442350	D.T. de la regresión	0,351281	
R-cuadrado	0,836355	R-cuadrado corregido	0,827264	
F(2, 36)	91,99414	Valor p (de F)	7,08e-15	
Log-verosimilitud	-12,97723	Criterio de Akaike	31,95446	
Criterio de Schwarz	36,94515	Crit. de Hannan-Quinn	33,74508	
rho	0,413989	Durbin-Watson	1,171458	

Nota. Fuente Software Gretl 2019a.

#### 4.2.1.6 Verificación del Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018)

**Verificación del Supuesto 1: El Modelo Es Lineal.** Los valores p de las pruebas LM de linealidad son mayores al  $\alpha$  de 0,05, por lo cual no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de relación lineal, por tanto el modelo no viola el supuesto de linealidad. Por un lado, en la Figura 75 se observa que el valor p de la versión cuadrados de la prueba LM, es 0,08844 ( $> 0,05$ ); con lo cual la relación es lineal.

Figura 75

Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba LM de Linealidad (Cuadrados)

	Coefficiente	Desv. típica	Estadístico t	valor p
const	-0,875072	0,502610	-1,741	0,0907 *
PBI	6,02027e-06	3,48214e-06	1,729	0,0929 *
INF1	0,000277610	0,000167699	1,655	0,1070
sq_PBI	-8,95029e-012	5,23841e-012	-1,709	0,0966 *
sq_INF1	-3,64947e-08	2,26822e-08	-1,609	0,1169

R-cuadrado = 0,124380

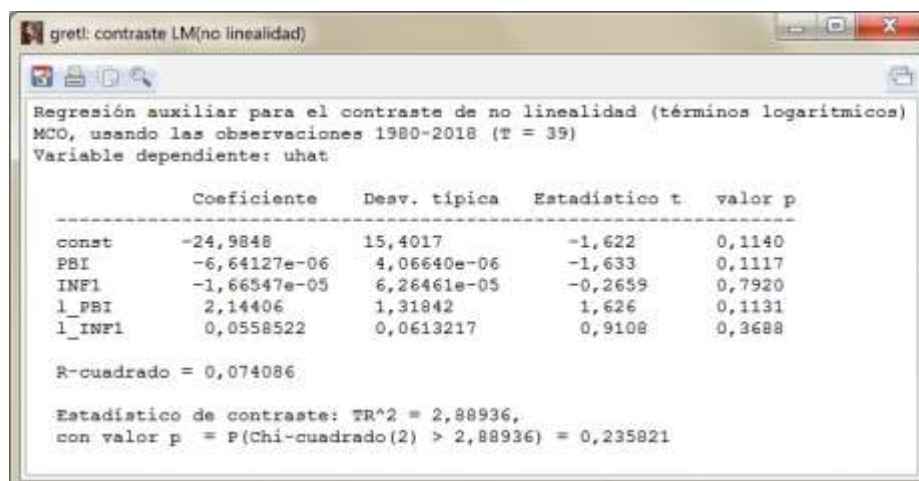
Estadístico de contraste:  $TR^2 = 4,85081$ ,  
con valor p =  $P(\text{Chi-cuadrado}(2) > 4,85081) = 0,0884421$

Nota. Fuente Software Gretl 2019a.

Por otro lado, la Figura 76 muestra que el valor p de la versión logaritmos de la prueba LM, es 0,2358 ( $> 0,05$ ), por tanto la relación es lineal.

### Figura 76

*Modelo MCO: L-CONG vs. PBI e INF1(1980-2018). Prueba LM de Linealidad (Logaritmos)*

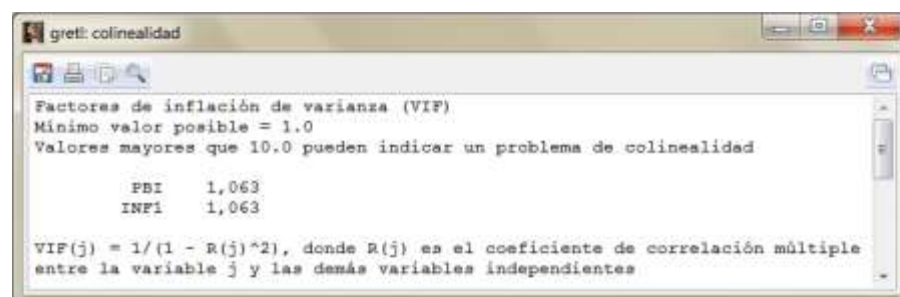


Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 2: No Existe Colinealidad Exacta.** Los FIVs de las variables independientes son menores a 10. El FIV de ambas variables, tanto PBI como INF1, es 1,063 (Figura 77); por lo tanto el modelo no viola el supuesto de inexistencia de colinealidad exacta.

### Figura 77

*Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Colinealidad: FIV*



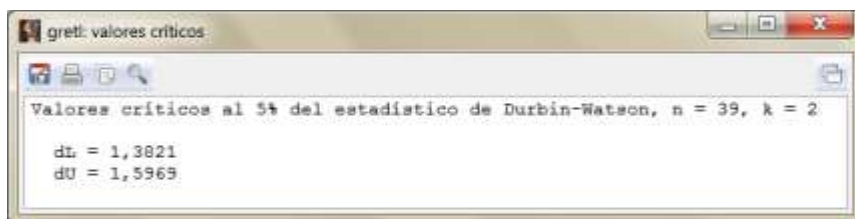
Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 3: El Promedio de los Errores Es Cero.** Para verificar el cumplimiento de este supuesto basta la inclusión del intercepto en el modelo de regresión.

**Verificación del Supuesto 4: Los Errores No Están Autocorrelacionados.** El estadístico  $d$  de DW del modelo MCO es 1,1715 (Figura 74, p. 126) y los valores  $d_L$  y  $d_U$  son 1,3821 y 1,5969, respectivamente (Figura 78).

### Figura 78

Valores de  $d_L$  y  $d_U$  para la Prueba  $d$  de DW de Autocorrelación ( $n=39$  y  $k=2$ )

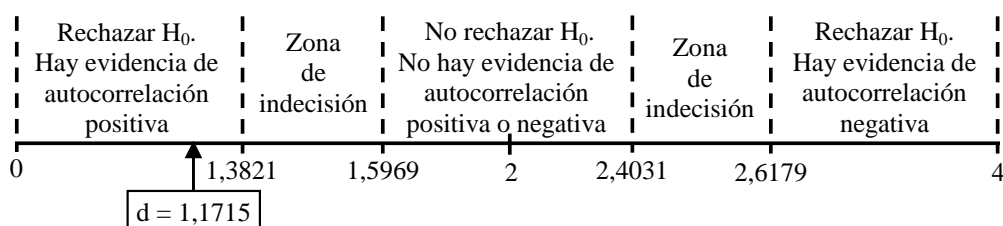


Nota. Fuente Software Gretl 2019a.

Con lo cual, al establecer los límites de las zonas de la prueba DW, el estadístico  $d$  cae en la zona delimitada por cero y 1,3821, por tanto existe suficiente evidencia que justifica el rechazo de la  $H_0$  de errores no autocorrelacionados (Figura 79), en consecuencia, el modelo viola el supuesto de no autocorrelación.

### Figura 79

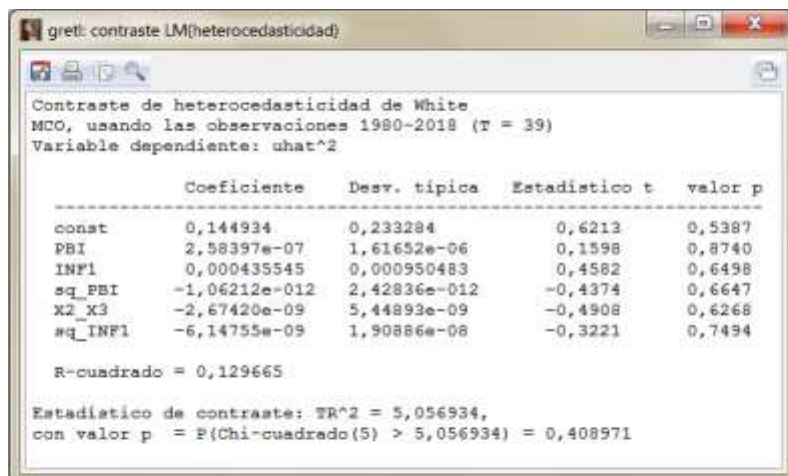
Modelo MCO: *L-CONG* vs. *PBI* e *INF1* (1980-2018). Prueba  $d$  de DW de Autocorrelación



**Verificación del Supuesto 5: Los Errores Son Homocedásticos.** El valor  $p$  de la prueba de heterocedasticidad de White es 0,4090 (Figura 80), es mayor al  $\alpha$  de 0,05, por lo tanto no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores homocedásticos y, en consecuencia, el modelo no viola el supuesto de homocedasticidad.

**Figura 80**

*Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba de Heterocedasticidad de White*

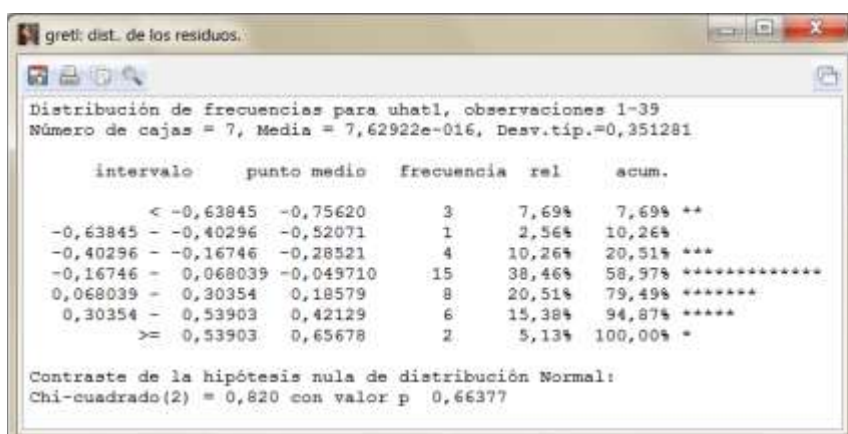


Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 6: Los Errores Se Distribuyen Normalmente.** El valor p de la prueba de normalidad de DH es 0,66377 (Figura 81), es superior al  $\alpha$  de 0,05 y, en consecuencia, no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores normalmente distribuidos, por tanto el modelo no viola el supuesto de normalidad.

**Figura 81**

*Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba de Normalidad de DH*



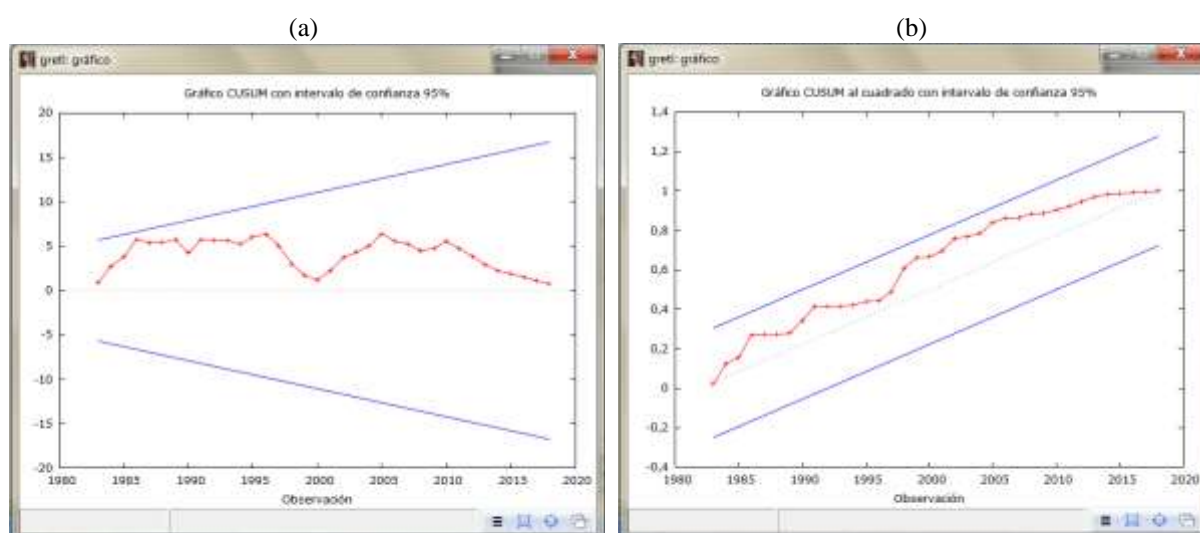
Nota. Fuente Software Gretl 2019a.

**Verificación del Supuesto 7: Ausencia de Cambio Estructural.** Las gráficas CUSUM (Figura 82a) y CUSUMQ (Figura 82b) no superan las bandas de confianza, por tanto no hay

suficiente evidencia que justifique el rechazo de la  $H_0$  de ausencia de cambio estructural y, en consecuencia, el modelo no viola el supuesto de estabilidad estructural.

### Figura 82

*Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Pruebas Gráficas de Cambio Estructural (a) CUSUM y (b) CUSUMQ*



*Nota.* Fuente Software Gretl versión 2019a.

**Verificación del Supuesto 8: Especificación Correcta.** El valor p de la prueba de especificación RR es 0,252 (Figura 83), es superior al  $\alpha$  de 0,05, por lo tanto no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de especificación correcta y, en consecuencia, el modelo no viola el supuesto de especificación correcta.

### Figura 83

*Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018). Prueba de Especificación RR*

	Coeficiente	Desv. típica	Estadístico t	valor p
const	2,31342	2,65086	0,8727	0,3889
PBI	3,65852e-05	4,86297e-05	0,7523	0,4570
INF1	0,000938141	0,00127173	0,7377	0,4658
yhat^2	-1,04082	2,01845	-0,5157	0,6094
yhat^3	0,0746136	0,179774	0,4150	0,6807

Estadístico de contraste:  $F = 1,437871$ ,  
con valor p =  $P(F(2,34) > 1,43787) = 0,252$

*Nota.* Fuente Software Gretl 2019a.

#### 4.2.1.7 Corrección del Modelo MCO: L-CONG vs. PBI e INF1 (1980-2018).

Después de verificar el modelo log-lineal MCO: L-CONG vs. PBI e INF1 (Figura 74, p. 126), se colige que únicamente se viola el supuesto de no autocorrelación. La medida remedial es la modelación PW. El estadístico  $d$  de DW del modelo PW es 1,8948 (Figura 84) y los valores  $d_L$  y  $d_U$  son 1,3821 y 1,5969, respectivamente (Figura 68, p. 122).

#### Figura 84

Modelo PW: L-CONG vs. PBI e INF1 (1980-2018)

	Coefficiente	Desv. típica	Estadístico t	valor p
const	1,62420	0,235407	6,900	4,45e-08 ***
PBI	6,34002e-06	7,74408e-07	8,187	9,74e-010 ***
INF1	0,000112818	4,14094e-05	2,724	0,0099 ***

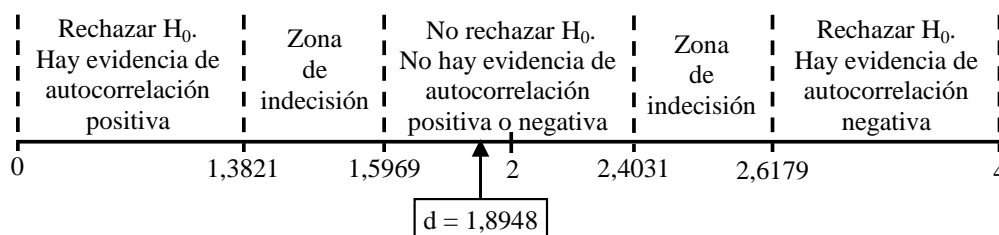
Estadísticos basados en los datos rho-diferenciados:			
Media de la vble. dep.	3,377823	D.T. de la vble. dep.	0,845207
Suma de cuad. residuos	3,513270	D.T. de la regresión	0,312395
R-cuadrado	0,870593	R-cuadrado corregido	0,863404
F(2, 36)	34,08303	Valor p (de F)	4,95e-09
rho	0,052350	Durbin-Watson	1,894763

Nota. Fuente Software Gretl 2019a.

Con lo cual, al establecer los límites de las zonas de la prueba DW, el estadístico  $d$  cae en la zona delimitada por 1,5969 y 2,4031, por tanto no existe suficiente evidencia que justifique el rechazo de la  $H_0$  de errores no autocorrelacionados (Figura 85) y, en consecuencia, el modelo no viola el supuesto de no autocorrelación.

#### Figura 85

Modelo PW: L-CONG vs. PBI e INF1 (1980-2018). Prueba  $d$  de DW de Autocorrelación



**4.2.1.8 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. PBI e INF1 (1980-2018).** Los valores p, tanto el de la prueba F como los de las pruebas t, son menores e iguales al  $\alpha$  de 0,05. Por un lado, el valor p de la prueba F es  $4,95 \times 10^{-9}$  (Figura 84, p. 131), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, globalmente, son iguales a cero. Por otro lado, los valores p de la prueba t son  $9,74 \times 10^{-10}$  (PBI) y 0,0099 (INF1) (Figura 84, p. 131), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, individualmente, son iguales a cero. Similarmente, el valor p de la constante es  $4,45 \times 10^{-8}$  (Figura 84, p. 131). En vista de los resultados, se concluye que el modelo es significativo y, en consecuencia, el modelo ha sido validado.

En lo referente a poder explicativo, el modelo tiene un  $R^2$  de 87,06% (Figura 84, p. 131). Es decir, el comportamiento de la variable dependiente L-CONG está siendo explicado en un 87,06% por las variables PBI e INF1. Por último, el modelo PW (Figura 84, p. 131) es:

$$L-CONG = 1,6242 + 0,00000634002 \text{ PBI} + 0,000112818 \text{ INF1}$$

Según el modelo, la demanda de congelados aumentará en 0,000634002% ( $0,00000634002 \times 100$ ) por cada millón de soles de incremento en el PBI, manteniendo constante la INF1. De igual manera, la demanda de congelados aumentará en 0,0112818% ( $0,000112818 \times 100$ ) por cada punto porcentual de incremento en INF1, manteniendo constante el PBI.

Adicionalmente se lograron validar otros tres modelos. La construcción, verificación, corrección y validación de los modelos se realizó según el procedimiento mostrado en el numeral 4.2.1 (p. 114). Las tablas siguientes también incluyen al modelo validado en dicho numeral. Antes, se codificaron los supuestos MCO (Tabla 15), de tal manera que en las tablas siguientes solo se consignó el código del supuesto o de los supuestos violados por los modelos.

**Tabla 15***Codificación de los Supuestos MCO*

Supuesto MCO	Código
El modelo es lineal	1
No existe multicolinealidad exacta	2
El promedio de los errores es cero	3
Los errores no están autocorrelacionados	4
Los errores son homocedásticos	5
Los errores se distribuyen normalmente	6
Ausencia de cambio estructural	7
Especificación correcta	8

Siguiendo con el procedimiento, primero se construyeron y verificaron los modelos MCO con variable dependiente CONG. Los modelos, aunque presentaron alto poder explicativo ( $R^2 \geq 70\%$ ), violaron entre tres y cuatro supuestos MCO (Tabla 16).

**Tabla 16***Verificación de los Modelos MCO con Variable Dependiente CONG (1980-2018)*

Nº	Método	Variable dependiente	Variabes independientes	$r^2$ o $R^2$	$r_c^2$ o $R_c^2$	Merma	FIV	d de DW	Supuestos violados
1	MCO	CONG	DI	87,04%	86,69%	0,35%	na	1,001	1, 4, 7, 8
2	MCO	CONG	DI - INF1	90,08%	89,53%	0,55%	1,061	1,265	4, 5, 7
3	MCO	CONG	PBIman - INF1	81,77%	80,76%	1,01%	1,078	0,787	1, 4, 5, 7
4	MCO	CONG	PBI - INF1	91,10%	90,60%	0,49%	1,063	1,321	1, 4, 7, 8

*Nota:* na = no aplica. Fuente Software Gretl 2019a.

Luego, los modelos fueron corregidos cambiando su forma funcional, se modificó la variable dependiente de CONG a L-CONG (Logaritmo natural de CONG). Los modelos corregidos solo violaron el supuesto de no autocorrelación (Tabla 17).

**Tabla 17**

*Verificación de los Modelos MCO con Variable Dependiente L-CONG (1980-2018)*

Nº	Método	Variable dependiente	Variabes independientes	$r^2$ o $R^2$	$r_c^2$ o $R_c^2$	Merma	FIV	d de DW	Supuestos violados
1	MCO	L-CONG	DI	74,58%	73,89%	0,69%	na	0,808	4
2	MCO	L-CONG	DI - INF1	80,81%	79,75%	1,07%	1,061	1,038	4
3	MCO	L-CONG	PBIman - INF1	84,00%	83,11%	0,89%	1,078	1,298	4
4	MCO	L-CONG	PBI - INF1	83,64%	82,73%	0,91%	1,063	1,171	4

*Nota:* na = no aplica. Fuente Software Gretl 2019a.

Finalmente, la Tabla 18 muestra que el problema de autocorrelación de los errores de los modelos MCO (Tabla 17), fue corregido mediante el empleo del método PW. En la Tabla se aprecia que el estadístico d de DW se acerca al valor de dos, a diferencia de los modelos MCO (Tabla 17).

**Tabla 18**

*Modelos Corregidos por PW con Variable Dependiente L-CONG (1980-2018)*

Nº	Método	Variable dependiente	Variabes independientes	$r^2$ o $R^2$	$r_c^2$ o $R_c^2$	Merma	FIV	d de DW
1	PW	L-CONG	DI	83,63%	83,18%	0,44%	na	1,760
2	PW	L-CONG	DI - INF1	85,97%	85,19%	0,78%	1,061	1,876
3	PW	L-CONG	PBIman - INF1	86,52%	85,77%	0,75%	1,078	1,879
4	PW	L-CONG	PBI - INF1	87,06%	86,34%	0,72%	1,063	1,895

*Nota:* na = no aplica. Fuente Software Gretl 2019a.

**4.2.1.9 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. DI (1980-2018).** Los valores p de la prueba F y los de las pruebas t, son menores e iguales al  $\alpha$  de 0,05. Por un lado, el valor p de la prueba F es  $1,37 \times 10^{-6}$  (Figura 86), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, globalmente, son iguales a cero. Por otro lado, el valor p de la prueba t es  $1,83 \times 10^{-6}$  (DI) (Figura 86), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que el coeficiente es igual a cero. De igual

manera, el valor p de la constante es  $3,19 \times 10^{-7}$  (Figura 86). Ante los resultados, se concluye que el modelo es significativo, por tanto el modelo ha sido validado.

### Figura 86

Modelo PW: L-CONG vs. DI (1980-2018)

	Coeficiente	Desv. típica	Estadístico t	valor p
const	1,87624	0,301789	6,217	3,19e-07 ***
DI	5,75274e-06	1,01677e-06	5,658	1,83e-06 ***

Estadísticos basados en los datos rho-diferenciados:			
Media de la vble. dep.	3,377823	D.T. de la vble. dep.	0,845207
Suma de cuad. residuos	4,445831	D.T. de la regresión	0,346638
R-cuadrado	0,836267	R-cuadrado corregido	0,831842
F(1, 37)	33,05868	Valor p (de F)	1,37e-06
rho	0,118015	Durbin-Watson	1,760263

Nota. Fuente Software Gretl 2019a.

En cuanto a poder explicativo, el modelo presenta un  $r^2$  de 83,63% (Figura 86). Es decir, el comportamiento de la variable dependiente L-CONG está siendo explicado en un 83,63% por la variable DI. Finalmente, el modelo PW (Figura 86) es:

$$\text{L-CONG} = 1,87624 + 0,00000575274 \text{ DI}$$

Según el modelo, la demanda de congelados aumentará en 0,000575274% ( $0,00000575274 \times 100$ ) por cada millón de soles de incremento en la DI.

**4.2.1.10 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs DI e INF1 (1980-2018).** Los valores p de la prueba F y los de las pruebas t, son menores e iguales al  $\alpha$  de 0,05. De un lado, el valor p de la prueba F es  $1,26 \times 10^{-7}$  (Figura 87), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, globalmente, son iguales a cero. De otro lado, los valores p de la prueba t son  $3,71 \times 10^{-8}$  (DI) y 0,0177 (INF1) (Figura 87), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, individualmente, son iguales a cero. Análogamente, el valor p de la constante es

$7,45 \times 10^{-8}$  (Figura 87). De acuerdo a los resultados, se colige que el modelo es significativo y, en consecuencia, el modelo ha sido validado.

### Figura 87

Modelo PW: L-CONG vs. DI e INF1 (1980-2018)

	Coefficiente	Desv. típica	Estadístico t	valor p
const	1,75004	0,260056	6,729	7,45e-08 ***
DI	6,10138e-06	8,76678e-07	6,960	3,71e-08 ***
INF1	0,000105973	4,26108e-05	2,487	0,0177 **

Estadísticos basados en los datos rho-diferenciados:			
Media de la vble. dep.	3,377823	D.T. de la vble. dep.	0,845207
Suma de cuad. residuos	3,809520	D.T. de la regresión	0,325300
R-cuadrado	0,859694	R-cuadrado corregido	0,851900
F(2, 36)	25,51223	Valor p (de F)	1,26e-07
rho	0,061325	Durbin-Watson	1,875769

Nota. Fuente Software Gretl 2019a.

En lo referente a poder explicativo, el modelo exhibe un  $R^2$  de 85,97% (Figura 87). Es decir, el comportamiento de la variable dependiente L-CONG está siendo explicado en un 85,97% por las variables DI e INF1. Por último, el modelo PW (Figura 87) es:

$$L-CONG = 1,75004 + 0,00000610138 DI + 0,000105973 INF1$$


Según el modelo, la demanda de congelados aumentará en 0,000610138% (0,00000610138 x 100) por cada millón de soles de incremento en la DI, manteniendo constante la INF1. Asimismo, la demanda de congelados aumentará en 0,0105973% (0,000105973 x 100) por cada punto porcentual de incremento en la INF1, manteniendo constante el PBI.

**4.2.1.11 Significancia, Poder Explicativo e Interpretación del Modelo PW: L-CONG vs. PBI<sub>man</sub> e INF1 (1980-2018).** Los valores p de la prueba F y los de las pruebas t, son menores e iguales al  $\alpha$  de 0,05. Por un lado, el valor p de la prueba F es  $1,88 \times 10^{-9}$  (Figura 88), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, globalmente, son iguales a cero. Por otro lado, los valores p de la prueba t son  $2,84 \times 10^{-10}$

(PBI<sub>man</sub>) y 0,0049 (INF1) (Figura 88), por tanto existe suficiente evidencia para rechazar la  $H_0$  de que los coeficientes, individualmente, son iguales a cero. De igual manera, el valor p de la constante es  $3,47 \times 10^{-5}$  (Figura 88). En vista de los resultados, se concluye que el modelo es significativo, por tanto el modelo ha sido validado.

### Figura 88

*Modelo PW: L-CONG vs. PBI<sub>man</sub> e INF1 (1980-2018)*



	Coeficiente	Desv. típica	Estadístico t	valor p	
const	1,24544	0,263577	4,725	3,47e-05	***
PBI <sub>man</sub>	5,05086e-05	5,86220e-06	8,616	2,84e-010	***
INF1	0,000128417	4,28523e-05	2,997	0,0049	***

Medida de la vble. dep.	Suma de cuad. residuos	R-cuadrado	F(2, 36)	rho	D.T. de la vble. dep.	D.T. de la regresión	R-cuadrado corregido	Valor p (de F)	Durbin-Watson
3,377823	3,663106	0,865179	36,96016	0,058038	0,845207	0,318987	0,857689	1,88e-09	1,878802

*Nota.* Fuente Software Gretl 2019a.

En cuanto a poder explicativo, el modelo validado presenta un  $R^2$  de 86,52% (Figura 88). Es decir, el comportamiento de la variable dependiente L-CONG está siendo explicado en un 86,52% por las variables PBI<sub>man</sub> e INF1. Finalmente, el modelo PW (Figura 88) es:

$$L\text{-CONG} = 1,24544 + 0,0000505086 \text{ PBI}_{\text{man}} + 0,000128417 \text{ INF1}$$

Según el modelo, la demanda de congelados aumentará en 0,00505086% (0,0000505086 x 100) por cada millón de soles de incremento en el PBI<sub>man</sub>, manteniendo constante la INF1. De igual manera, la demanda de congelados aumentará en 0,0128417% (0,000128417 x 100) por cada punto porcentual de incremento en la INF1, manteniendo constante el PBI<sub>man</sub>.

#### 4.2.1.12 Construcción, Verificación y Corrección de los Modelos 1980-2013.

Considerando la información del periodo 1980-2013 y, siguiendo el procedimiento del numeral 4.2.1 (p. 114), no solo se replicaron los cuatro modelos validados (Tabla 18, p. 134),

sino también se incluyeron tres modelos de regresión simple con las variables independientes PBI<sub>man</sub>, PBI e INF1. Primero se construyeron y verificaron los modelos con variable dependiente CONG, los modelos violaron entre dos y cuatro supuestos (Tabla 19).

**Tabla 19**

*Verificación de los Modelos MCO con Variable Dependiente CONG (1980-2013)*

Nº	Método	Variable dependiente	Variables independientes	$r^2$ o $R^2$	$r_c^2$ o $R_c^2$	Merma	FIV	d de DW	Supuestos violados
1	MCO	CONG	DI	73,59%	72,76%	0,83%	na	1,261	4, 6, 7
2	MCO	CONG	DI - INF1	80,65%	79,40%	1,25%	1,066	1,631	6, 7
3	MCO	CONG	PBI <sub>man</sub>	73,85%	73,03%	0,82%	na	1,328	1, 4, 6, 7
4	MCO	CONG	PBI <sub>man</sub> - INF1	81,65%	80,47%	1,18%	1,075	1,769	6, 7
5	MCO	CONG	PBI	75,01%	74,23%	0,78%	na	1,282	1, 4, 6
6	MCO	CONG	PBI - INF1	82,47%	81,34%	1,13%	1,070	1,727	6, 7
7	MCO	CONG	INF1	0,19%	-2,93%	3,12%	na	0,305	1, 4, 6, 7

*Nota:* na = no aplica. Fuente Software Gretl 2019a.

Seguidamente, y para intentar corregir las violaciones MCO (Tabla 19), se cambió la variable dependiente a L-CONG, variando la forma funcional del modelo; lo cual provocó que cinco modelos violaran únicamente el supuesto 4 (Tabla 20).

**Tabla 20**

*Verificación de los Modelos MCO con Variable Dependiente L-CONG (1980-2013)*

Nº	Método	Variable dependiente	Variables independientes	$r^2$ o $R^2$	$r_c^2$ o $R_c^2$	Merma	FIV	d de DW	Supuestos violados
1	MCO	L-CONG	DI	58,29%	56,99%	1,30%	na	0,815	4
2	MCO	L-CONG	DI - INF1	69,04%	67,04%	2,00%	1,066	1,081	4
3	MCO	L-CONG	PBI <sub>man</sub>	63,51%	62,37%	1,14%	na	0,923	4, 7
4	MCO	L-CONG	PBI <sub>man</sub> - INF1	75,70%	74,14%	1,57%	1,075	1,347	4
5	MCO	L-CONG	PBI	62,23%	61,05%	1,18%	na	0,862	4
6	MCO	L-CONG	PBI - INF1	73,75%	72,05%	1,69%	1,07	1,225	4
7	MCO	L-CONG	INF1	1,61%	-1,46%	3,07%	na	0,294	1, 4, 7

*Nota:* na = no aplica. Fuente Software Gretl 2019a.

Por último, luego de la aplicación del método PW a los modelos que únicamente violaron el supuesto 4 (Tabla 20), se observa que los d de DW están muy cercanos a dos, lo

cual evidencia que los errores de los modelos ya no violan el supuesto de autocorrelación (Tabla 21).

**Tabla 21**

*Modelos PW con Variable Dependiente L-CONG (1980-2013)*

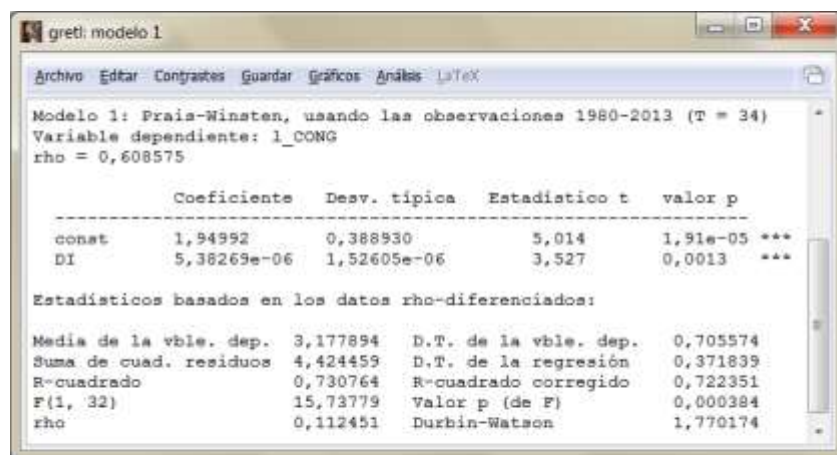
Nº	Método	Variable dependiente	Variabes independientes	$r^2$ o $R^2$	$r_c^2$ o $R_c^2$	Merma	FIV	d de DW
1	PW	L-CONG	DI	73,08%	72,24%	0,84%	na	1,770
2	PW	L-CONG	DI - INF1	76,87%	75,38%	1,49%	1,066	1,877
3	PW	L-CONG	PBIman - INF1	79,06%	77,71%	1,35%	1,075	1,902
4	PW	L-CONG	PBI	74,47%	73,67%	0,80%	na	1,786
5	PW	L-CONG	PBI - INF1	78,70%	77,33%	1,37%	1,070	1,888

*Nota:* na = no aplica. Fuente Software Gretl 2019a.

**4.2.1.13 Significancia de los Modelos PW 1980-2013.** Se evaluó la significancia de los cinco modelos PW presentados en la Tabla 21. Para el modelo PW: L-CONG vs. DI (1980-2013), el valor p de la prueba F es 0,00038 y el valor p de la prueba t es 0,0013 (DI), además el valor p de la constante es  $1,91 \times 10^{-5}$  (Figura 89). Como los valores p son menores e iguales al  $\alpha$  de 0,05, existe suficiente evidencia para rechazar las  $H_0$  de que los coeficientes, globalmente e individualmente, son iguales a cero. Por tanto el modelo es significativo y se colige que el modelo ha sido validado.

**Figura 89**

*Modelo PW: L-CONG vs. DI (1980-2013)*



```

gretl: modelo 1
-----
Modelo 1: Prais-Winsten, usando las observaciones 1980-2013 (T = 34)
Variable dependiente: l_CONG
rho = 0,608575

-----
                Coeficiente  Desv. típica  Estadístico t  valor p
-----
const          1,94992      0,388930      5,014          1,91e-05 ***
DI              5,38269e-06    1,52605e-06    3,527          0,0013 ***

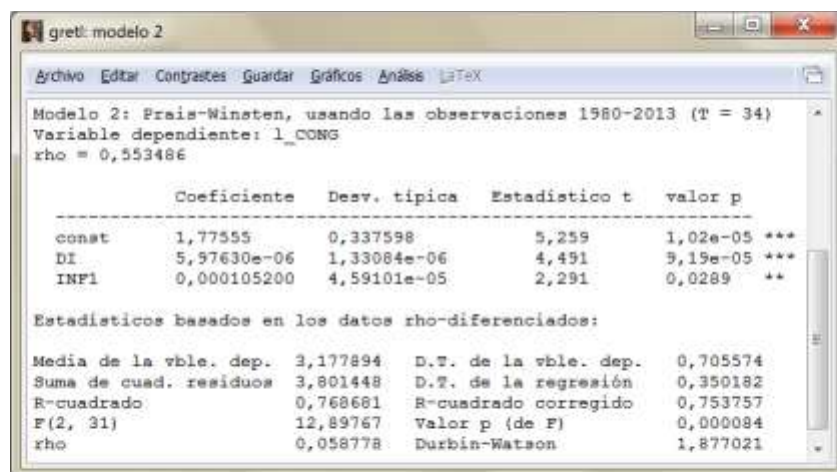
Estadísticos basados en los datos rho-diferenciados:
-----
Media de la vble. dep.  3,177894  D.T. de la vble. dep.  0,705574
Suma de cuad. residuos  4,424459  D.T. de la regresión   0,371839
R-cuadrado             0,730764  R-cuadrado corregido   0,722351
F(1, 32)              15,73779  Valor p (de F)         0,000384
rho                   0,112451  Durbin-Watson          1,770174
  
```

*Nota.* Fuente Software Gretl 2019a.

En cuanto al modelo PW: L-CONG vs. DI e INF1 (1980-2013), el valor p de la prueba F es  $0,000084$  y los valores p de la prueba t son  $9,19 \times 10^{-5}$  (DI) y  $0,0289$  (INF1), además el valor p de la constante es  $1,02 \times 10^{-5}$  (Figura 90). Dado que los valores p son menores e iguales al  $\alpha$  de  $0,05$ , existe suficiente evidencia para rechazar las  $H_0$  de que los coeficientes, globalmente e individualmente, son iguales a cero. Por tanto el modelo es significativo y, en consecuencia, el modelo ha sido validado.

### Figura 90

Modelo PW: L-CONG vs. DI e INF1 (1980-2013)



	Coefficiente	Desv. típica	Estadístico t	valor p
const	1,77595	0,337598	5,259	1,02e-05 ***
DI	5,97630e-06	1,33084e-06	4,491	9,19e-05 ***
INF1	0,000105200	4,59101e-05	2,291	0,0289 **

Estadísticos basados en los datos rho-diferenciados:

Media de la vble. dep.	3,177894	D.T. de la vble. dep.	0,705574
Suma de cuad. residuos	3,801448	D.T. de la regresión	0,350182
R-cuadrado	0,768681	R-cuadrado corregido	0,753757
F(2, 31)	12,89767	Valor p (de F)	0,000084
rho	0,058778	Durbin-Watson	1,877021

Nota. Fuente Software Gretl 2019a.

En lo que respecta al modelo PW: L-CONG vs. PBIman e INF1 (1980-2013), el valor p de la prueba F es  $1,55 \times 10^{-6}$  y los valores p de la prueba t son  $5,76 \times 10^{-7}$  (PBIman) y  $0,0079$  (INF1), además el valor p de la constante es  $4,33 \times 10^{-5}$  (Figura 91). Dado que los valores p son menores e iguales al  $\alpha$  de  $0,05$ , existe suficiente evidencia para rechazar las  $H_0$  de que los coeficientes, globalmente e individualmente, son iguales a cero. Por tanto el modelo es significativo y, en consecuencia, el modelo ha sido validado.

**Figura 91**

*Modelo PW: L-CONG vs. PBIman e INF1 (1980-2013)*

```

gretl: modelo 3
-----
Modelo 3: Fraiss-Winsten, usando las observaciones 1980-2013 (T = 34)
Variable dependiente: l_CONG
rho = 0,424031

-----
                Coeficiente   Desv. típica   Estadístico t   valor p
-----
const          1,41069         0,296675       4,755           4,33e-05 ***
PBIman         4,52003e-05           7,21322e-06    6,266           5,76e-07 ***
INF1           0,000127642          4,49789e-05    2,838           0,0079 ***

Estadísticos basados en los datos rho-diferenciados:

Media de la vble. dep.  3,177894   D.T. de la vble. dep.  0,705574
Suma de cuad. residuos  3,440795   D.T. de la regresión   0,333157
R-cuadrado              0,790635   R-cuadrado corregido   0,777128
F(2, 31)                21,23473   Valor p (de F)         1,55e-06
rho                      0,046749   Durbin-Watson           1,901672
  
```

*Nota.* Fuente Software Gretl 2019a.

En relación al modelo PW: L-CONG vs. PBI (1980-2013), se observa que el valor p de la prueba F es 0,00009 y el valor p de la prueba t es 0,0002 (PBI), además el valor p de la constante es  $5,04 \times 10^{-5}$  (Figura 92). Todos los valores p son menores e iguales al  $\alpha$  de 0,05, por tanto existe suficiente evidencia para rechazar las  $H_0$  de que los coeficientes, globalmente e individualmente, son iguales a cero. Por lo tanto el modelo es significativo y, en consecuencia, el modelo ha sido validado.

**Figura 92**

*Modelo PW: L-CONG vs. PBI (1980-2013)*

```

gretl: modelo 1
-----
Modelo 1: Fraiss-Winsten, usando las observaciones 1980-2013 (T = 34)
Variable dependiente: l_CONG
rho = 0,573617

-----
                Coeficiente   Desv. típica   Estadístico t   valor p
-----
const          1,75900         0,375951       4,679           5,04e-05 ***
PBI            5,97251e-06     1,44277e-06    4,140           0,0002 ***

Estadísticos basados en los datos rho-diferenciados:

Media de la vble. dep.  3,177894   D.T. de la vble. dep.  0,705574
Suma de cuad. residuos  4,194411   D.T. de la regresión   0,362043
R-cuadrado              0,744720   R-cuadrado corregido   0,736743
F(1, 32)                20,05127   Valor p (de F)         0,000090
rho                      0,105408   Durbin-Watson           1,786341
  
```

*Nota.* Fuente Software Gretl 2019a.

Finalmente, en el modelo PW: L-CONG vs. PBI e INF1 (1980-2013), el valor p de la prueba F es  $7,10 \times 10^{-6}$  y los valores p de la prueba t son  $4,15 \times 10^{-6}$  (PBI) y 0,0154 (INF1), además el valor p de la constante es  $1,47 \times 10^{-5}$  (Figura 93). Debido a que los valores p son menores e iguales al  $\alpha$  de 0,05, existe suficiente evidencia para rechazar las  $H_0$  de que los coeficientes, globalmente e individualmente, son iguales a cero. Por tanto el modelo es significativo y, en consecuencia, el modelo ha sido validado.

### Figura 93

*Modelo PW: L-CONG vs. PBI e INF1 (1980-2013)*

	Coefficiente	Desv. típica	Estadístico t	valor p
const.	1,57379	0,306579	5,133	1,47e-05 ***
PBI	6,57529e-06	1,17978e-06	5,573	4,15e-06 ***
INF1	0,000114727	4,47543e-05	2,563	0,0154 **

Estadísticos basados en los datos rho-diferenciados:			
Media de la vble. dep.	3,177894	D.T. de la vble. dep.	0,705574
Suma de cuad. residuos	3,499665	D.T. de la regresión	0,335995
R-cuadrado	0,787045	R-cuadrado corregido	0,773306
F(2, 31)	17,80610	Valor p (de F)	7,10e-06
rho	0,053295	Durbin-Watson	1,888386

*Nota.* Fuente Software Gretl 2019a.

## 4.2.2 Extensión de las Series Temporales y Precisión del Pronóstico

**4.2.2.1 Modelos PW: L-CONG vs. DI.** El pronóstico del modelo PW: L-CONG vs. DI (1980-2013) presenta un EMAP de 13,44% y un sesgo de 16,44 (Tabla 22). En 2014, el Error absoluto porcentual (EAP) y el error son muy pequeños, luego experimentan una tendencia creciente que alcanza su máximo en 2018 con 18,17% y 25,47, respectivamente (Tabla 22). La tendencia creciente del EAP se aprecia en la separación ascendente entre el CONG real y el pronosticado, y el sesgo positivo se visualiza en la medida que el CONG pronosticado siempre subestima al CONG real (Figura 94).

**Tabla 22**

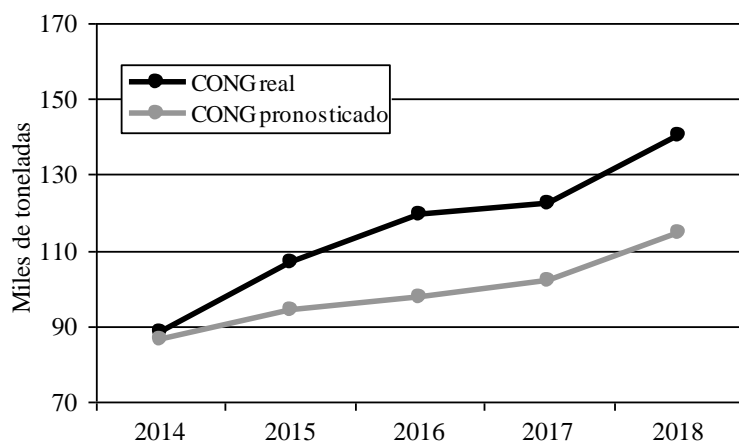
*Modelo PW: L-CONG vs. DI (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018)*

Año	CONG real (miles de t)	CONG pronosticado		Error absoluto porcentual	Error
		(Logaritmo)	(miles de t)		
2014	88,6	4,4585	86,4	2,53%	2,24
2015	106,9	4,5456	94,2	11,89%	12,71
2016	119,3	4,5818	97,7	18,11%	21,61
2017	122,2	4,6251	102,0	16,52%	20,19
2018	140,2	4,7426	114,7	18,17%	25,47
				13,44%	16,44
				EMAP	Sesgo

*Nota:* Fuente PRODUCE, Software Gretl 2019a.

**Figura 94**

*Modelo PW: L-CONG vs. DI (1980-2013). CONG Real y Pronosticado (2014-2018)*



*Nota.* Fuente PRODUCE, Software Gretl 2019a.

A continuación, la Tabla 23 muestra los 22 modelos PW: L-CONG vs. DI validados, de los cuales ocho presentaron precisión Muy buena ( $EMAP \leq 10\%$ ), 13 exhibieron precisión Buena ( $10\% < EMAP \leq 20\%$ ) y un modelo tuvo precisión Moderada ( $20\% < EMAP \leq 30\%$ ). El EMAP y el sesgo, de los 21 modelos adicionales, se calculó según lo mostrado en la Tabla 22.

**Tabla 23**

*Modelos PW: L-CONG vs. DI. EMAP, Precisión, Sesgo (2014-2018) y Merma Según*

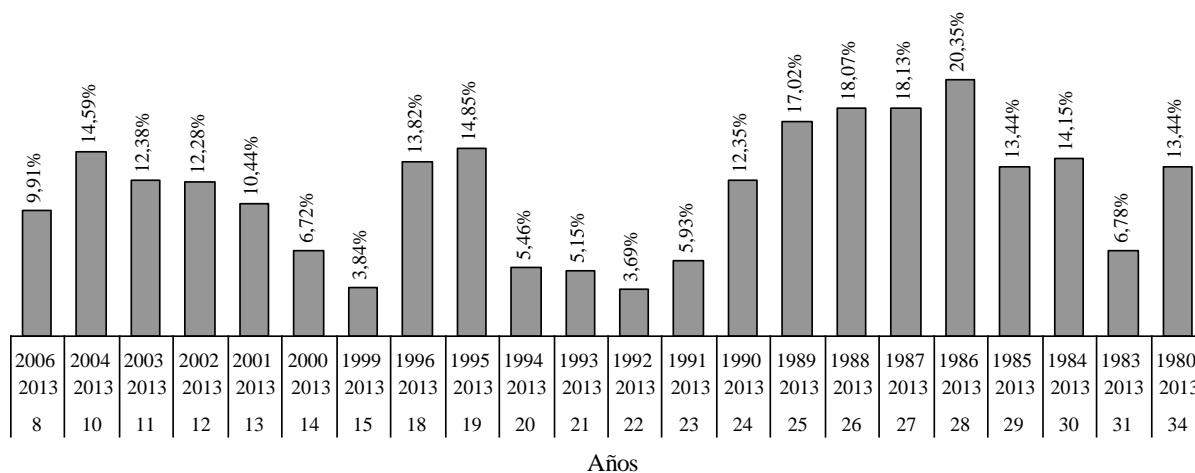
*Extensión de las Series Temporales*

Nº	Periodo	Extensión (Años)	EMAP	Precisión	Sesgo	$r^2$	$r_c^2$	Merma
1	2006-2013	8	9,91%	Muy buena	9,88	64,31%	58,36%	5,95%
2	2004-2013	10	14,59%	Buena	16,30	58,73%	53,57%	5,16%
3	2003-2013	11	12,38%	Buena	12,12	65,29%	61,43%	3,86%
4	2002-2013	12	12,28%	Buena	11,89	68,40%	65,24%	3,16%
5	2001-2013	13	10,44%	Buena	8,06	72,16%	69,63%	2,53%
6	2000-2013	14	6,72%	Muy buena	0,84	68,64%	66,03%	2,61%
7	1999-2013	15	3,84%	Muy buena	-0,75	71,49%	69,29%	2,19%
8	1996-2013	18	13,82%	Buena	16,91	76,73%	75,28%	1,45%
9	1995-2013	19	14,85%	Buena	18,15	76,86%	75,49%	1,36%
10	1994-2013	20	5,46%	Muy buena	5,31	76,64%	75,34%	1,30%
11	1993-2013	21	5,15%	Muy buena	4,89	78,41%	77,28%	1,14%
12	1992-2013	22	3,69%	Muy buena	2,64	80,02%	79,02%	1,00%
13	1991-2013	23	5,93%	Muy buena	5,98	78,65%	77,64%	1,02%
14	1990-2013	24	12,35%	Buena	15,22	73,33%	72,12%	1,21%
15	1989-2013	25	17,02%	Buena	20,76	73,51%	72,35%	1,15%
16	1988-2013	26	18,07%	Buena	22,02	73,37%	72,26%	1,11%
17	1987-2013	27	18,13%	Buena	22,09	73,01%	71,93%	1,08%
18	1986-2013	28	20,35%	Moderada	24,74	72,01%	70,93%	1,08%
19	1985-2013	29	13,44%	Buena	16,50	71,23%	70,17%	1,07%
20	1984-2013	30	14,15%	Buena	17,34	71,70%	70,69%	1,01%
21	1983-2013	31	6,78%	Muy buena	7,07	69,50%	68,45%	1,05%
22	1980-2013	34	13,44%	Buena	16,44	73,08%	72,24%	0,84%

Seguidamente, la Figura 95 muestra, para los modelos PW: L-CONG vs. DI, la evolución del EMAP en relación a la extensión de las series. En general, la relación es negativa entre precisión y extensión, sin embargo, en el análisis se observa una evolución mixta. Por un lado, la relación es positiva para series de entre 10 (2004-2013) y 15 años (1999-2013). Por otro lado, la relación es negativa para extensiones de entre 22 (1992-2013) y 28 años (1986-2013).

**Figura 95**

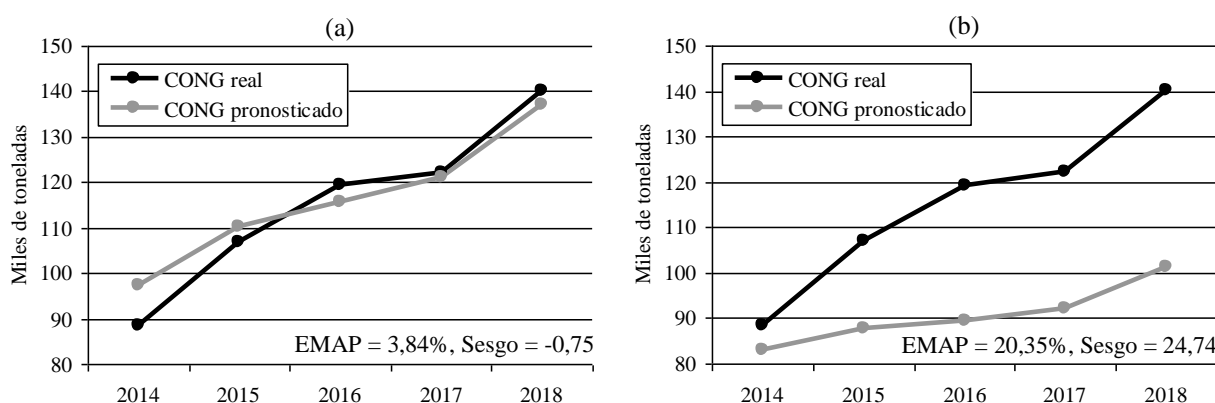
*Modelos PW: L-CONG vs. DI. EMAP (2014-2018) Según Extensión de las Series Temporales*



Finalmente, de todos los modelos PW: L-CONG vs. DI, el de mayor precisión fue el construido con 15 años (1999-2013), el EMAP fue 3,84% (precisión Muy buena) y el sesgo fue -0,75 (Figura 96a). Mientras que el modelo menos preciso se construyó con 28 años (1986-2013), el EMAP fue 20,35% (precisión Moderada) y el sesgo fue 24,74 (Figura 96b).

**Figura 96**

*CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. DI con Extensión de (a) 15 Años (1999-2013) y (b) 28 Años (1986-2013)*



**4.2.2.2 Modelos PW: L-CONG vs. DI e INF1.** El pronóstico del modelo PW: L-CONG vs. DI e INF1 (1980-2013) muestra un EMAP de 5,78% y un sesgo de 5,92 (Tabla 24). En 2014 y 2015, el EAP y el error son muy pequeños, posteriormente alcanzan su

máximo en 2016 con 9,14% y 10,90, respectivamente, para luego continuar con una tendencia decreciente (Tabla 24). Este modelo inicia con un error negativo (-2,57) que luego se torna positivo para el resto del periodo (Tabla 24). Los EAPs pequeños se aprecian en la cercanía entre el CONG real y el pronosticado, y, el cambio de signo en el error, se visualiza en 2014, donde el CONG pronosticado sobrestima al CONG real, mientras que el comportamiento se invierte en el resto del periodo (Figura 97).

**Tabla 24**

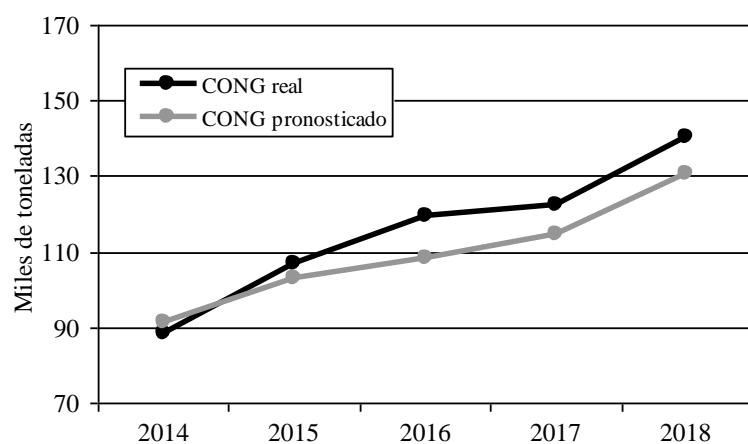
*Modelo PW: L-CONG vs. DI e INF1 (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018)*

Año	CONG real (miles de t)	CONG pronosticado		Error absoluto porcentual	Error
		(Logaritmo)	(miles de t)		
2014	88,6	4,5127	91,2	2,90%	-2,57
2015	106,9	4,6334	102,9	3,81%	4,07
2016	119,3	4,6858	108,4	9,14%	10,90
2017	122,2	4,7399	114,4	6,36%	7,77
2018	140,2	4,8737	130,8	6,71%	9,40
				5,78%	5,92
				EMAP	Sesgo

*Nota:* Fuente PRODUCE, Software Gretl 2019a.

**Figura 97**

*Modelo PW: L-CONG vs. DI e INF1 (1980-2013). CONG Real y Pronosticado (2014-2018)*



A continuación, la Tabla 25 muestra que se lograron validar 11 modelos PW: L-CONG vs. DI e INF1, de los cuales nueve presentaron precisión Muy buena ( $EMAP \leq 10\%$ ) y dos exhibieron precisión Buena ( $10\% < EMAP \leq 20\%$ ). El EMAP y el sesgo, de los 10 modelos adicionales, se calculó según lo mostrado en la Tabla 24.

**Tabla 25**

*Modelos PW: L-CONG vs. DI e INF1. EMAP, Precisión, Sesgo (2014-2018) y Merma Según*

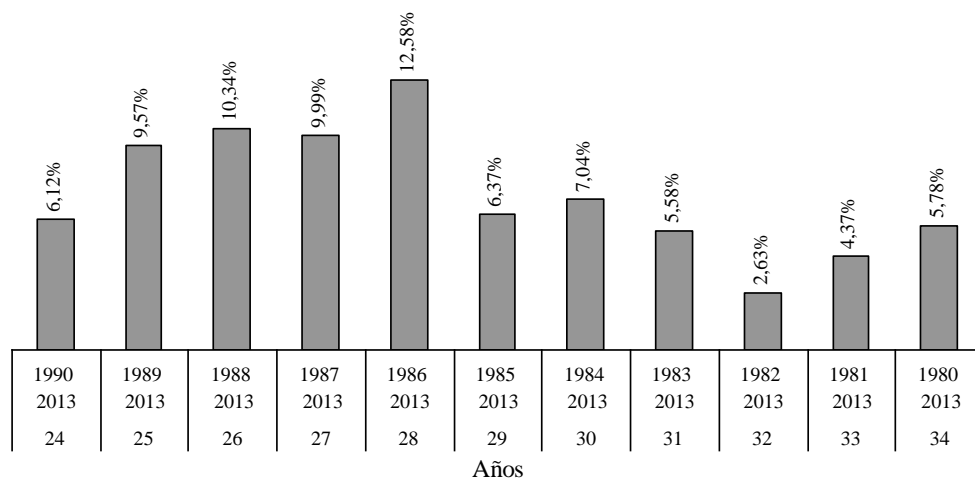
*Extensión de las Series Temporales*

Nº	Periodo	Extensión (Años)	EMAP	Precisión	Sesgo	R <sup>2</sup>	R <sub>c</sub> <sup>2</sup>	Merma
1	1990-2013	24	6,12%	Muy buena	6,31	78,91%	76,90%	2,01%
2	1989-2013	25	9,57%	Muy buena	11,76	78,11%	76,12%	1,99%
3	1988-2013	26	10,34%	Buena	12,78	78,06%	76,15%	1,91%
4	1987-2013	27	9,99%	Muy buena	12,36	78,10%	76,27%	1,83%
5	1986-2013	28	12,58%	Buena	15,47	76,82%	74,97%	1,85%
6	1985-2013	29	6,37%	Muy buena	6,49	76,25%	74,43%	1,83%
7	1984-2013	30	7,04%	Muy buena	7,56	76,58%	74,85%	1,73%
8	1983-2013	31	5,58%	Muy buena	-5,97	74,24%	72,40%	1,84%
9	1982-2013	32	2,63%	Muy buena	-1,23	75,46%	73,77%	1,69%
10	1981-2013	33	4,37%	Muy buena	3,67	76,64%	75,09%	1,56%
11	1980-2013	34	5,78%	Muy buena	5,92	76,87%	75,38%	1,49%

Seguidamente, en la Figura 98 se muestra, para los modelos PW: L-CONG vs. DI e INF1, el desempeño del EMAP en relación a la extensión de las series. En general, se aprecia una relación positiva entre precisión y extensión. Se observa que los menores EMAPs (mejores precisiones) se obtuvieron al utilizar extensiones de entre 29 (1985-2013) y 34 años (1980-2013). Y para extensiones de entre 30 (1984-2013) y 32 años (1982-2013) se aprecia una relación positiva entre EMAP y extensión. Sin embargo, la relación es negativa para series de entre 24 (1990-2013) y 28 años (1986-2013). Y también lo es para series de entre 32 (1982-2013) y 34 años (1980-2013).

**Figura 98**

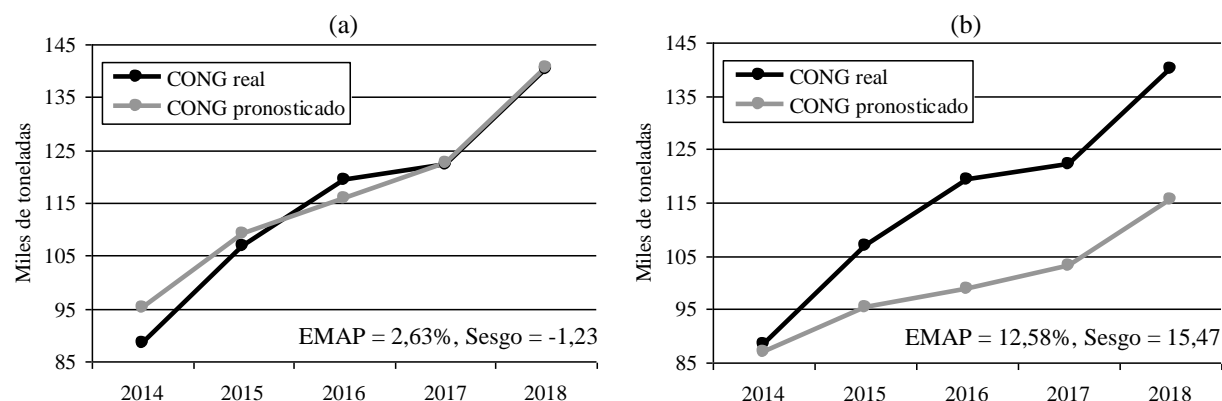
*Modelos PW: L-CONG vs DI e INF1. EMAP (2014-2018) Según Extensión de las Series Temporales*



Finalmente, de los modelos PW: L-CONG vs. DI e INF1, el de mayor precisión fue el construido con 32 años (1982-2013), el EMAP fue 2,63% (precisión Muy buena) y el sesgo fue -1,23 (Figura 99a). Mientras que el modelo de menor precisión se obtuvo con 28 años (1986-2013), el EMAP fue 12,58% (precisión Buena) y el sesgo fue 15,47 (Figura 99b).

**Figura 99**

*CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. DI e INF1 con Extensión de (a) 32 Años (1982-2013) y (b) 28 Años (1986-2013)*



#### 4.2.2.3 Modelos PW: L-CONG vs. PBIman e INF1. El pronóstico del modelo PW:

L-CONG vs. PBIman e INF1 (1980-2013) presenta un EMAP de 31,34% y un sesgo de 37,64

(Tabla 26). En el periodo 2014-2018, el EAP y del error oscila entre mediano y muy grande; el EAP alcanza su máximo en 2017 con 39,62% y el error lo hace en 2018 con 51,80 (Tabla 26). Los EAPs medianos y muy grandes se aprecian en la creciente separación entre el CONG real y el pronosticado; y el sesgo positivo se visualiza cuando el CONG pronosticado subestima consistentemente al CONG real (Figura 100).

**Tabla 26**

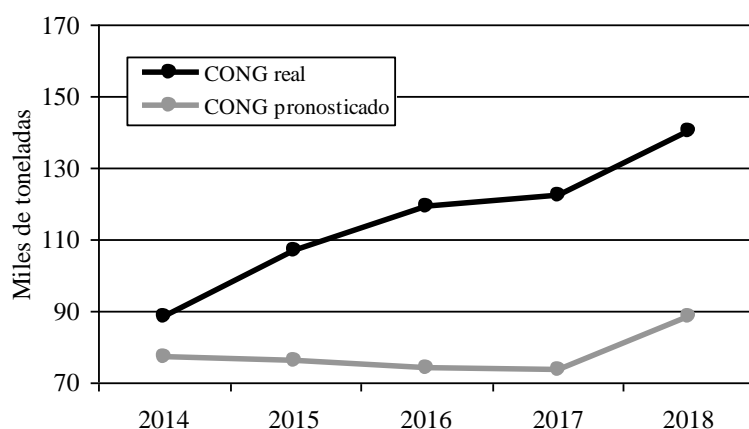
*Modelo PW: L-CONG vs. PBI<sub>man</sub> e INF1 (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018)*

Año	CONG real (miles de t)	CONG pronosticado		Error absoluto porcentual	Error
		(Logaritmo)	(miles de t)		
2014	88,6	4,3433	77,0	13,14%	11,64
2015	106,9	4,3304	76,0	28,95%	30,96
2016	119,3	4,3029	73,9	38,04%	45,38
2017	122,2	4,3012	73,8	39,62%	48,41
2018	140,2	4,4819	88,4	36,95%	51,80
				31,34%	37,64
				EMAP	Sesgo

*Nota:* Fuente PRODUCE, Software Gretl 2019a.

**Figura 100**

*Modelo PW: L-CONG vs. PBI<sub>man</sub> e INF1 (1980-2013). CONG Real y Pronosticado (2014-2018)*



A continuación, la Tabla 27 muestra que se lograron validar 12 modelos PW: L-CONG vs. PBI<sub>man</sub> e INF1, de los cuales cuatro presentaron precisión Moderada ( $20\% < \text{EMAP} \leq 30\%$ ) y ocho exhibieron precisión Pobre ( $\text{EMAP} > 30\%$ ). El EMAP y el sesgo, de los 11 modelos adicionales, se calculó según lo mostrado en la Tabla 26.

**Tabla 27**

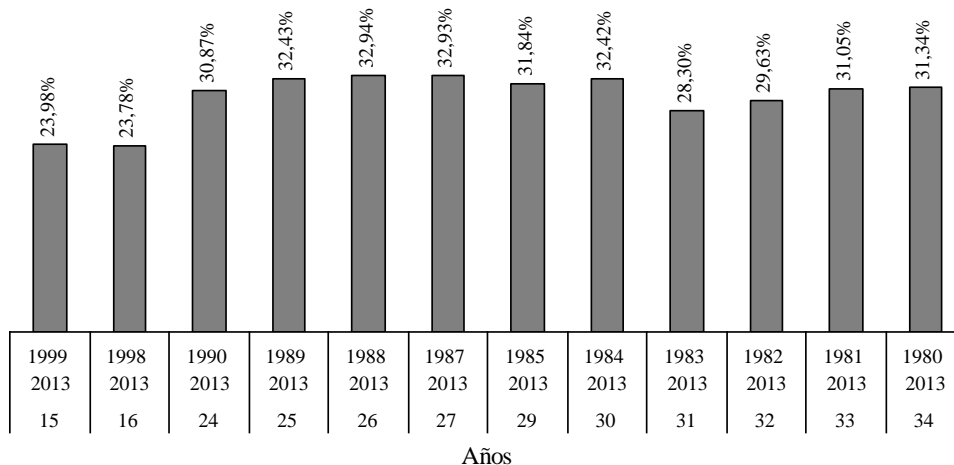
*Modelos PW: L-CONG vs. PBI<sub>man</sub> e INF1. EMAP, Precisión, Sesgo (2014-2018) y Merma Según Extensión de las Series Temporales*

Nº	Periodo	Extensión (Años)	EMAP	Precisión	Sesgo	R <sup>2</sup>	R <sub>c</sub> <sup>2</sup>	Merma
1	1999-2013	15	23,98%	Moderada	28,58	83,98%	81,32%	2,67%
2	1998-2013	16	23,78%	Moderada	28,16	87,58%	85,67%	1,91%
3	1990-2013	24	30,87%	Pobre	37,13	81,26%	79,47%	1,78%
4	1989-2013	25	32,43%	Pobre	38,95	79,99%	78,17%	1,82%
5	1988-2013	26	32,94%	Pobre	39,54	80,00%	78,26%	1,74%
6	1987-2013	27	32,93%	Pobre	39,53	80,05%	78,39%	1,66%
7	1985-2013	29	31,84%	Pobre	38,29	78,55%	76,91%	1,65%
8	1984-2013	30	32,42%	Pobre	38,97	78,67%	77,09%	1,58%
9	1983-2013	31	28,30%	Moderada	34,26	77,80%	76,21%	1,59%
10	1982-2013	32	29,63%	Moderada	35,71	77,71%	76,17%	1,54%
11	1981-2013	33	31,05%	Pobre	37,31	78,73%	77,31%	1,42%
12	1980-2013	34	31,34%	Pobre	37,64	79,06%	77,71%	1,35%

Seguidamente, la Figura 101 muestra, para los modelos PW: L-CONG vs. PBI<sub>man</sub> e INF1, el comportamiento del EMAP con respecto a la extensión de las series. Se aprecia una relación negativa entre precisión y extensión. Para extensiones de entre 24 (1990-2013) y 27 años (1987-2013) el EMAP es creciente. Y también lo es para series de entre 31 (1983-2013) y 34 años (1980-2013).

**Figura 101**

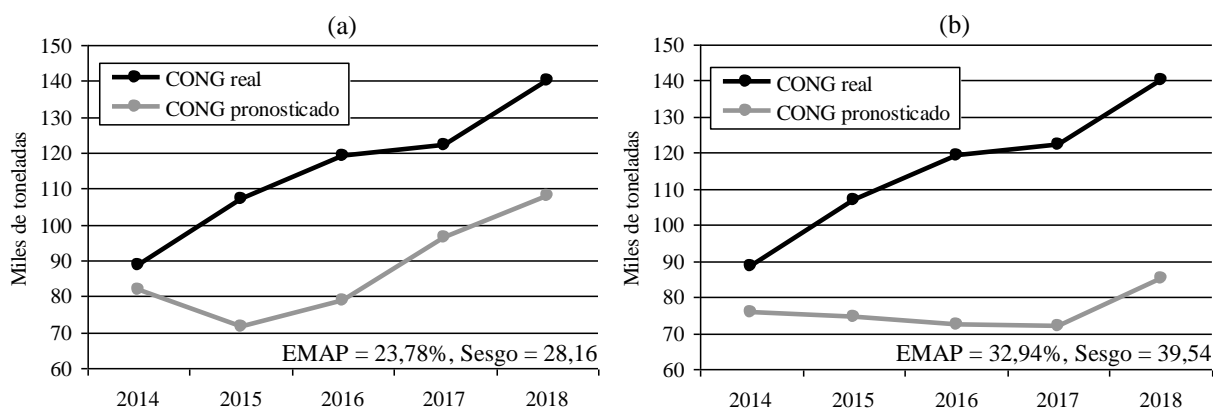
*Modelos PW: L-CONG vs. PBIman e INF1. EMAP (2014-2018) Según Extensión de las Series Temporales*



Finalmente, de todos los modelos PW: L-CONG vs. PBIman e INF1, el de mayor precisión fue el construido con 16 años (1998-2013), el EMAP fue 23,78% (precisión Moderada) y el sesgo fue 28,16 (Figura 102a). Mientras que el modelo de menor precisión se obtuvo con 26 años (1988-2013), el EMAP fue 32,94% (precisión Pobre) y el sesgo fue 39,54 (Figura 102b).

**Figura 102**

*CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. PBIman e INF1 con Extensión de (a) 16 Años (1998-2013) y (b) 26 Años (1988-2013)*



#### 4.2.2.4 Modelos PW: L-CONG vs. PBI. El pronóstico del modelo PW: L-CONG vs.

PBI (1980-2013) tiene un EMAP de 2,66% y un sesgo de 1,65 (Tabla 28). En 2014, el EAP y

el error son muy pequeños (0,72% y -0,63, respectivamente), al año siguiente alcanzan su máximo con un EAP de 6,26% y un error de 6,66, para luego mostrar una tendencia a la baja; los mínimos ocurren en 2018 con un EAP de 0,59% y un error de 0,82 (Tabla 28). La evolución del EAP se observa en la superposición del CONG real y el pronosticado (2014, 2017 y 2018), mientras que en los años 2015-2016 la diferencia es reducida; y el sesgo pequeño se aprecia cuando el CONG pronosticado, en algunos años, sobrestima al CONG real y en otros lo subestima (Figura 103).

**Tabla 28**

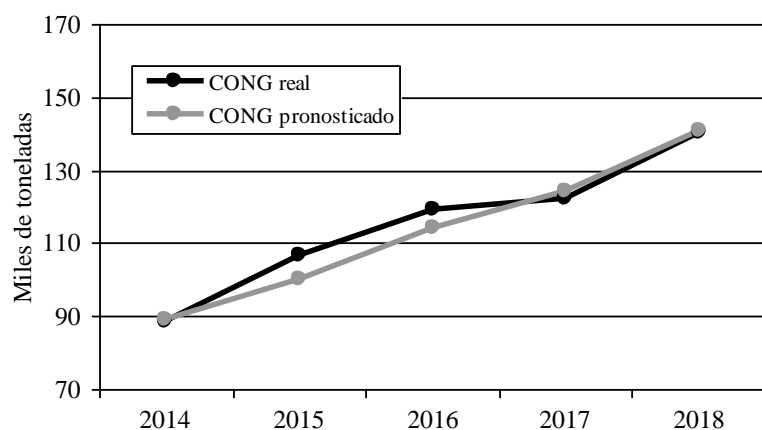
*Modelo PW: L-CONG vs. PBI (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018)*

Año	CONG real (miles de t)	CONG pronosticado		Error absoluto porcentual	Error
		(Logaritmo)	(miles de t)		
2014	88,6	4,4913	89,2	0,72%	-0,63
2015	106,9	4,6079	100,3	6,23%	6,66
2016	119,3	4,7389	114,3	4,19%	4,99
2017	122,2	4,8214	124,1	1,59%	-1,94
2018	140,2	4,9489	141,0	0,59%	-0,82
				2,66%	1,65
				EMAP	Sesgo

*Nota:* Fuente PRODUCE, Software Gretl 2019a.

**Figura 103**

*Modelo PW: L-CONG vs PBI (1980-2013). CONG Real y Pronosticado (2014-2018)*



A continuación, en la Tabla 29 se muestran los 20 modelos PW: L-CONG vs. PBI que se lograron validar, de ellos 16 presentaron precisión Muy buena ( $EMAP \leq 10\%$ ) y cuatro exhibieron precisión Buena ( $10\% < EMAP \leq 20\%$ ). El EMAP y el sesgo, de los 19 modelos adicionales, se calculó según lo mostrado en la Tabla 28.

**Tabla 29**

*Modelos PW: L-CONG vs. PBI. EMAP, Precisión, Sesgo (2014-2018) y Merma Según*

*Extensión de las Series Temporales*

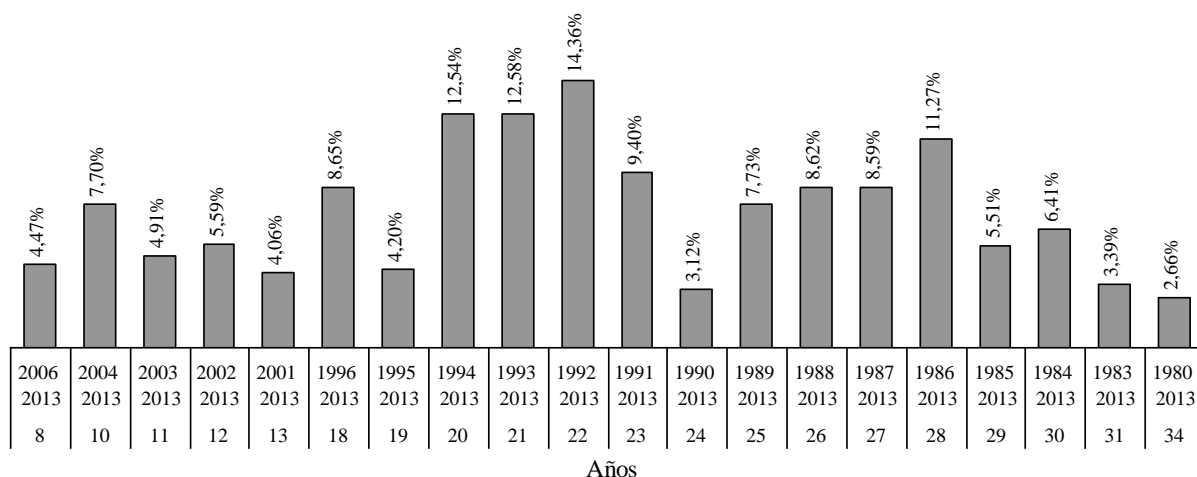
Nº	Periodo	Extensión (Años)	EMAP	Precisión	Sesgo	$r^2$	$r_c^2$	Merma
1	2006-2013	8	4,47%	Muy buena	-4,96	66,48%	60,89%	5,59%
2	2004-2013	10	7,70%	Muy buena	6,96	60,80%	55,90%	4,90%
3	2003-2013	11	4,91%	Muy buena	2,06	68,39%	64,88%	3,51%
4	2002-2013	12	5,59%	Muy buena	3,15	71,09%	68,20%	2,89%
5	2001-2013	13	4,06%	Muy buena	-0,71	75,78%	73,58%	2,20%
6	1996-2013	18	8,65%	Muy buena	-10,52	79,21%	77,91%	1,30%
7	1995-2013	19	4,20%	Muy buena	-4,21	78,94%	77,70%	1,24%
8	1994-2013	20	12,54%	Buena	-15,19	79,19%	78,04%	1,16%
9	1993-2013	21	12,58%	Buena	-15,25	80,81%	79,80%	1,01%
10	1992-2013	22	14,36%	Buena	-17,40	82,34%	81,45%	0,88%
11	1991-2013	23	9,40%	Muy buena	-11,32	80,67%	79,75%	0,92%
12	1990-2013	24	3,12%	Muy buena	3,14	74,38%	73,22%	1,16%
13	1989-2013	25	7,73%	Muy buena	9,24	74,65%	73,55%	1,10%
14	1988-2013	26	8,62%	Muy buena	10,30	74,62%	73,56%	1,06%
15	1987-2013	27	8,59%	Muy buena	10,25	74,36%	73,34%	1,03%
16	1986-2013	28	11,27%	Buena	13,51	73,26%	72,23%	1,03%
17	1985-2013	29	5,51%	Muy buena	6,57	72,44%	71,42%	1,02%
18	1984-2013	30	6,41%	Muy buena	7,65	72,86%	71,89%	0,97%
19	1983-2013	31	3,39%	Muy buena	-2,90	71,14%	70,14%	1,00%
20	1980-2013	34	2,66%	Muy buena	1,65	74,47%	73,67%	0,80%

Seguidamente, en la Figura 104 se observa, para los modelos PW: L-CONG vs. PBI, la evolución del EMAP en relación a la extensión de las series. Por un lado, se muestra una relación negativa para series de entre 19 (1995-2013) y 22 años (1992-2013), observándose la

misma relación para series de entre 24 (1990-2013) y 28 años (1986-2013). Por otro lado, la relación es positiva para extensiones de entre 29 (1985-2013) y 31 años (1983-2013).

### Figura 104

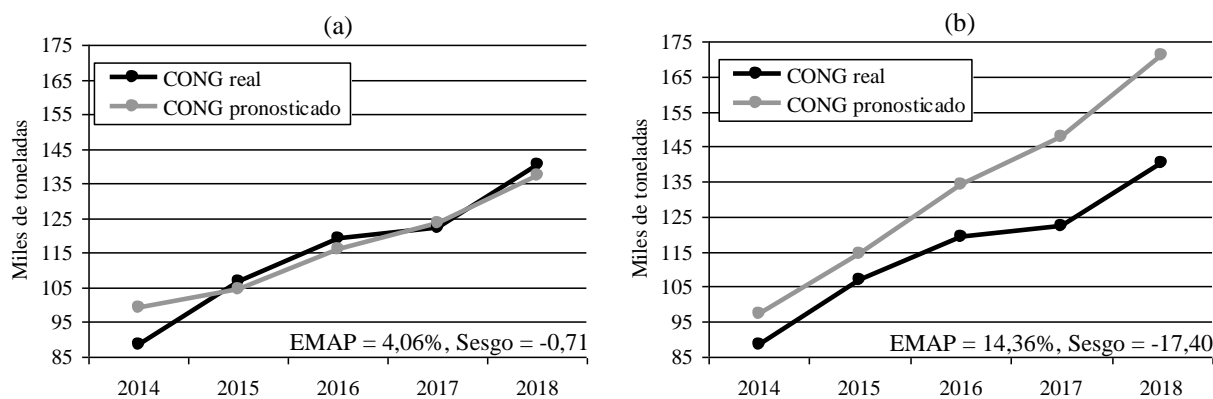
Modelos PW: L-CONG vs. PBI. EMAP (2014-2018) Según Extensión de las Series Temporales



En suma, de todos los modelos PW: L-CONG vs. PBI, el de mayor precisión fue el construido con 13 años (2001-2013), el EMAP fue 4,06% (precisión Muy buena) y el sesgo fue -0,71 (Figura 105a). Mientras que el modelo de menor precisión se obtuvo con 22 años (1992-2013), el EMAP fue 14,36% (precisión Buena) y el sesgo fue -17,40 (Figura 105b).

### Figura 105

CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. PBI con Extensión de (a) 13 Años (2001-2013) y (b) 22 Años (1992-2013)



**4.2.2.5 Modelos PW: L-CONG vs. PBI e INF1.** El pronóstico del modelo PW: L-CONG vs. PBI e INF1 (1980-2013) presenta un EMAP de 9,70% y un sesgo de -11,80 (Tabla 30). Los mínimos se observan en 2015 con un EAP de 3,07% y un error de -3,28; mientras que los máximos ocurren en 2018 con 15,57% y -21,83, respectivamente (Tabla 30). El comportamiento del EAP se aprecia en la pequeña distancia entre el CONG real y el pronosticado (2014-2016), para luego mostrar una tendencia creciente; y el error negativo se visualiza cuando el CONG pronosticado siempre sobrestima al CONG real (Figura 106).

**Tabla 30**

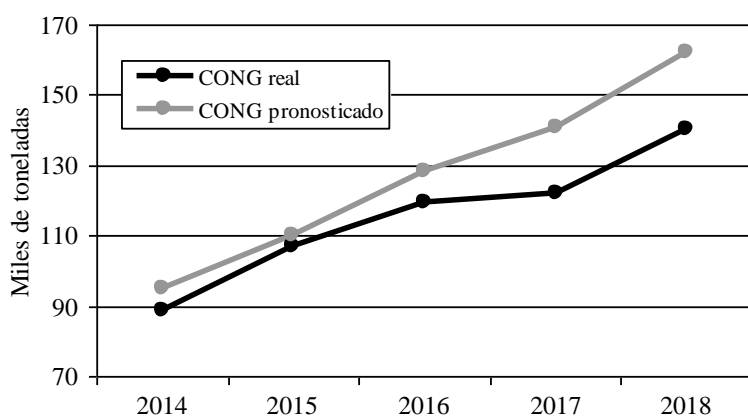
*Modelo PW: L-CONG vs. PBI e INF1 (1980-2013). Pronóstico Fuera de la Muestra, EMAP y Sesgo (2014-2018)*

Año	CONG real (miles de t)	CONG pronosticado		Error absoluto porcentual	Error
		(Logaritmo)	(miles de t)		
2014	88,6	4,5533	94,9	7,16%	-6,35
2015	106,9	4,7024	110,2	3,07%	-3,28
2016	119,3	4,8541	128,3	7,51%	-8,96
2017	122,2	4,9470	140,8	15,18%	-18,55
2018	140,2	5,0878	162,0	15,57%	-21,83
				9,70%	-11,80
				EMAP	Sesgo

*Nota:* Fuente PRODUCE, Software Gretl 2019a.

**Figura 106**

*Modelo PW: L-CONG vs. PBI e INF1 (1980-2013). CONG Real y Pronosticado (2014-2018)*



A continuación, la Tabla 31 muestra que se lograron validar 11 modelos PW: L-CONG vs. PBI e INF1, de los cuales ocho presentaron precisión Muy buena ( $EMAP \leq 10\%$ ) y tres exhibieron precisión Buena ( $10\% < EMAP \leq 20\%$ ). El EMAP y el sesgo, de los 10 modelos adicionales, se calculó según lo mostrado en la Tabla 30.

**Tabla 31**

*Modelos PW: L-CONG vs. PBI e INF1. EMAP, Precisión, Sesgo (2014-2018) y Merma*

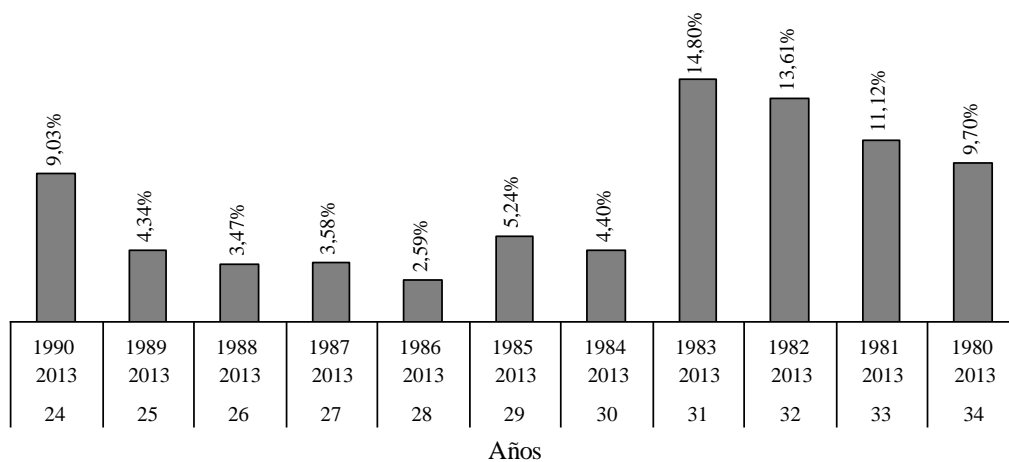
*Según Extensión de las Series Temporales*

Nº	Periodo	Extensión (Años)	EMAP	Precisión	Sesgo	R <sup>2</sup>	R <sub>c</sub> <sup>2</sup>	Merma
1	1990-2013	24	9,03%	Muy buena	-10,91	80,94%	79,13%	1,81%
2	1989-2013	25	4,34%	Muy buena	-4,66	79,88%	78,05%	1,83%
3	1988-2013	26	3,47%	Muy buena	-3,01	79,87%	78,12%	1,75%
4	1987-2013	27	3,58%	Muy buena	-3,21	79,99%	78,32%	1,67%
5	1986-2013	28	2,59%	Muy buena	0,80	78,51%	76,79%	1,72%
6	1985-2013	29	5,24%	Muy buena	-6,30	78,02%	76,33%	1,69%
7	1984-2013	30	4,40%	Muy buena	-4,89	78,28%	76,67%	1,61%
8	1983-2013	31	14,80%	Buena	-17,60	76,48%	74,80%	1,68%
9	1982-2013	32	13,61%	Buena	-16,40	77,45%	75,90%	1,56%
10	1981-2013	33	11,12%	Buena	-13,51	78,50%	77,06%	1,43%
11	1980-2013	34	9,70%	Muy buena	-11,80	78,70%	77,33%	1,37%

A continuación, la Figura 107 muestra, para los modelos PW: L-CONG vs. PBI e INF1, el comportamiento del EMAP con respecto a la extensión de las series. En general, se aprecia una relación negativa entre precisión y extensión. Sin embargo, tras el análisis, se observa una relación positiva para extensiones de 24 (1990-2013) y 28 años (1986-2013). Luego, en la serie de 31 años (1983-2013) ocurre un incremento abrupto del EMAP (14,80%). Seguidamente la relación es positiva hasta la serie de 34 años (1980-2013).

**Figura 107**

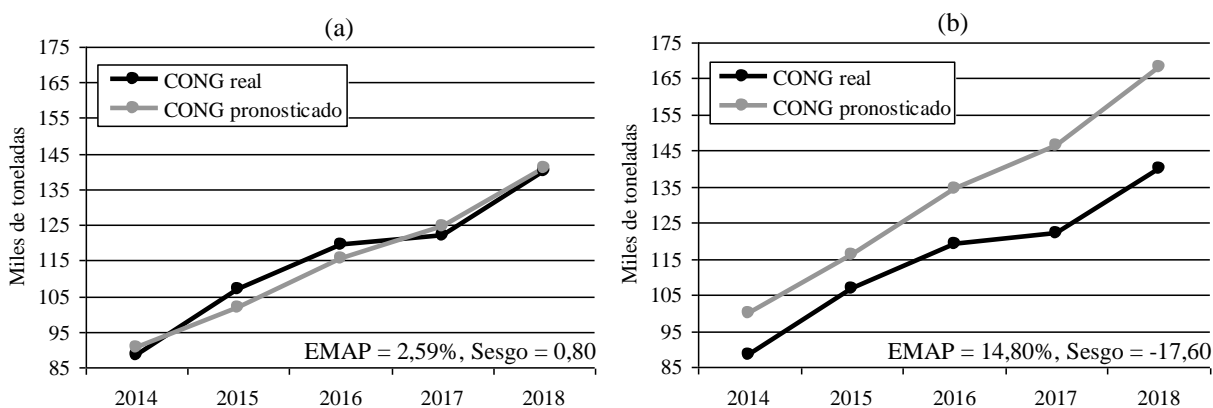
*Modelos PW: L-CONG vs. PBI e INF1. EMAP (2014-2018) Según Extensión de las Series Temporales*



De todos los modelos PW: L-CONG vs. PBI e INF1, el mayor precisión fue el construido con 28 años (1986-2013), el EMAP fue 2,59% (precisión Muy buena) y el sesgo fue 0,80 (Figura 108a). Mientras que el modelo de menor precisión se obtuvo con 31 años (1983-2013), el EMAP fue 14,80% (precisión Buena) y el sesgo fue -17,60 (Figura 108b).

**Figura 108**

*CONG Real y Pronosticado (2014-2018) de los Modelos PW: L-CONG vs. PBI e INF1 con Extensión de (a) 28 Años (1986-2013) y (b) 31 Años (1983-2013)*



A manera de resumen, la Tabla 32 presenta una selección de modelos PW con EMAP menor e igual a 5%, así como también el sesgo y la extensión de las series por cada variable independiente. La Tabla muestra los 16 modelos que cumplieron con la condición previa, es

decir, los modelos que exhibieron un EMAP menor e igual a 5% fueron dos modelos PW: L-CONG vs. DI, dos modelos PW: L-CONG vs. DI e INF1, siete modelos PW: L-CONG vs. PBI y cinco modelos PW: L-CONG vs. PBI e INF1.

En la construcción de los modelos se utilizaron entre ocho años (modelo N° 10) y 34 años (modelo N° 5) por cada variable independiente.

**Tabla 32**

*EMAP, Sesgo y Extensión por Variable Independiente de Modelos PW Seleccionados*

N°	Variable dependiente	Variables independientes	Periodo	Extensión (Años)	EMAP	Sesgo	Extensión por variable independiente (Años)
1	L-CONG	DI	1992-2013	22	3,69%	2,64	22
2			<b>1999-2013</b>	<b>15</b>	<b>3,84%</b>	<b>-0,75</b>	<b>15</b>
3	L-CONG	DI - INF1	<b>1982-2013</b>	<b>32</b>	<b>2,63%</b>	<b>-1,23</b>	<b>16</b>
4			1981-2013	33	4,37%	3,67	17
5			1980-2013	34	2,66%	1,65	34
6			1990-2013	24	3,12%	3,14	24
7			1983-2013	31	3,39%	-2,90	31
8	L-CONG	PBI	<b>2001-2013</b>	<b>13</b>	<b>4,06%</b>	<b>-0,71</b>	<b>13</b>
9			1995-2013	19	4,20%	-4,21	19
10			2006-2013	8	4,47%	-4,96	8
11			2003-2013	11	4,91%	2,06	11
12			<b>1986-2013</b>	<b>28</b>	<b>2,59%</b>	<b>0,80</b>	<b>14</b>
13			1988-2013	26	3,47%	-3,01	13
14	L-CONG	PBI - INF1	1987-2013	27	3,58%	-3,21	14
15			1989-2013	25	4,34%	-4,66	13
16			1984-2013	30	4,40%	-4,89	15

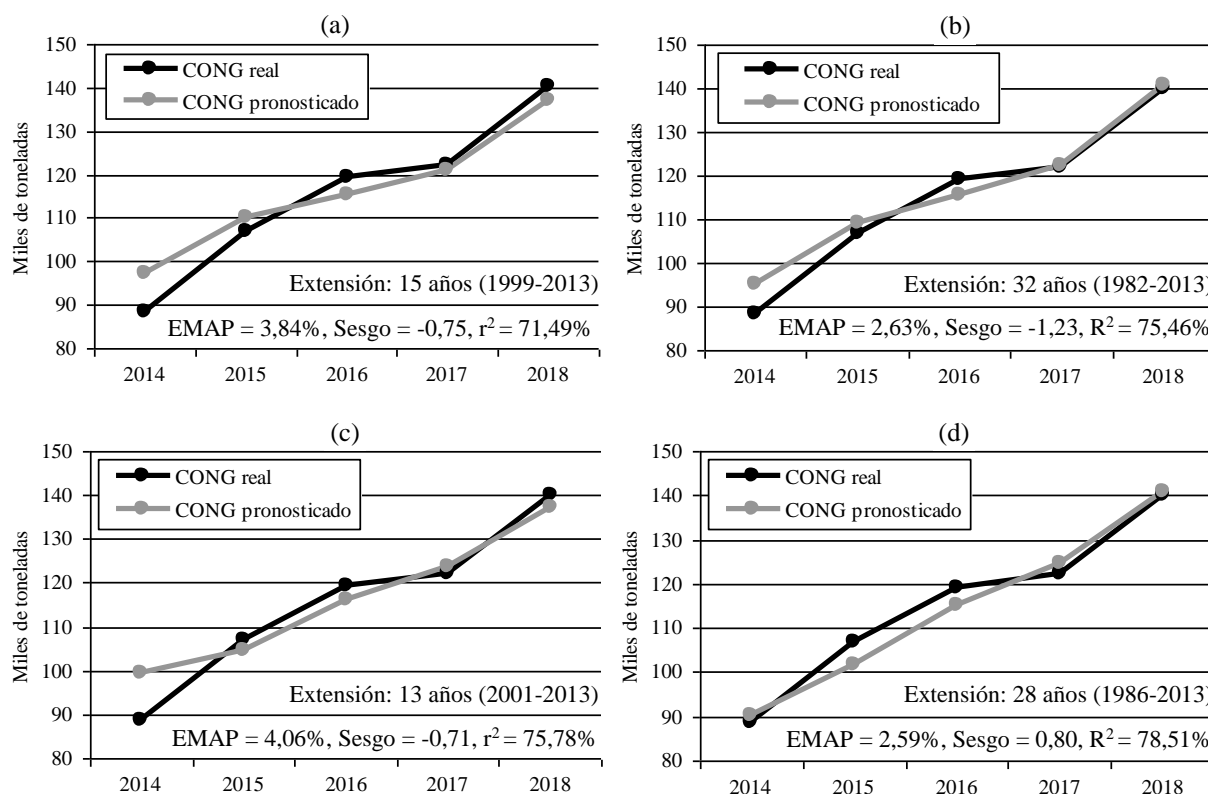
*Nota:* Los modelos con sesgo más pequeño están en negritas.

Finalmente, la Figura 109 muestra la cercanía y el insesgamiento entre el CONG real y el pronosticado (2014-2018) para los cuatro modelos más precisos, en términos de EMAP y sesgo, seleccionados de la Tabla 32 (Modelos N° 2, 3, 8 y 12). Así, el modelo N° 2 (PW: L-CONG vs. DI) se construyó con series temporales de 15 años de extensión, obteniéndose un EMAP de 3,84% y un sesgo de -0,75 (Figura 109a). Por su parte, para el modelo N° 3 (PW:

L-CONG vs. DI e INF1) se utilizaron series de 16 años por cada variable independiente, resultando un EMAP de 2,63% y un sesgo de -1,23 (Figura 109b). Seguidamente, el modelo N° 8 (PW: L-CONG vs. PBI) se construyó con 13 años, alcanzándose un EMAP de 4,06% y un sesgo de -0,71 (Figura 109c). Y el modelo N° 8 (PW: L-CONG vs. PBI e INF1) se elaboró con series de 14 años por cada variable independiente, lográndose un EMAP de 2,59% y un sesgo de 0,80 (Figura 109d).

### Figura 109

*CONG Real y Pronosticado (2014-2018) con los Modelos PW (a) L-CONG vs. DI (1999-2013), (b) L-CONG vs. DI e INF1 (1982-2013), (c) L-CONG vs. PBI (2001-2013) y (d) L-CONG vs. PBI e INF1 (1986-2013)*



#### 4.2.3 Número de Variables Independientes y Precisión del Pronóstico

De los modelos PW validados en el numeral 4.2.2 (p. 142), se seleccionaron a los modelos L-CONG vs. DI (Tabla 23, p. 144) y a sus versiones aumentadas L-CONG vs. DI e INF1 (Tabla 25, p. 147) y a los modelos L-CONG vs. PBI (Tabla 29, p. 153) y a sus

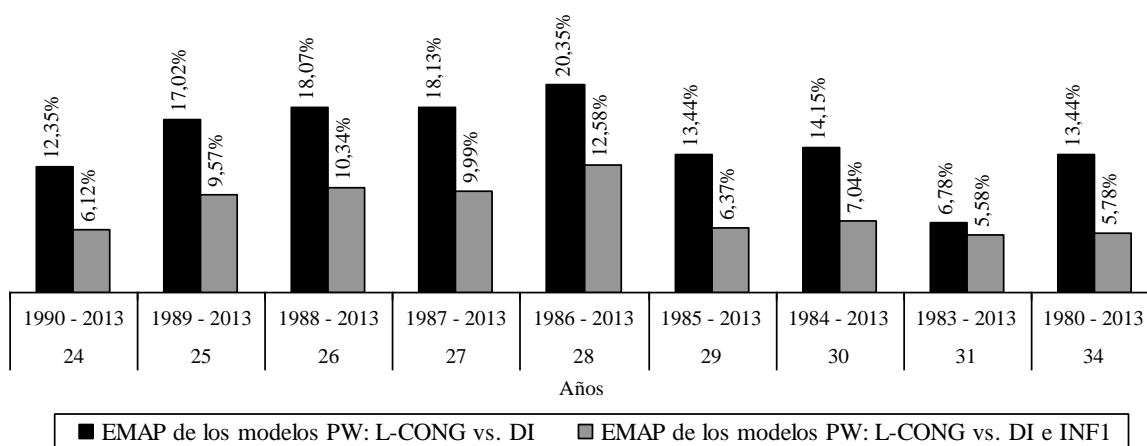
versiones ampliadas L-CONG vs. PBI e INF1 (Tabla 31, p. 156). Los modelos se construyeron con extensiones de 24, 25, 26, 27, 28, 29, 30, 31 y 34 años, respectivamente.

#### 4.2.3.1 Modelo PW: L-CONG vs. DI Aumentado a PW: L-CONG vs. DI e INF1.

En la Figura 110 se observa que, cuando los modelos PW: L-CONG vs. DI son aumentados a los modelos PW: L-CONG vs. DI e INF1, el EMAP decrece y, por tanto, la precisión se incrementa. Es decir, en estos modelos la relación es positiva entre el número de variables independientes y la precisión del pronóstico de demanda.

**Figura 110**

*Modelos PW: L-CONG vs. DI y PW: L-CONG vs. DI e INF1. EMAP (2014-2018) Según Número de Variables Independientes*

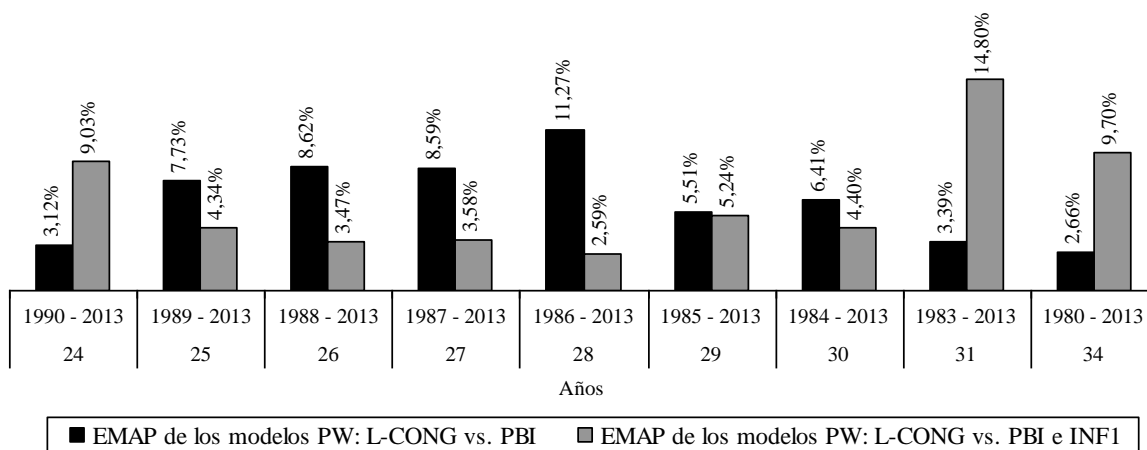


#### 4.2.3.2 Modelo PW: L-CONG vs. PBI Aumentado a PW: L-CONG vs. PBI e INF1.

En la Figura 111 se observa que, cuando los modelo PW: L-CONG vs. PBI son ampliados a los modelos PW: L-CONG vs. PBI e INF1, el EMAP presenta resultados mixtos. No obstante, en seis casos de nueve, la relación es positiva entre el número de variables independientes y la precisión del pronóstico de demanda.

**Figura 111**

*Modelos PW: L-CONG vs. PBI y PW: L-CONG vs. PBI e INF1. EMAP (2014-2018) Según Número de Variables Independientes*



#### **4.2.4 Modelo de Regresión Lineal Validado y Precisión del Pronóstico**

La Tabla 33 presenta EMAP, precisión, sesgo y merma tanto de los modelos validados PW con variable dependiente L-CONG (1980-2013) (Tabla 21, p. 139), como los de los modelos no validados MCO con variable dependiente CONG (1980-2013). Además, debido a que en los pronósticos de demanda, es frecuente el uso del tiempo como variable independiente, se ha incluido un modelo MCO con dicha variable (Modelo N° 1). La Tabla muestra como sistemáticamente, el EMAP y el sesgo, son menores en los modelos validados. También se observa como los modelos no validados 4, 8 y 10 presentan  $R^2$  superiores a 80%.

**Tabla 33**

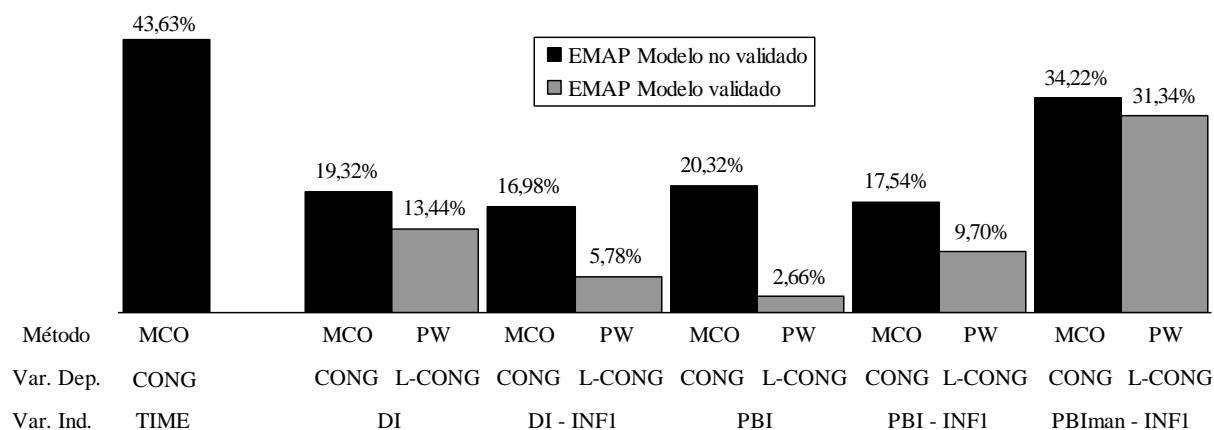
*EMAP, Precisión, Sesgo (2014-2018) y Poder Explicativo de Modelos No Validados y de Modelos Validados (1980-2013)*

N°	Método	Variable Dependiente	Variables Independientes	Supuestos violados	EMAP	Precisión	Sesgo	R <sup>2</sup>	R <sub>c</sub> <sup>2</sup>
1	MCO	CONG	TIME	1, 4, 7, 8	43,63%	Pobre	51,51	55,45%	54,06%
2	MCO	CONG	DI	4, 6, 7	19,32%	Buena	23,94	73,59%	72,76%
3	PW	L-CONG	DI	-----	13,44%	Buena	16,44	73,08%	72,24%
4	MCO	CONG	DI - INF1	6, 7, 8	16,98%	Buena	20,82	80,65%	79,40%
5	PW	L-CONG	DI - INF1	-----	5,78%	Muy buena	5,92	76,87%	75,38%
6	MCO	CONG	PBI	1, 4, 6	20,32%	Moderada	24,78	75,01%	74,23%
7	PW	L-CONG	PBI	-----	2,66%	Muy buena	1,65	74,47%	73,67%
8	MCO	CONG	PBI - INF1	6, 7	17,54%	Buena	21,57	82,47%	81,34%
9	PW	L-CONG	PBI - INF1	-----	9,70%	Muy buena	-11,80	78,70%	77,33%
10	MCO	CONG	PBIman - INF1	6, 7	34,22%	Pobre	41,21	81,65%	80,47%
11	PW	L-CONG	PBIman - INF1	-----	31,34%	Pobre	37,64	79,06%	77,71%

A continuación, la Figura 112 muestra el EMAP de los modelos validados y el de los modelos no validados. En particular, el modelo MCO: CONG vs. TIME (1980-2013) presenta el mayor EMAP (43,63%). Y, en las comparaciones, los EMAPs de los modelos validados son menores al de los modelos no validados. Por ejemplo, el EMAP del modelo no validado MCO: CONG vs. PBI es 20,32% (precisión Moderada), mientras que el EMAP del modelo validado PW: L-CONG vs. PBI es 2,66% (precisión Muy buena).

**Figura 112**

*EMAP (2014-2018) de Modelos No Validados y de Modelos Validados (1980-2013)*



#### 4.2.4.1 Modelo MCO: CONG vs. TIME (1980-2013). El modelo no validado MCO:

CONG vs. TIME (1980-2013) presenta un EMAP de 43,63% (precisión Pobre) y un sesgo de 51,51 (Tabla 34), lo cual revela la poca utilidad del tiempo como única variable independiente, para fines de pronóstico.

**Tabla 34**

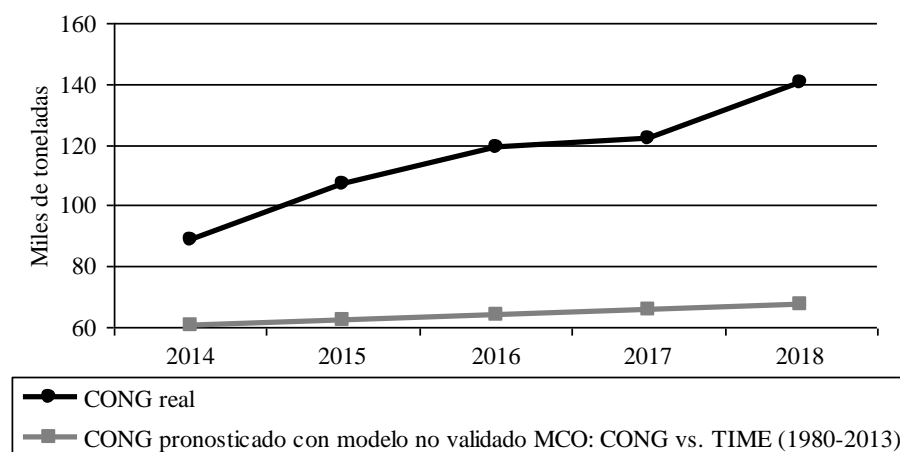
*EMAP y Sesgo (2014-2018) con el Modelo No Validado MCO: CONG vs. TIME (1980-2013)*

Año	CONG real (miles de t)	Modelo No Validado MCO: CONG vs. TIME (1980-2013) Viola supuestos 1, 4, 7, 8		
		CONG pronosticado	Error absoluto	Error
		(miles de t)	porcentual	
2014	88,6	60,5	31,69%	28,08
2015	106,9	62,2	41,80%	44,70
2016	119,3	63,9	46,41%	55,36
2017	122,2	65,6	46,28%	56,55
2018	140,2	67,4	51,96%	72,85
			43,63%	51,51
			EMAP	Sesgo

Los altos valores del EMAP y sesgo se visualizan en el distanciamiento, cada vez mayor, entre el CONG real y el CONG pronosticado; además, el signo positivo del sesgo indica que el CONG pronosticado subestima persistentemente al CONG real (Figura 113).

**Figura 113**

*CONG Real y Pronosticado (2014-2018) con el Modelo No validado MCO: CONG vs. TIME (1980-2013)*



**4.2.4.2 Modelos PW: L-CONG vs. DI (1980-2013) y MCO: CONG vs. DI (1980-2013).** El modelo validado PW: L-CONG vs. DI (1980-2013) presenta un EMAP de 13,44% (precisión Buena) y un sesgo de 16,44, mientras que el modelo no validado MCO: CONG vs. DI (1980-2013) tiene un EMAP de 19,32% (precisión Buena) y un sesgo de 23,94 (Tabla 35).

**Tabla 35**

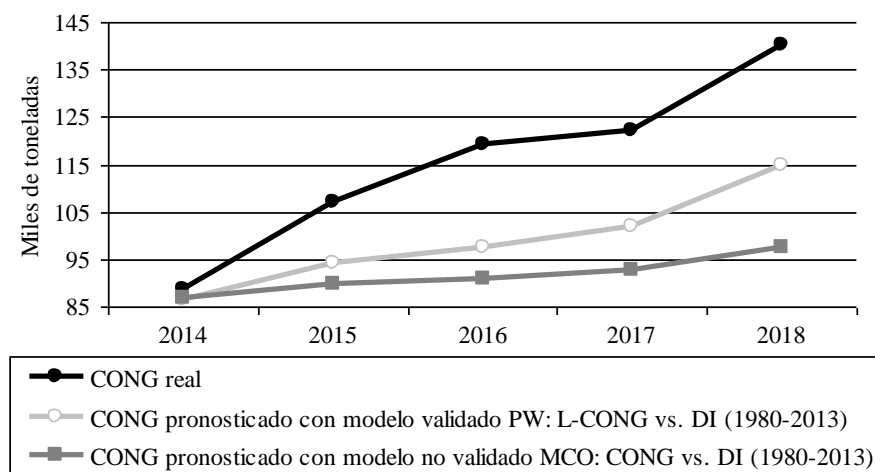
*EMAP y Sesgo (2014-2018) con el Modelo Validado PW: L-CONG vs. DI (1980-2013) y con el Modelo No Validado MCO: CONG vs. DI (1980-2013)*

Año	CONG real (miles de t)	Modelo Validado PW: L-CONG vs. DI (1980-2013)			Modelo No Validado MCO: CONG vs. DI (1980-2013) Viola supuestos 4, 6, 7		
		CONG pronosticado (miles de t)	Error absoluto porcentual	Error	CONG pronosticado (miles de t)	Error absoluto porcentual	Error
		2014	88,6	86,4	2,53%	2,24	86,7
2015	106,9	94,2	11,89%	12,71	89,8	16,01%	17,12
2016	119,3	97,7	18,11%	21,61	91,0	23,74%	28,32
2017	122,2	102,0	16,52%	20,19	92,6	24,23%	29,61
2018	140,2	114,7	18,17%	25,47	97,4	30,51%	42,77
			13,44%	16,44		19,32%	23,94
			EMAP	Sesgo		EMAP	Sesgo

La diferencia entre EMAP y sesgo se visualiza cuando, el CONG pronosticado con el modelo validado, está más próximo al CONG real que el CONG pronosticado con el modelo no validado; aunque ambos pronósticos se alejan cada vez más del CONG real (Figura 114).

**Figura 114**

*CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. DI (1980-2013) y con el Modelo No Validado MCO: CONG vs. DI (1980-2013)*



**4.2.4.3 Modelos PW: L-CONG vs. DI e INF1 (1980-2013) y MCO: CONG vs. DI e INF1 (1980-2013).** El modelo validado PW: L-CONG vs. DI e INF1 (1980-2013) presenta un EMAP de 5,78% (precisión Muy buena) y un sesgo de 5,92, por su parte el modelo no validado MCO: CONG vs. DI e INF1 (1980-2013) tiene un EMAP de 16,98% (precisión Buena) y un sesgo de 20,82 (Tabla 36). La diferencia entre EMAP y sesgo se visualiza cuando el CONG pronosticado con el modelo validado está muy cerca al CONG real y, por el contrario, el CONG pronosticado con el modelo no validado se aleja persistentemente del CONG real (Figura 115).

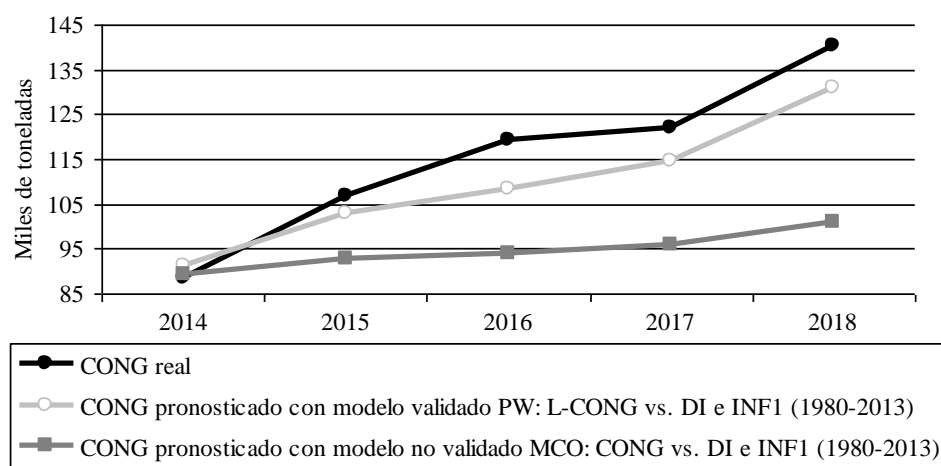
**Tabla 36**

*EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. DI e INF1 (1980-2013) y del Modelo No Validado MCO: CONG vs. DI e INF1 (1980-2013)*

Año	CONG real (miles de t)	Modelo Validado PW: L-CONG vs. DI e INF1 (1980-2013)			Modelo No Validado MCO: CONG vs. DI e INF1 (1980-2013) Viola supuestos 6, 7, 8		
		CONG pronosticado (miles de t)	Error absoluto porcentual	Error	CONG pronosticado (miles de t)	Error absoluto porcentual	Error
		2014	88,6	91,2	2,90%	-2,57	89,5
2015	106,9	102,9	3,81%	4,07	92,8	13,21%	14,13
2016	119,3	108,4	9,14%	10,90	94,1	21,16%	25,24
2017	122,2	114,4	6,36%	7,77	95,8	21,61%	26,41
2018	140,2	130,8	6,71%	9,40	101,0	27,95%	39,18
			5,78%	5,92		16,98%	20,82
			EMAP	Sesgo		EMAP	Sesgo

**Figura 115**

*CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. DI e INF1 (1980-2013) y con el Modelo No Validado MCO: CONG vs. DI e INF1 (1980-2013)*



#### 4.2.4.4 Modelos PW: L-CONG vs. PBI (1980-2013) y MCO: CONG vs. PBI

(1980-2013). El modelo validado PW: L-CONG vs. PBI (1980-2013) presenta un EMAP de 2,66% (precisión Muy buena) y un sesgo de 1,65, mientras que el modelo no validado MCO: CONG vs. PBI (1980-2013) tiene un EMAP de 20,32% (precisión Moderada) y un sesgo de 24,78 (Tabla 37).

**Tabla 37**

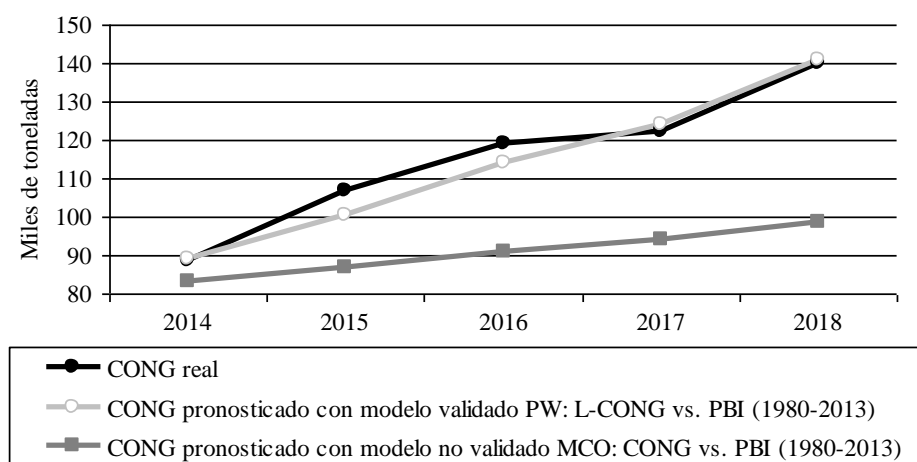
*EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. PBI (1980-2013) y del Modelo No Validado MCO: CONG vs. PBI (1980-2013)*

Año	CONG real (miles de t)	Modelo Validado PW: L-CONG vs. PBI (1980-2013)			Modelo No Validado MCO: CONG vs. PBI (1980-2013) Viola supuestos 1, 4, 6		
		CONG pronosticado (miles de t)	Error absoluto porcentual	Error	CONG pronosticado (miles de t)	Error absoluto porcentual	Error
		2014	88,6	89,2	0,72%	-0,63	83,2
2015	106,9	100,3	6,23%	6,66	86,6	18,98%	20,29
2016	119,3	114,3	4,19%	4,99	91,1	23,66%	28,23
2017	122,2	124,1	1,59%	-1,94	93,9	23,16%	28,30
2018	140,2	141,0	0,59%	-0,82	98,6	29,69%	41,63
			2,66%	1,65		20,32%	24,78
		EMAP	Sesgo		EMAP	Sesgo	

La diferencia del EMAP y sesgo se aprecia en que el CONG pronosticado con el modelo validado está extremadamente cerca al CONG real, en contraste, el CONG pronosticado con el modelo no validado se aleja sistemáticamente del CONG real (Figura 116).

**Figura 116**

*CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. PBI (1980-2013) y con el Modelo No Validado MCO: CONG vs. PBI (1980-2013)*



**4.2.4.5 Modelos PW: L-CONG vs. PBI e INF1 (1980-2013) y MCO: CONG vs. PBI e INF1 (1980-2013).** El modelo validado PW: L-CONG vs. PBI e INF1 (1980-2013) presenta un EMAP de 9,70% (precisión Muy buena) y un sesgo de -11,80, a su vez el modelo no validado MCO: CONG vs. PBI e INF1 (1980-2013) tiene un EMAP de 17,54% (precisión Buena) y un sesgo de 21,57 (Tabla 38).

**Tabla 38**

*EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. PBI e INF1 (1980-2013) y del Modelo No Validado MCO: CONG vs. PBI e INF1 (1980-2013)*

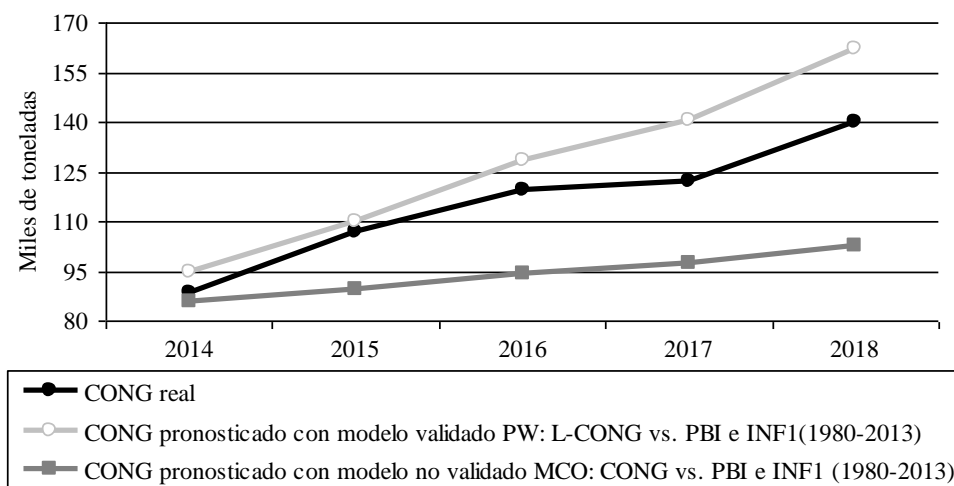
Año	CONG real (miles de t)	Modelo Validado PW: L-CONG vs. PBI e INF1 (1980-2013)			Modelo No Validado MCO: CONG vs. PBI e INF1 (1980-2013) Viola supuestos 6, 7		
		CONG pronosticado (miles de t)	Error absoluto porcentual	Error	CONG pronosticado (miles de t)	Error absoluto porcentual	Error
		2014	88,6	94,9	7,16%	-6,35	85,7
2015	106,9	110,2	3,07%	-3,28	89,5	16,29%	17,42
2016	119,3	128,3	7,51%	-8,96	94,3	20,94%	24,99
2017	122,2	140,8	15,18%	-18,55	97,4	20,32%	24,83
2018	140,2	162,0	15,57%	-21,83	102,4	26,94%	37,77
		9,70%		-11,80	17,54%		21,57
		EMAP		Sesgo	EMAP		Sesgo

La diferencia del EMAP y sesgo se debe a que el CONG pronosticado con el modelo validado está más cerca al CONG real, en contraste, el CONG pronosticado con el modelo no validado se distancia cada vez más del CONG real (Figura 117).

**Figura 117**

*CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. PBI e*

*INF1 (1980-2013) y con el Modelo No Validado MCO: CONG vs. PBI e INF1 (1980-2013)*



#### 4.2.4.6 Modelos PW: L-CONG vs. PBIman e INF1 (1980-2013) y MCO: CONG

**vs. PBIman e INF1 (1980-2013).** El modelo validado PW: L-CONG vs. PBIman e INF1 (1980-2013) presenta un EMAP de 31,34% (precisión Pobre) y un sesgo de 37,64, mientras que el modelo no validado MCO: CONG vs. PBIman e INF1 (1980-2013) tiene un EMAP de 34,22% (precisión Pobre) y un sesgo de 41,21 (Tabla 39).

**Tabla 39**

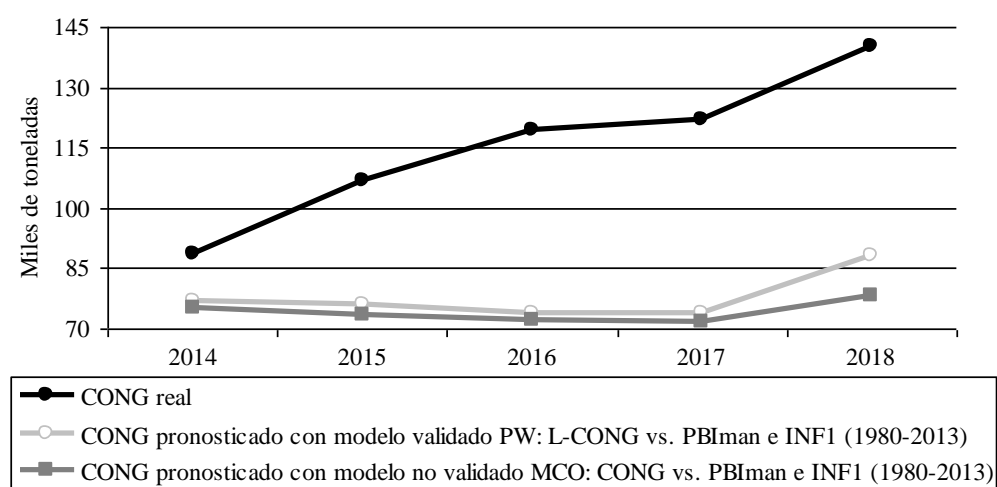
*EMAP y Sesgo (2014-2018) del Modelo Validado PW: L-CONG vs. PBIman e INF1 (1980-2013) y del Modelo No Validado MCO: CONG vs. PBIman e INF1 (1980-2013)*

Año	CONG real (miles de t)	Modelo Validado PW: L-CONG vs. PBIman e INF1 (1980-2013)			Modelo No Validado MCO: CONG vs. PBIman e INF1 (1980-2013) Viola supuestos 6, 7		
		CONG pronosticado (miles de t)	Error absoluto porcentual	Error	CONG pronosticado (miles de t)	Error absoluto porcentual	Error
		2014	88,6	77,0	13,14%	11,64	75,2
2015	106,9	76,0	28,95%	30,96	73,6	31,16%	33,32
2016	119,3	73,9	38,04%	45,38	72,1	39,54%	47,17
2017	122,2	73,8	39,62%	48,41	71,9	41,19%	50,33
2018	140,2	88,4	36,95%	51,80	78,4	44,10%	61,83
			31,34%	37,64		34,22%	41,21
			EMAP	Sesgo		EMAP	Sesgo

En este caso, las diferencias entre EMAP y sesgo son pequeñas, el CONG pronosticado de ambos modelos, está prácticamente a la misma distancia del CONG real, excepto en el último año (Figura 118).

### Figura 118

*CONG Real y Pronosticado (2014-2018) con el Modelo Validado PW: L-CONG vs. PBIman e INF1 (1980-2013) y con el Modelo No Validado MCO: CONG vs. PBIman e INF1 (1980-2013)*



## V Discusión de Resultados

### 5.1 Extensión de las Series Temporales y Precisión del Pronóstico

Muchos investigadores han afirmado que la precisión de los pronósticos mejora cuando se incrementa la extensión de las series temporales (Armstrong, 2002; Groebner et al., 2018; Wheelen et al., 2018).

No obstante, en lo referente a la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú, el resultado de la prueba de hipótesis revela que la extensión de las series temporales no tiene relación con la precisión del pronóstico. Aunque, la relación positiva se cumple para ciertos rangos de años, mientras que para otros la relación se torna negativa. Por ejemplo, para el modelo PW: L-CONG vs. DI (Figura 95, p. 145) la relación es positiva para extensiones de entre 10 años (2004-2013) y 15 años (1999-2013) y la relación es negativa cuando la extensión varía de 22 años (1992-2013) a 28 años (1986-2013). Y, en el caso del modelo PW: L-CONG vs. PBI (Figura 104, p. 154), la relación es positiva cuando la extensión varía de 10 años (2004-2013) a 13 años (2001-2013) y cuando varía de 30 años (1992-2013) a 31 años (1986-2013). Y, por el contrario, la relación es negativa para extensiones de entre 19 años (1995-2013) y 22 años (1992-2013) y para extensiones de entre 24 años (1990-2013) y 28 años (1986-2013).

Además, aunque no hay acuerdo en cuanto al número de datos a emplear por cada variable independiente, muchos autores han sugerido usar al menos 10 datos (Behar y Grima, 2018; Doane y Seward, 2016; Hanke y Wichern, 2014; Wilson y Keating, 2009). No obstante, Chase (2013) propuso usar cinco datos y Keith (2019) recomendó utilizar entre 10 y 20 datos.

Sin embargo, en lo concerniente a la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú, los cuatro modelos de regresión lineal de mayor precisión, en términos de EMAP (2014-2018) y sesgo, se construyeron utilizando

series temporales de entre 13 y 16 años de extensión por cada variable independiente (Figura 109, p. 159). Estos resultados son coherentes por lo recomendado por Keith (2019).

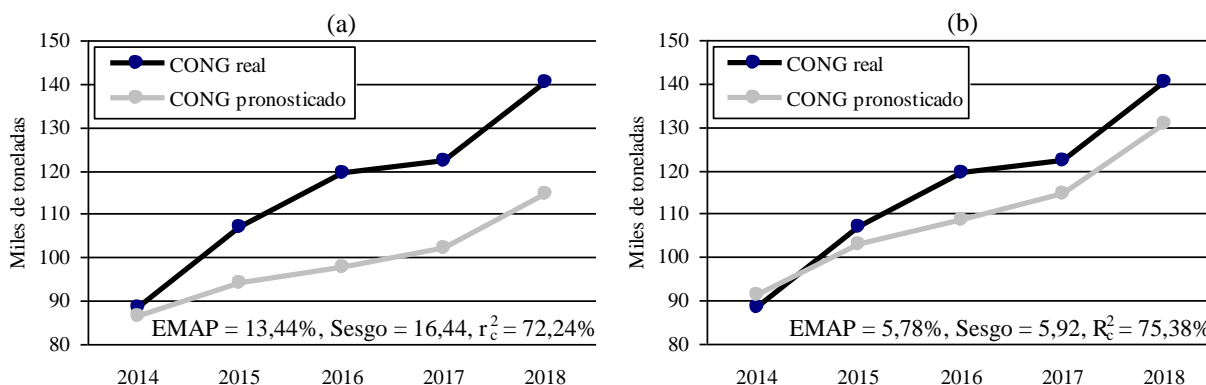
## 5.2 Número de Variables Independientes y Precisión del Pronóstico

Varios autores han coincidido en que los pronósticos obtenidos a partir de un modelo de regresión lineal múltiple son más precisos, comparados con los obtenidos de una regresión lineal simple (Anderson et al., 2017; Baca, 2013; Carlberg, 2016; Levine et al., 2017; Lilja, 2017).

En cuanto al pronóstico de demanda de productos hidrobiológicos congelados en el Perú, el resultado de la prueba de hipótesis revela que, efectivamente, el número de variables independientes tiene relación positiva con la precisión del pronóstico. Por ejemplo, cuando el modelo de regresión lineal simple PW: L-CONG vs. DI (1980-2013) es ampliado al modelo múltiple PW: L-CONG vs. DI e INF1 (1980-2013), el EMAP (2014-2018) decrece de 13,44% (precisión Buena) a 5,78% (precisión Muy buena); la misma tendencia se aprecia en el sesgo que cae de 16,44 a 5,92 (Figura 119).

### Figura 119

*CONG Real y Pronosticado (2014-2018) (a) con el Modelo Simple PW: L-CONG vs. DI (1980-2013) y (b) con el Modelo Aumentado PW: L-CONG vs. DI e INF1 (1980-2013)*

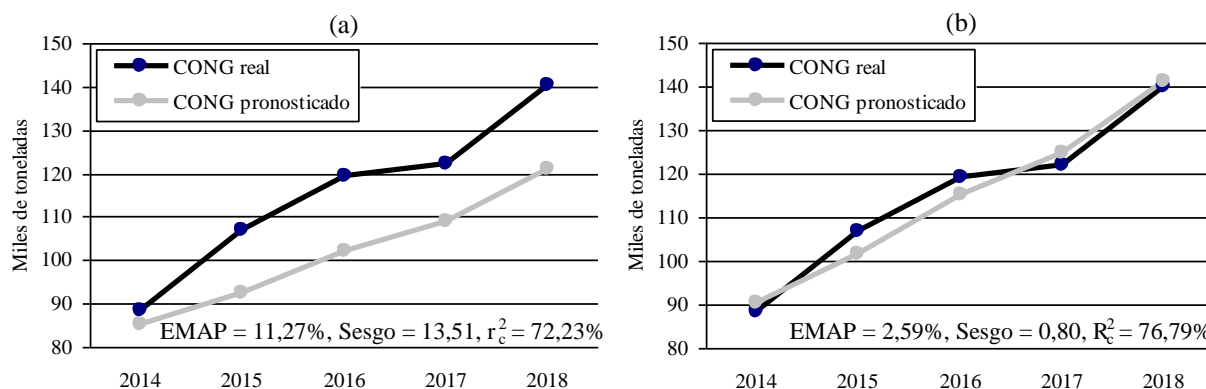


De igual manera, cuando el modelo de regresión lineal simple PW: L-CONG vs. PBI (1986-2013) es aumentado al modelo múltiple PW: L-CONG vs. PBI e INF1 (1986-2013), el

EMAP (2014-2018) disminuye de 11,27% (precisión Buena) a 2,59% (precisión Muy buena); igual tendencia se observa en el sesgo que decrece de 13,51 a 0,80 (Figura 120).

### Figura 120

*CONG Real y Pronosticado (2014-2018) (a) con el Modelo Simple PW: L-CONG vs. PBI (1986-2013) y (b) con el Modelo Aumentado PW: L-CONG vs. PBI e INF1 (1986-2013)*



### 5.3 Modelo de Regresión Lineal Validado y Precisión del Pronóstico

Un modelo de regresión lineal validado, es decir, un modelo que satisface los supuestos MCO y que es significativo, presenta pronósticos más precisos. Por un lado, los pronósticos pierden precisión si se violan los supuestos de ausencia de cambio estructural (Castle y Hendry, 2019; Ghysels y Marcellino, 2018; Marcellino, 2016), de linealidad (Young, 2017), de normalidad (Gujarati y Porter, 2010; Pedace, 2013; Welc y Rodriguez, 2018), de no autocorrelación y de homocedasticidad (Rosales et al., 2013; Welc y Rodriguez, 2018) y de especificación correcta (Welc y Rodriguez, 2018).

Por otro lado, también se pierde precisión cuando el modelo no es significativo (Lind et al., 2018), en ese caso, las variables independientes no son útiles para pronosticar la variable dependiente (Steele et al., 2016; Vu, 2015).

En este caso tampoco existe controversia, el resultado de la prueba de hipótesis revela que la precisión de los pronósticos de demanda de productos hidrobiológicos congelados en el Perú, obtenida de un modelo de regresión lineal validado es mayor frente a su contraparte

no validada. Por ejemplo, el modelo de regresión lineal validado PW: L-CONG vs. PBI (1980-2013) exhibe un EMAP de 2,66% (precisión Muy buena) y un sesgo de 1,65, y su contraparte, el modelo de regresión lineal no validado MCO: CONG vs. PBI (1980-2013), resulta con un EMAP de 20,32% (precisión Moderada) y un sesgo de 24,78 (Tabla 37 y Figura 116, p. 167).

Similarmente, el modelo de regresión lineal validado PW: L-CONG vs. DI e INF1 (1980-2013) presenta un EMAP de 5,78% (precisión Muy buena) y un sesgo de 5,92, y su contraparte, el modelo de regresión lineal no validado MCO: CONG vs. DI e INF1 (1980-2013), resulta con un EMAP de 16,98% (precisión Buena) y un sesgo de 20,82 (Tabla 36 y Figura 115, p. 166).

## VI Conclusiones

1. La extensión de las series temporales no se relaciona con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.
2. El número de variables independientes se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.
3. Un modelo validado presenta mayor precisión en el pronóstico de demanda de productos hidrobiológicos congelados en el Perú, frente a un modelo no validado.
4. Luego del resultado de las tres hipótesis específicas, es posible afirmar que el análisis de regresión lineal se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.

## VII Recomendaciones

1. En atención a lo expresado por Konasani y Kadre (2015), en cuanto a que el EMAP debe ser menor e igual a 5% y, tomando como base la clasificación del EMAP propuesta por Ghiani et al. (2013), se recomienda la siguiente clasificación del EMAP (Tabla 40):

**Tabla 40**

*Clasificación del EMAP Propuesta por el Autor*

Rango	Precisión
$0 < \text{EMAP} \leq 5\%$	Excelente
$5\% < \text{EMAP} \leq 10\%$	Muy buena
$10\% < \text{EMAP} \leq 15\%$	Buena
$15\% < \text{EMAP} \leq 20\%$	Moderada
$\text{EMAP} > 20\%$	Pobre

2. Debido a que es frecuente que los modelos de regresión lineal con series temporales violen el supuesto de no autocorrelación del error (Tabla 16, p. 133 y Tabla 20, p. 138), se debe empezar construyendo modelos PW. Por un lado, si un modelo PW es significativo y posee suficiente poder explicativo ( $r^2$  o  $R^2 \geq 0,7$ ), entonces construir la versión MCO del modelo para verificarlo y, eventualmente, validarlo. Por otro lado, si un modelo PW no es significativo y/o no posee suficiente poder explicativo, entonces el modelo debe ser descartado. Lo recomendado ahorrará tiempo valioso en el proceso de construcción de los modelos.
3. Construir, corregir y validar modelos regresión lineal, con series temporales muy extensas, no asegura pronósticos de gran precisión (EMAPs pequeños). Por lo cual se debe construir, corregir y validar un modelo base con toda la información disponible. Luego, el modelo base debe ser replicado, disminuyendo la extensión en un año cada vez, tantas veces como sea posible. Mediante este procedimiento y, según la extensión de las series, se obtendrá cierto número de modelos validados y no validados. Los modelos no validados deben ser descartados y los modelos validados se diferenciarán por su precisión (EMAP) y sesgo.

4. En la elección del mejor modelo de regresión lineal, con fines de pronóstico, se recomienda seleccionar a aquel modelo validado que presente un EMAP menor e igual a 5% (precisión Excelente), que su sesgo varíe entre -1 y +1, que exhiba un poder explicativo de al menos 70% y, además, posea el menor número de variables independientes (principio de parsimonia).
5. Dado que la demanda de productos hidrobiológicos congelados está atravesando un ciclo expansivo, es necesaria la realización de investigaciones que apunten a establecer si la oferta futura de recursos hidrobiológicos, la materia prima, será suficiente para sostener el crecimiento de la demanda de productos congelados. No hay que olvidar, que si bien los recursos hidrobiológicos son renovables, existe cierto nivel de extracción más allá del cual se pone en serio peligro la sostenibilidad de las especies hidrobiológicas y, por ende, el de la actividad económica.
6. Realizar estudios que establezcan el impacto del nivel de precisión de los pronósticos de demanda, tanto de modelos validados como de modelos no validados, en los indicadores de rentabilidad VAN y TIR y, por ende, en la decisión acerca de la materialización de un proyecto de inversión.

### VIII Referencias

- Adkins, L. (2018). *Using gretl for principles of econometrics* (5th ed.).  
[http://www.learneconometrics.com/gretl/poe5/using\\_gretl\\_for\\_POE5.pdf](http://www.learneconometrics.com/gretl/poe5/using_gretl_for_POE5.pdf)
- Agresti, A., Franklin, C. & Klingenberg, B. (2018). *Statistics. The art and science of learning from data* (4th ed.). Pearson.
- Albright, S. & Winston, W. (2017). *Business analytics. Data analysis and decision making* (6th ed.). Cengage.
- Aljandali, A. (2017). *Multivariate methods and forecasting with IBM® SPSS® statistics*. Springer.
- Amendola, A. & Francq, C. (2009). *Concepts of and tools for nonlinear time-series modeling*. En D. Belsley, & E. Kontoghiorghes (Eds.), *Handbook of computational econometrics* (pp. 377-427). Wiley.
- Anderson, D., Sweeney, D., Williams, T., Camm, J. & Cochran, J. (2017). *Statistics for business and economics* (13th ed.). USA: Cengage. Recuperado de gen.lib.rus.ec/
- Andren, T. (2007). *Econometrics*. Bookboon.
- Armstrong, J. (2002). *Extrapolation for time series and cross sectional data*. En J. Armstrong, (Ed.), *Principles of forecasting: A handbook of researchers and practitioners* (pp. 217-243). Kluwer Academic.
- Armstrong, G. & Kotler, P. (2017). *Marketing. An introduction* (13th ed.). Pearson.
- Asteriou, D. & Hall, S. (2011). *Applied econometrics* (2nd ed.). Macmillan.
- Baca, G. (2013). *Evaluación de proyectos* (7a ed.). McGraw-Hill.
- Ballou, R. (2004). *Logística. Administración de la cadena de suministro* (5a ed.). Pearson.
- Baltagi, B. (2011). *Econometrics* (5th ed.). Springer.

- Barreto, C. (2018). Modelo de regresión del consumo final privado y el producto bruto interno nacional del Perú durante el periodo 2000 – 2013. *In Crescendo*, 9(1), 13-32. <https://dialnet.unirioja.es/servlet/articulo?codigo=6508879>
- Barrow, M. (2017). *Statistics for economics, accounting and business studies* (7th ed.). Pearson.
- Behar, R. y Grima, P. (2018). *55 respuestas a dudas típicas de estadística* (2a ed.). Díaz de Santos.
- Bisgaard, S. & Kulahci, M. (2011). *Time series analysis and forecasting by example*. Wiley.
- Bowerman, B., O'Connell, R. & Murphree, E. (2015). *Regression Analysis: Unified concepts, practical applications and computer implementation*. Business Expert Press.
- Box, G., Jenkins, G., Reinsel, G. & Ljung, G. (2016). *Time series analysis. Forecasting and control* (5th ed.). Wiley.
- Bozarth, C. & Handfield, R. (2019). *Introduction to operations and supply chain management* (5th ed.). Pearson.
- Brooks, C. (2014). *Introductory econometrics for finance* (3rd ed.). Cambridge University Press.
- Brooks, C. & Tsolacos, S. (2010). *Real state modelling and forecasting*. Cambridge University Press.
- Caballero, A. (2014). *Metodología integral innovadora para planes y tesis. La metodología del cómo formularlos*. Cengage.
- Camm, J., Cochran, J., Fry, M., Ohlmann, J., Anderson, D., Sweeney, D. & Williams, T. (2019). *Business analytics* (3rd ed.). Cengage.
- Carbonel, J. (2016). *Formulación y evaluación de proyectos de inversión*. Macro.
- Caridad y Ocerin, J. (2012). *Econometría: Modelos econométricos y series temporales con los paquetes  $\mu$ TSP y TSP. Tomo 1: Modelos econométricos uniecuacionales*. Reverté.

- Carlberg, C. (2016). *Excel® sales forecasting for dummies* (2nd ed.). Wiley.
- Case, K., Fair, R. & Oster, S. (2017). *Principles of economics* (12th ed.). Pearson.
- Castillo, R. y Torres, H. (2016). *Estudio de prefactibilidad para la instalación de una planta productora de jabón líquido a base de jaboncillo de campo (Cucumis dipsaceus)*. [Tesis de pregrado]. Universidad de Lima
- Castle, J. & Hendry, D. (2019). *Modelling our changing world*. Macmillan.
- Chapman, S., Arnold, T., Gatewood, A. & Clive, L. (2017). *Introduction to materials management* (8th ed.). Pearson.
- Chase, C. (2013). *Demand-driven forecasting. A structured approach to forecasting* (2nd ed.). Wiley.
- Chase, C. (2016). *Next generation demand management. People, process, analytics and technology*. Wiley.
- Chatfield, C. (2015). *Confessions of a pragmatic forecaster*. En M. Gilliland, L. Tashman, & U. Sglavo (Eds.), *Business forecasting: Practical problems and solutions* (pp. 82-92). Wiley.
- Chatterjee, S. & Hadi, A. (2012). *Regression analysis by example* (5th ed.). Wiley.
- Chávez, R. y Juscamaita, M. (2013). *Estudio de pre-factibilidad para la exportación de cacao en grano tostado al mercado estadounidense*. [Tesis de pregrado, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP. <http://tesis.pucp.edu.pe/repositorio/handle/123456789/4563>
- Clements, M. (2005). *Evaluating econometric forecast of economic and financial variables*. Macmillan.
- Comisión de Promoción del Perú para la Exportación y el Turismo. (2019). *Desarrollo del comercio exterior pesquero y acuícola en el Perú. Informe anual 2018*. Comisión de Promoción del Perú para la Exportación y el Turismo (PROMPERU).

[https://www.siicex.gob.pe/siicex/resources/sectoresproductivos/DP2018%20VFinal .pdf](https://www.siicex.gob.pe/siicex/resources/sectoresproductivos/DP2018%20VFinal.pdf)

Cooper, D. & Schindler, P. (2014). *Business research methods* (12th ed.). McGraw-Hill.

Cottrell, A. & Lucchetti, R. (2022a). *Gretl command reference*.

<https://sourceforge.net/projects/gretl/files/manual/gretl-ref-a4.pdf/download>

Cottrell, A. & Lucchetti, R. (2022b). *Gretl User's guide*.

<https://sourceforge.net/projects/gretl/files/manual/gretl-guide-a4.pdf/download>

Court, E. y Rengifo, E. (2011). *Estadísticas y econometría financiera*. Cengage.

Curo, J. e Ybañez, S. (2017). *Parámetros óptimos para la obtención de un néctar de copoazú (*Theobroma grandiflorum*) y maracuyá (*Passiflora edulis*) y su estudio a nivel de pre-factibilidad*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos].

Repositorio Institucional UNMSM. <https://core.ac.uk/reader/323344780>

Dannecker, L. (2015). *Energy time series forecasting. Efficient and accurate forecasting of evolving time series from the energy domain*. Springer.

Darlington, R. & Hayes, A. (2017). *Regression analysis and linear models. Concepts, applications, and implementation*. Guilford.

Davidson, J. (2018). *An introduction to econometric theory*. Wiley.

De La Cruz, L. (2014). *Estudio de pre-factibilidad para la producción y comercialización de un complemento nutricional elaborado a base de productos naturales*. [Tesis de pregrado, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP.

<http://tesis.pucp.edu.pe/repositorio/handle/123456789/6020>

Del Castillo, J. (2016). *Análisis sectorial de la industria de pota y perico congelados en el Perú*. [Tesis de maestría, Universidad de Piura]. Repositorio Institucional UDEP.

[https://pirhua.udep.edu.pe/bitstream/handle/11042/2940/MDE\\_1615.pdf?sequence=1&isAllowed=y](https://pirhua.udep.edu.pe/bitstream/handle/11042/2940/MDE_1615.pdf?sequence=1&isAllowed=y)

- Devore, J. (2016). *Probability and statistics for engineering and the sciences* (9th ed.). Cengage.
- Dhrymes, P. (2017). *Introductory econometrics*. Springer.
- Díaz, R., Talley, W. & Tulpule, M. (2011). Forecasting empty container volumes. *The Asian Journal of Shipping and Logistics*, 27(2), 217-236.  
[https://www.researchgate.net/publication/257737372\\_Forecasting\\_Empty\\_Container\\_Volumes](https://www.researchgate.net/publication/257737372_Forecasting_Empty_Container_Volumes)
- Diebold, F. (2017). *Forecasting in economics, business, finance and beyond*. University of Pennsylvania.  
<https://www.sas.upenn.edu/~fdiebold/Teaching221/Forecasting.pdf>
- Doane, D. & Seward, L. (2016). *Applied statistics in business and economics* (5th ed.). McGraw-Hill.
- Domínguez, S., Sánchez, E. y Sánchez, G. (2009). *Guía para elaborar una tesis*. McGraw-Hill.
- Dougherty, C. (2011). *Introduction to econometrics* (4th ed.). Oxford University Press.
- Ducoing, C. (2015). Aplicación del paquete estadístico Gretl en la enseñanza de la historia económica. *Revista Uruguaya de Historia Económica*, 5(7), 81-100.  
<https://www.researchgate.net/publication/280085617>
- Ebert, R., Griffin, R., Dracopoulos, G. & Starke, F. (2020). *Business essentials* (9th ed.). Pearson.
- Elff, M. (2015). *Estimation techniques: Ordinary least squares and maximum likelihood*. En H. Best, & C. Wolf, (Eds.), *The SAGE handbook of regression analysis and causal inference* (pp. 7-30). SAGE.
- Elliott, G. & Timmermann, A. (2016). *Economic forecasting*. Princeton University.
- Enders, W. (2015). *Applied econometric time series* (4th ed.). Wiley.

- Evans, J. (2017). *Business analytics* (2nd ed.). Pearson.
- Fabozzi, F., Focardi, S., Rachev, S. & Arshanapalli, B. (2014). *The basics of financial econometrics. Tools, concepts, and asset management applications*. Wiley.
- Ferrell, O., Hirt, G. & Ferrell, L. (2018). *Business foundations. A changing world* (11th ed.). McGraw-Hill.
- Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). SAGE.
- Freund, R., Wilson, W. & Mohr, D. (2010). *Statistical methods* (3th ed.). Academic Press.
- Gee, S. (2015). *Fraud and fraud detection. A data analytics approach*. Wiley.
- Ghiani, G., Laporte, G. & Musmanno, R. (2004). *Introduction to logistics systems planning and control*. Wiley.
- Ghiani, G., Laporte, G. & Musmanno, R. (2013). *Introduction to logistics systems management* (2nd ed.). Wiley.
- Ghysels, E. & Marcellino, M. (2018). *Applied economic forecasting using time series methods* (2nd ed.). Oxford University Press.
- Gilliland, M. (2015). *Defining "Demand" for demand forecasting*. En M. Gilliland, L. Tashman, & U. Sglavo (Eds.), *Business forecasting: Practical problems and solutions* (pp. 60-66). Wiley.
- González-Rivera, G. (2016). *Forecasting for economics and business*. Routledge.
- Goodwin, P. (2018). *Profit from your forecasting software. A best practice guide for sales forecaster*. Wiley.
- Greene, W. (2018). *Econometric analysis* (8th ed.). Pearson.
- Groebner, D., Shannon, P. & Fry, P. (2018). *Business statistics: A decision-making approach* (10th ed.). Pearson.
- Guell, R. (2018). *Issues in economics today* (8th ed.). McGraw-Hill.
- Gujarati, D. (2015). *Econometrics by example* (2nd ed.). Macmillan.

- Gujarati, D. y Porter, D. (2010). *Econometría* (5a ed.). McGraw-Hill.
- Gutiérrez, A. (2012). *Probabilidad y estadística. Enfoque por competencias*. McGraw-Hill.
- Hahs-Vaughn, D. (2017). *Applied multivariate statistical concepts*. Routledge.
- Hanke, J. & Wichern, D. (2014). *Business forecasting* (9th ed.). Pearson.
- Harris, M. (2015). *Inside the Crystal Ball. How to make and use forecast*. Wiley.
- Harrison, A., Van Hoek, R. & Skipworth, H. (2014). *Logistics management and strategy. Competing through the supply chain* (5th ed.). Pearson.
- Heij, C., De Boer, P., Franses, P., Kloek, T. & Van Dijk, H. (2004). *Econometric methods with applications in business and economics*. Oxford University.
- Heizer, J., Render, B. & Munson, C. (2017). *Operations management. Sustainability and supply chain Management* (12th ed.). Pearson.
- Hernández Sampieri, R. y Mendoza, C. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Hill, C., Griffiths, W. & Lim, G. (2018). *Principles of econometrics* (5th ed.). Wiley.
- Hoekstra R., Kiers H. & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in psychology*, 3(137), 1-9.  
<https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00137/full>
- Hoffmann, J. (2010). *Linear regression analysis. Applications and assumptions* (2nd ed.). Brigham Young University.
- Hollensen, S. (2017). *Global marketing* (7th ed.). Pearson.
- Hoshmand, R. (2010). *Business forecasting: A practical approach* (2nd ed.). Routledge.
- Hyndman, R. & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
- Instituto Nacional de Estadística e Informática. (1996). *Perú: Compendio estadístico 1995-96*. Instituto Nacional de Estadística e Informática (INEI).

- Instituto Nacional de Estadística e Informática. (1998). *Perú: Compendio de estadísticas económicas y financieras 1997-1998*. Instituto Nacional de Estadística e Informática (INEI).
- Instituto Nacional de Estadística e Informática. (2002). *Perú: Compendio estadístico 2002*. Instituto Nacional de Estadística e Informática (INEI).
- Instituto Nacional de Estadística e Informática. (2009a). *Perú: Compendio estadístico 2009*. Instituto Nacional de Estadística e Informática (INEI).
- Instituto Nacional de Estadística e Informática. (2009b). *Perú: Estimaciones y proyecciones de población, 1950-2050*. Instituto Nacional de Estadística e Informática (INEI).
- Instituto Nacional de Estadística e Informática. (2017). *Compendio estadístico Perú 2017*. Instituto Nacional de Estadística e Informática (INEI).
- Ivanov, D., Tsipoulanidis, A. & Schönberger, J. (2017). *Global supply chain and operations management*. Springer.
- Jacobs, R. & Chase, R. (2018). *Operations and supply chain management* (15th ed.). McGraw-Hill.
- Jaggia, S. & Kelly, A. (2020). *Essentials of business statistics. Communicating with numbers* (2nd ed.). McGraw-Hill.
- Jobber, D. & Lancaster, G. (2015). *Selling and sales management* (10th ed.). Pearson.
- Karlsson, S. (2013). *Forecasting with bayesian vector autoregression*. En G. Elliott, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2B, pp. 791-897). Elsevier.
- Keith, T. (2019). *Multiple regression and beyond. An introduction to multiple regression and structural equation modeling* (3rd ed.). Routledge.
- Keller, G. (2018). *Statistics for management and economics* (11th ed.). Cengage.
- Kennedy, P. (2008). *A guide to econometrics* (6th ed.). Blackwell.

- Kolassa, S. & Siemsen, E. (2016). *Demand forecasting for managers*. Business Expert Press.
- Konasani, V. & Kadre, S. (2015). *Practical business analytics using SAS: A hands-on guide*. Apress.
- Krajewski, L., Malhotra, M. & Ritzman, L. (2016). *Operations Management. Processes and supply chains* (11th ed.). Pearson.
- Krajewski, L., Ritzman, L. y Malhotra, M. (2008). *Administración de operaciones. Procesos y cadenas de valor* (8a ed.). Pearson.
- Kumar, A. & Suresh, N. (2009). *Operations management*. New Age International.
- Kutner, M., Nachtsheim, C., Neter, J. & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill.
- Lawrence, K., Klimberg, R. & Lawrence, S. (2009). *Fundamentals of forecasting using Excel*. Industrial Press.
- Lee, C. & Lee, M. (2017). Demand forecasting in the early stage of the technology's life cycle using a bayesian update. *Sustainability*, 9(1378), 1-15.  
<https://www.mdpi.com/2071-1050/9/8/1378>
- Lee, C., Lee, J. & Lee, A. (2013). *Statistics for Business and financial economics* (3rd ed.). Springer.
- Levine, D., Stephan, D. & Szabat, K. (2017). *Statistics for managers using Microsoft® Excel* (8th ed.). Pearson.
- Lilja, D. (2017). *Linear regression using R: An introduction to data modeling*. University of Minnesota.
- Lind, D., Marchal, W. & Wathen, S. (2018). *Statistical techniques in business & economics* (17th ed.). McGraw-Hill.
- Lysons, K. & Farrington, B. (2016). *Procurement and supply chain management* (9th ed.). Pearson.

- Maldonado, H. y Puertas, M. (2011). La pesca industrial peruana antes de la anchoveta (1923-1955). *Investigaciones sociales*, 15(27), pp. 559-573.  
<https://revistasinvestigacion.unmsm.edu.pe/index.php/sociales/article/view/7703/6708>
- Marcellino, M. (2016). *Applied econometrics. An introduction*. Bocconi University.
- Martínez, E. (2005). Errores frecuentes en la interpretación del coeficiente de determinación lineal. *Anuario jurídico y económico escurialense*, 38(2005), 315-332.  
<https://dialnet.unirioja.es/descarga/articulo/1143023.pdf>
- Matloff, N. (2017). *Statistical regression and classification*. CRC Press.
- Mendenhall, W. & Sincich, T. (2012). *A second course in statistics. Regression analysis* (7th ed.). Pearson.
- McClave, J. & Sincich, T. (2018). *Statistics* (13th ed.). Pearson.
- McCormick, K. & Salcedo, J. (2015). *SPSS for dummies* (3rd ed.). Wiley.
- McEvoy, D. (2018). *A guide to business statistics*. Wiley.
- Mendez, J. (2018). *Modelo matemático predictivo del comportamiento de la exportación de fresa (Fragaria vesca L.) congelada del Perú al mercado asiático frente a factores externos e internos*. [Tesis de pregrado]. Universidad Privada Antenor Orrego
- Meuleman, B., Loosvedlt, G. & Emonds, V. (2015). *Regression analysis: Assumptions and diagnostics*. En H. Best, & C. Wolf, (Eds.), *The SAGE handbook of regression analysis and causal inference* (pp. 83-110). SAGE.
- Min, H. (2015). *The essentials of supply chain management. New business, concepts and applications*. Pearson.
- Ministerio de la Producción. (2005). *Anuario estadístico del sector producción 2004*. Ministerio de la Producción (PRODUCE).
- Ministerio de la Producción. (2007). *Anuario estadístico del sector producción 2006*. Ministerio de la Producción (PRODUCE).

- Ministerio de la Producción. (2010). *Anuario estadístico del sector producción 2010*.  
Ministerio de la Producción (PRODUCE).
- Ministerio de la Producción. (2018a). *Boletín del sector pesquero- Noviembre 2018*.  
Ministerio de la Producción (PRODUCE).
- Ministerio de la Producción. (2018b). *Boletín del sector pesquero- Diciembre 2018*.  
Ministerio de la Producción (PRODUCE).
- Ministerio de la Producción. (2019). *Anuario estadístico pesquero y acuícola 2017*.  
Ministerio de la Producción (PRODUCE).
- Montgomery, D., Jennings, C. & Kulahci, M. (2015). *Introduction to time series analysis and forecasting* (2nd ed.). Wiley.
- Montgomery, D., Peck, E. & Vining, G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
- Muñoz, C. (2011). *Cómo elaborar y asesorar una investigación de tesis* (2a ed.). Pearson.
- Nahmias, S. & Olsen, T. (2015). *Production and operations analysis* (7th ed.). Waveland Press.
- Nash, J. & Nash, M. (2001). *Practical forecasting for managers*. Arnold.
- Nayal, L. (2015). *Regression analysis to forecast the demand of new single family houses in USA* (Starred paper). [Tesis de maestría, Saint Cloud State University]. Repositorio Institucional Saint Cloud State University.  
[https://repository.stcloudstate.edu/cgi/viewcontent.cgi?referer=https://www.google.com.pe/&httpsredir=1&article=1018&context=mme\\_etds](https://repository.stcloudstate.edu/cgi/viewcontent.cgi?referer=https://www.google.com.pe/&httpsredir=1&article=1018&context=mme_etds)
- Newbold, P., Carlson, W. & Thorne, B. (2013). *Statistics for Business and Economics* (8th ed.). Pearson.

- Palacios, C., Salcedo, O. y Coronel, F. (2021). *Determinantes del incremento del número de clientes residenciales de gas natural en la concesión de Lima y Callao: 2007-2019*. [Tesis de maestría]. Universidad del Pacífico
- Pasupathy, K. (2010). *Forecasting model for strategic and operations planning of a nonprofit health care organization*. En K. Lawrence, & R. Klimberg (Eds.), *Advances in business and management forecasting* (Vol. 7, pp. 59-69). Emerald.
- Pedace, R. (2013). *Econometrics for dummies*. Wiley.
- Pesaran, M. (2015). *Time series and panel data econometrics*. Oxford University Press.
- Pino, C. (2017). *Estudio de prefactibilidad para el procesamiento y exportación de aceitunas rellenas de pimienta a Estados Unidos*. [Tesis de pregrado, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP. [https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/8873/PINO\\_GON\\_ZALEZ\\_CLAUDIA\\_ESTUDIO\\_TESIS.pdf?sequence=8&isAllowed=y](https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/8873/PINO_GON_ZALEZ_CLAUDIA_ESTUDIO_TESIS.pdf?sequence=8&isAllowed=y)
- Pirc, J., Desanto, D., Davison, I. & Gragido, W. (2016). *Threat forecasting. Leveraging big data for predictive analysis*. Elsevier.
- Quintana, L. y Mendoza, M. (2016). *Econometría aplicada utilizando R*. Universidad Autónoma de México.
- Reid, D. & Sanders, N. (2013). *Operations management* (5th ed.). Wiley.
- Render, B., Stair, R., Hanna, M. & Hale, T. (2018). *Quantitative analysis for management* (13th ed.). Pearson.
- Richardson, R. (2015). *Business applications of multiple regression* (2nd ed.). Business Expert Press.
- Rodríguez, D. y González, G. (2017). *Principios de econometría*. Instituto Tecnológico Metropolitano.

- Rokunuzzaman, M. (2013). *Multivariate analysis of EEG data: Some aspects of diagnostic of MANOVA model*. [Tesis de maestría, Universidad de Estocolmo].  
[http://www.statistics.su.se/polopoly\\_fs/1.136892.1369899902!/menu/standard/file/VT13%20Masteruppsats%20Md%20Rokonuzzaman.pdf](http://www.statistics.su.se/polopoly_fs/1.136892.1369899902!/menu/standard/file/VT13%20Masteruppsats%20Md%20Rokonuzzaman.pdf)
- Rosales, R., Perdomo, J., Morales, C. y Urrego, J. (2013). *Fundamentos de econometría intermedia: teoría y aplicaciones*. Uniandes.
- Ryan, T. (2013). *Sample size determination and power*. Wiley.
- Sánchez, Y., Matus, J., García, J., Martínez, M. y Gómez, M. (2011). Estimación de la demanda de importaciones de limón persa (*Citrus latifolia Tanaka*) en Estados Unidos procedentes de México (1994-2008). *Tropical and Subtropical Agroecosystems*, 14 (2011), 819-827. <https://www.scielo.org.mx/pdf/tsa/v14n3/v14n3a12.pdf>
- Sankaran, G., Sasso, F., Kepcczynski, R. & Chiaraviglio, A. (2019). *Improving forecasts with integrated business planning*. Springer.
- Sapag, N., Sapag, R. y Sapag, J. (2014). *Preparación y evaluación de proyectos* (6a ed.). McGraw-Hill.
- Schönsleben, P. (2016). *Integral logistics management. Operations and supply chain management within and across companies* (5th ed.). CRC Press.
- Schroeder, R. & Goldstein, S. (2018). *Operations management in the supply chain. Decisions and cases* (7th ed.). McGraw-Hill.
- Shahabuddin, S. (2009). Forecasting automobile sales. *Management Research News*, 32(7), 670-682.  
<https://www.emerald.com/insight/content/doi/10.1108/01409170910965260/full/html?skipTracking=true>
- Shim, J. (2009). *Strategic business forecasting*. Global Professional Publishing.
- Siegel, A. (2016). *Practical business statistics* (7th ed.). Academic Press.

- Silver, E., Pyke, D. & Thomas, D. (2017). *Inventory and production management in supply chains* (4th ed.). CRC Press.
- Soto, R. (2015). Modelo econométrico para estimar la demanda de sandías en USA, 2000 – 2010. *El buzón de Pacioli*, 15(90), 50-62.  
<https://www.itson.mx/publicaciones/pacioli/Documents/no90/Pacioli-90-eBook.pdf>
- Steele, B., Chandler, J. & Reddy, S. (2016). *Algorithms for data science*. Springer.
- Stevenson, W. (2018). *Operations management* (13th ed.). McGraw-Hill.
- Stock, J. & Watson, M. (2020). *Introduction to econometrics* (4th ed.). Pearson.
- Studenmund, A. (2017). *Using econometrics: A practical guide* (7th ed.). Pearson.
- Suárez, E., Pérez, C., Rivera, R. & Martínez, M. (2017). *Applications of regression models in epidemiology*. Wiley.
- Tattar, P., Ramaiah, S. & Manjunath, B. (2016). *A course in statistics with R*. Wiley.
- Taylor, B. (2016). *Introduction to management science* (12th ed.). Pearson.
- Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8), 1352-1362.  
[www.researchgate.net/publication/280222125\\_A\\_better\\_measure\\_of\\_relative\\_prediction\\_accuracy\\_for\\_model\\_selection\\_and\\_model\\_estimation](http://www.researchgate.net/publication/280222125_A_better_measure_of_relative_prediction_accuracy_for_model_selection_and_model_estimation)
- Triola, M. (2018). *Estadística* (12a ed.). Pearson.
- Verbeek, M. (2017). *A guide to modern econometrics* (5th ed.). Wiley.
- Vu, T. (2015). *Seeing the future. How to build basic forecasting models*. Business Expert Press.
- Walpole, R., Myers, R., Myers, S. y Ye, K. (2016). *Probability & Statistics for engineers & scientists* (9th ed.). Pearson.
- Waller, D. (2008). *Statistics for Business*. Elsevier.

- Warner, R. (2013). *Applied statistics. From bivariate through multivariate techniques* (2nd ed.). SAGE.
- Waters, D. (2011). *Quantitative methods for business* (5th ed.). Pearson.
- Webb, R. y Fernández, G. (1992). *Perú en números 1992*. Cuánto.
- Weisberg, S. (2014). *Applied linear regression* (4th ed.). Wiley.
- Welc, J. & Rodriguez, P. (2018). *Applied regression analysis for business*. Springer.
- Wheelen, T., Hunger, D., Hoffman, A. & Bamford, C. (2018). *Concepts in strategic management and business policy. Globalization, innovation and sustainability* (15th ed.). Pearson.
- Wilson, H. & Keating, B. (2009). *Business forecasting with ForecastX* (6th ed.). McGraw-Hill.
- Wisner, J., Tan, K. & Leong, G. (2019). *Principles of supply chain management. A balanced approach* (5th ed.). Cengage.
- Wooldridge, J. (2020). *Introductory econometrics. A modern approach* (7th ed.). Cengage.
- Yan, X; & Su, X. (2009). *Linear regression analysis. Theory and computing*. World Scientific.
- Young, D. (2017). *Handbook of regression methods*. CRC Press.
- Zax, J. (2011). *Introductory econometrics. Intuition, proof, and practice*. Stanford University Press.

## IX Anexos

## Anexo A: Matriz de Consistencia

EL ANÁLISIS DE REGRESIÓN LINEAL Y LA PRECISIÓN DEL PRONÓSTICO DE DEMANDA DE PRODUCTOS HIDROBIOLÓGICOS CONGELADOS EN EL PERÚ			
PROBLEMAS	OBJETIVOS	HIPÓTESIS	MÉTODO
<b>GENERAL</b>	<b>GENERAL</b>	<b>GENERAL</b>	<b>TIPO DE INVESTIGACIÓN</b>
¿Qué relación existe entre el análisis de regresión lineal y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú?	Evaluar la relación existente entre el análisis de regresión lineal y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.	El análisis de regresión lineal se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.	La investigación es de tipo cuantitativa y aplicada.
			<b>NIVEL DE LA INVESTIGACIÓN</b>
			El nivel de la investigación es el correlacional.
<b>ESPECÍFICO 1</b>	<b>ESPECÍFICO 1</b>	<b>ESPECÍFICA 1</b>	<b>DISEÑO DE LA INVESTIGACIÓN</b>
¿Cómo se relaciona la extensión de las series temporales con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú?	Precisar la relación que existe entre la extensión de las series temporales y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.	Existe una relación positiva entre la extensión de las series temporales y la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.	El diseño de la investigación es no experimental, correlacional-longitudinal.
<b>ESPECÍFICO 2</b>	<b>ESPECÍFICO 2</b>	<b>ESPECÍFICA 2</b>	
¿En qué forma se relaciona el número de variables independientes con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú?	Determinar la relación del número de variables independientes con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.	El número de variables independientes se relaciona positivamente con la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú.	
<b>ESPECÍFICO 3</b>	<b>ESPECÍFICO 3</b>	<b>ESPECÍFICA 3</b>	
¿Es mayor la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú, de un modelo de regresión lineal validado frente a un modelo de regresión lineal no validado?	Comparar la precisión del pronóstico de demanda de productos hidrobiológicos congelados en el Perú, entre un modelo de regresión lineal validado y un modelo de regresión lineal no validado.	Un modelo de regresión lineal validado presenta mayor precisión en el pronóstico de demanda de productos hidrobiológicos congelados en el Perú, frente a un modelo de regresión lineal no validado.	<b>POBLACIÓN</b>
			La población está compuesta por las series temporales CONG (1980-2018), POB, PBI, PBIpes, PBIman, DI, IP e INF (1950-2018).
			<b>MUESTRA</b>
			La muestra la componen las series temporales listadas en el ítem población, pero para el periodo 1980-2018.

**Anexo B: Series Temporales Empleadas (1980-2018)**

Año	CONG (Miles de t)	POB (Habitantes)	PBI (Millones de soles a precios constantes de 2007)	PBIpes	PBIman	DI	IP	INF (%)	INF1 (%)
1980	14,3	17 324 179	167 596	412	29 628	163 099	21 770	60,82	61,82
1981	10,8	17 760 219	176 901	450	29 827	177 828	25 700	72,67	73,67
1982	7,1	18 197 198	176 507	528	29 477	174 624	23 835	72,93	73,93
1983	7,0	18 635 588	158 136	371	24 128	152 822	14 651	125,07	126,07
1984	17,0	19 075 874	163 842	543	25 510	152 816	13 639	111,46	112,46
1985	13,7	19 518 555	167 219	636	26 659	152 380	12 079	158,26	159,26
1986	27,8	19 965 797	182 981	840	30 817	176 112	16 802	62,90	63,90
1987	20,7	20 417 262	200 778	740	34 765	197 356	21 260	114,51	115,51
1988	26,6	20 869 717	181 822	876	30 869	177 840	19 321	1 722,32	1 723,32
1989	31,0	21 319 883	159 436	922	26 034	147 158	15 242	2 774,98	2 775,98
1990	39,4	21 764 515	151 492	916	24 524	144 696	17 159	7 649,65	7 650,65
1991	19,6	22 203 931	154 854	818	26 038	149 775	17 154	139,23	140,23
1992	11,3	22 640 305	154 017	923	25 421	149 741	16 450	56,73	57,73
1993	12,7	23 073 150	162 093	1 117	26 643	157 824	18 028	39,48	40,48
1994	13,6	23 501 974	182 044	1 440	30 583	178 776	25 088	15,38	16,38
1995	25,3	23 926 300	195 536	1 223	32 312	198 931	31 946	10,23	11,23
1996	24,2	24 348 132	201 009	1 195	32 879	201 177	31 247	11,84	12,84
1997	15,1	24 767 794	214 028	1 191	34 365	213 839	36 241	6,46	7,46
1998	9,4	25 182 269	213 190	1 076	33 425	211 514	35 373	6,01	7,01
1999	10,6	25 588 546	216 377	1 428	32 962	203 985	29 981	3,73	4,73
2000	13,7	25 983 588	222 207	1 710	34 792	207 189	29 461	3,73	4,73
2001	23,5	26 366 533	223 580	1 488	35 094	205 923	28 089	-0,13	0,87
2002	33,1	26 739 379	235 773	1 529	37 424	214 586	28 145	1,52	2,52
2003	27,3	27 103 457	245 593	1 417	38 883	222 243	29 915	2,48	3,48
2004	33,4	27 460 073	257 770	1 988	41 778	228 321	32 335	3,48	4,48
2005	54,7	27 810 540	273 971	2 086	44 529	237 916	36 217	1,49	2,49
2006	29,9	28 151 443	294 598	2 163	47 766	264 934	43 482	1,14	2,14
2007	43,6	28 481 901	319 693	2 364	52 807	296 928	53 626	3,93	4,93
2008	41,9	28 807 034	348 923	2 436	57 354	336 314	66 440	6,65	7,65
2009	61,9	29 132 013	352 584	2 352	53 502	324 587	60 566	0,25	1,25
2010	98,6	29 461 933	382 380	1 891	59 255	373 071	76 167	2,08	3,08
2011	68,3	29 797 694	407 052	2 892	64 330	402 317	84 518	4,74	5,74
2012	73,5	30 135 875	431 273	1 960	65 265	431 810	97 722	2,65	3,65
2013	80,0	30 475 144	456 449	2 445	68 508	462 747	104 660	2,86	3,86
2014	88,6	30 814 175	467 376	1 762	66 047	472 691	102 337	3,22	4,22
2015	106,9	31 151 643	482 676	2 042	65 079	486 269	98 062	4,40	5,40
2016	119,3	31 488 625	502 191	1 836	64 187	491 417	92 762	3,23	4,23
2017	122,2	31 826 018	514 618	1 921	64 033	498 486	92 960	1,36	2,36
2018	140,2	32 162 184	535 171	2 685	67 976	519 730	97 059	2,19	3,19

Fuente: PRODUCE, INEI, BCRP

**Anexo C: Datos Para Probar la Hipótesis Específica 1**

N°	Método	Variable		Extensión (Años)	EMAP
		Dependiente	Independientes		
1	PW	L-CONG	DI	8	9,91%
2	PW	L-CONG	DI	10	14,59%
3	PW	L-CONG	DI	11	12,38%
4	PW	L-CONG	DI	12	12,28%
5	PW	L-CONG	DI	13	10,44%
6	PW	L-CONG	DI	14	6,72%
7	PW	L-CONG	DI	15	3,84%
8	PW	L-CONG	DI	18	13,82%
9	PW	L-CONG	DI	19	14,85%
10	PW	L-CONG	DI	20	5,46%
11	PW	L-CONG	DI	21	5,15%
12	PW	L-CONG	DI	22	3,69%
13	PW	L-CONG	DI	23	5,93%
14	PW	L-CONG	DI	24	12,35%
15	PW	L-CONG	DI	25	17,02%
16	PW	L-CONG	DI	26	18,07%
17	PW	L-CONG	DI	27	18,13%
18	PW	L-CONG	DI	28	20,35%
19	PW	L-CONG	DI	29	13,44%
20	PW	L-CONG	DI	30	14,15%
21	PW	L-CONG	DI	31	6,78%
22	PW	L-CONG	DI	34	13,44%
23	PW	L-CONG	DI - INF1	24	6,12%
24	PW	L-CONG	DI - INF1	25	9,57%
25	PW	L-CONG	DI - INF1	26	10,34%
26	PW	L-CONG	DI - INF1	27	9,99%
27	PW	L-CONG	DI - INF1	28	12,58%
28	PW	L-CONG	DI - INF1	29	6,37%
29	PW	L-CONG	DI - INF1	30	7,04%
30	PW	L-CONG	DI - INF1	31	5,58%
31	PW	L-CONG	DI - INF1	32	2,63%
32	PW	L-CONG	DI - INF1	33	4,37%
33	PW	L-CONG	DI - INF1	34	5,78%
34	PW	L-CONG	PBIman - INF1	15	23,98%
35	PW	L-CONG	PBIman - INF1	16	23,78%
36	PW	L-CONG	PBIman - INF1	24	30,87%
37	PW	L-CONG	PBIman - INF1	25	32,43%
38	PW	L-CONG	PBIman - INF1	26	32,94%

**Anexo C: Continuación**

N°	Método	Variable		Extensión (Años)	EMAP
		Dependiente	Independientes		
39	PW	L-CONG	PBIman - INF1	27	32,93%
40	PW	L-CONG	PBIman - INF1	29	31,84%
41	PW	L-CONG	PBIman - INF1	30	32,42%
42	PW	L-CONG	PBIman - INF1	31	28,30%
43	PW	L-CONG	PBIman - INF1	32	29,63%
44	PW	L-CONG	PBIman - INF1	33	31,05%
45	PW	L-CONG	PBIman - INF1	34	31,34%
46	PW	L-CONG	PBI	8	4,47%
47	PW	L-CONG	PBI	10	7,70%
48	PW	L-CONG	PBI	11	4,91%
49	PW	L-CONG	PBI	12	5,59%
50	PW	L-CONG	PBI	13	4,06%
51	PW	L-CONG	PBI	18	8,65%
52	PW	L-CONG	PBI	19	4,20%
53	PW	L-CONG	PBI	20	12,54%
54	PW	L-CONG	PBI	21	12,58%
55	PW	L-CONG	PBI	22	14,36%
56	PW	L-CONG	PBI	23	9,40%
57	PW	L-CONG	PBI	24	3,12%
58	PW	L-CONG	PBI	25	7,73%
59	PW	L-CONG	PBI	26	8,62%
60	PW	L-CONG	PBI	27	8,59%
61	PW	L-CONG	PBI	28	11,27%
62	PW	L-CONG	PBI	29	5,51%
63	PW	L-CONG	PBI	30	6,41%
64	PW	L-CONG	PBI	31	3,39%
65	PW	L-CONG	PBI	34	2,66%
66	PW	L-CONG	PBI - INF1	24	9,03%
67	PW	L-CONG	PBI - INF1	25	4,34%
68	PW	L-CONG	PBI - INF1	26	3,47%
69	PW	L-CONG	PBI - INF1	27	3,58%
70	PW	L-CONG	PBI - INF1	28	2,59%
71	PW	L-CONG	PBI - INF1	29	5,24%
72	PW	L-CONG	PBI - INF1	30	4,40%
73	PW	L-CONG	PBI - INF1	31	14,80%
74	PW	L-CONG	PBI - INF1	32	13,61%
75	PW	L-CONG	PBI - INF1	33	11,12%
76	PW	L-CONG	PBI - INF1	34	9,70%

Elaboración propia

**Anexo D: Datos Para Probar la Hipótesis Específica 2**

Nº	Método	Extensión (Años)	Variable Dependiente	Variable Independiente	EMAP	Variables Independientes	EMAP
1 - 2	PW	24	L-CONG	DI	12,35%	DI - INF1	6,12%
3 - 4	PW	25	L-CONG	DI	17,02%	DI - INF1	9,57%
5 - 6	PW	26	L-CONG	DI	18,07%	DI - INF1	10,34%
7 - 8	PW	27	L-CONG	DI	18,13%	DI - INF1	9,99%
9 - 10	PW	28	L-CONG	DI	20,35%	DI - INF1	12,58%
11 - 12	PW	29	L-CONG	DI	13,44%	DI - INF1	6,37%
13 - 14	PW	30	L-CONG	DI	14,15%	DI - INF1	7,04%
15 - 16	PW	31	L-CONG	DI	6,78%	DI - INF1	5,58%
17 - 18	PW	34	L-CONG	DI	13,44%	DI - INF1	5,78%
19 - 20	PW	24	L-CONG	PBI	3,12%	PBI - INF1	9,03%
21 - 22	PW	25	L-CONG	PBI	7,73%	PBI - INF1	4,34%
23 - 24	PW	26	L-CONG	PBI	8,62%	PBI - INF1	3,47%
25 - 26	PW	27	L-CONG	PBI	8,59%	PBI - INF1	3,58%
27 - 28	PW	28	L-CONG	PBI	11,27%	PBI - INF1	2,59%
29 - 30	PW	29	L-CONG	PBI	5,51%	PBI - INF1	5,24%
31 - 32	PW	30	L-CONG	PBI	6,41%	PBI - INF1	4,40%
33 - 34	PW	31	L-CONG	PBI	3,39%	PBI - INF1	14,80%
35 - 36	PW	34	L-CONG	PBI	2,66%	PBI - INF1	9,70%