



# Universidad Autónoma de Querétaro

## Facultad de Ingeniería

PREDICCIÓN DE PRECIOS: DOS MÉTODOS CONFRONTADOS

### **TESIS**

Que como parte de los requisitos para obtener el grado de  
Licenciado en Matemáticas Aplicadas

**Presentan:**

Daniel Gómez Torres

**Dirigido por:**

M. en C. Wilfrido Jacobo Paredes García

QUERÉTARO, QRO, 2021.



Universidad Autónoma de Querétaro  
Facultad de Ingeniería  
Licenciatura en Matemáticas Aplicadas

PREDICCIÓN DE PRECIOS: DOS MÉTODOS CONFRONTADOS  
**TESIS**

Que como parte de los requisitos para obtener el grado de  
Licenciado en Matemáticas Aplicadas

**Presentan:**

Daniel Gómez Torres

**Dirigido por:**

M. en C. Wilfrido Jacobo Paredes García

M. en C. Wilfrido Jacobo Paredes García  
Presidente

M. en C. Christopher Alexis Cedillo Jiménez  
Secretario

Dr. Eduardo Castaño Tostado  
Vocal

M. en C. Juan Antonio Villeda Resendiz  
Suplente

Centro Universitario, Querétaro, Qro.  
Junio de 2021  
México

## Resumen

En la actualidad, la volatilidad de los precios de las frutas y verduras es un tema apremiante para la sociedad. De ahí que el objetivo de esta tesis es determinar la metodología que tiene una mejor capacidad de predicción para los precios de los productos a través del error cuadrado medio.

El conocimiento del futuro es de gran importancia en cuestiones de economía debido a que permite realizar una planeación adecuada y evitar pérdidas monetarias. En este trabajo se emplearon los modelos autoregresivos integrados de media móvil y las redes neuronales recurrentes para realizar predicciones de 18 series de tiempo de precios semanales de frutas y hortalizas comercializadas en el mercado de abastos del estado de Querétaro. Por último, se analizó el desempeño de la estimación de los precios usando criterios de evaluación de errores.

**Palabras clave: Volatilidad, Predicción, Series de tiempo, Modelos autoregresivos integrados de media móvil, Redes neuronales recurrentes**

## Abstract

Currently, the volatility of fruit and vegetables prices represent a pressing issue for the society. Therefore, the aim of this thesis is to determinate the methodology which has a better price forecasting capability of the products through the mean square error.

A future knowing is greatly important in economic matters because it allows to perform an adequate planning and avoid monetary losses. In this work, the auto-regressive integrated moving average models and recurrent neuronal network were applied to made forecasting of 18 weekly time series of fruit and vegetables prices sold in the Querétaro state food market. To conclude, future price estimation performance was analyzed through error evaluation criteria.

**Key words: Volatility, Forecasting, Time series, Auto-regressive integrated moving average models, Recurrent neuronal networks**

Esta tesis está dedicada a mis padres que me han cuidado, apoyado, educado y guiado cada día de mi vida porque gracias a ellos he podido llegar hasta esta etapa de mi vida. A las personas que contribuyeron a mi formación académica y a mi crecimiento como persona.

Dirección General de Bibliotecas UAQ

## Agradecimientos

En primer lugar deseo expresar mi agradecimiento a mi director de tesis, al M en C. Wilfrido J. Paredes García, por su guía, su tiempo, su compromiso a lo largo de estos meses y por todo el apoyo prestado en la planeación y la escritura de este trabajo.

Asimismo, agradezco a la Universidad Autónoma de Querétaro y a la Facultad de Ingeniería por las herramientas brindadas a lo largo de estos años formándome como profesional y como ser humano. En particular, al fondo FOPER-2020-FIN01831 que permitió la obtención y de datos que permitieron la realización de este trabajo de tesis.

De igual manera quiero agradecer a mis compañeros del grupo 81 de la carrera de matemáticas aplicadas su ayuda dentro del aula de clase, su apoyo personal y por todas experiencias compartidas en estos 5 años.

A mis sinodales les agradezco por tomarse el tiempo para revisar la tesis y comentar sus observaciones para mejorar el trabajo.

Por último, le doy gracias a mis padres por el apoyo brindado, por permitirme tener una excelente educación a lo largo de mi vida y por los valores inculcados.

## Índice general

<b>Resumen</b>	<b>I</b>
<b>Abstract</b>	<b>II</b>
<b>Dedicatoria</b>	<b>III</b>
<b>Agradecimientos</b>	<b>IV</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Hipótesis y objetivos . . . . .	3
1.1.1. Hipótesis . . . . .	3
1.1.2. Objetivos . . . . .	3
<b>2. Marco teórico</b>	<b>4</b>
2.1. Web scraping . . . . .	4
2.2. Métodos de predicción . . . . .	5
2.2.1. Regresión lineal . . . . .	5
2.2.2. Series de tiempo . . . . .	5
2.2.3. Redes neuronales . . . . .	8
2.3. Modelos . . . . .	15
2.3.1. Modelo autorregresivo integrado de media móvil . . . . .	15
2.3.2. Red neuronal recurrente ELMAN . . . . .	17
2.4. Simulación numérica . . . . .	22
2.5. Método de remuestreo . . . . .	24
2.6. Prueba F de Fisher . . . . .	25
<b>3. Metodología</b>	<b>27</b>
3.1. Extracción de datos . . . . .	27
3.2. Elección de productos . . . . .	29
3.3. Análisis exploratorio de los datos . . . . .	30
3.4. Estimación de modelos . . . . .	32
3.4.1. Modelos ARIMA . . . . .	32
3.4.2. Red neuronal recurrente ELMAN . . . . .	34
3.5. Comparación de modelos . . . . .	38
<b>4. Resultados</b>	<b>39</b>
4.1. Extracción de la base de datos . . . . .	39
4.2. Análisis de datos . . . . .	40
4.3. Gráficas de las series . . . . .	40
4.4. Selección de frutas y hortalizas . . . . .	43
4.5. Transformación de los datos . . . . .	44
4.6. ARIMA . . . . .	46
4.7. ERNN . . . . .	49
4.8. Errores cuadrados medios . . . . .	51

4.9. Red neuronal recurrente Elman . . . . .	56
<b>5. Discusión y conclusiones</b>	<b>61</b>
5.1. Discusión . . . . .	61
5.2. Conclusiones . . . . .	62
<b>A. Productos de la base de datos</b>	<b>63</b>
<b>B. Análisis descriptivo</b>	<b>64</b>
<b>C. Gráficas: Series de tiempo</b>	<b>68</b>
<b>D. Tablas: Series de tiempo semanales</b>	<b>89</b>
<b>E. Tendencia lineal de las series de tiempo</b>	<b>96</b>
<b>F. Series de tiempo estacionarias</b>	<b>103</b>
<b>G. Funciones de autocorrelación muestral</b>	<b>110</b>
<b>H. Análisis de los residuales</b>	<b>119</b>
<b>I. Predicciones de las series de tiempo</b>	<b>128</b>
<b>J. Aprendizaje de la red neuronal</b>	<b>135</b>
<b>K. Algoritmo: Construcción de base de datos</b>	<b>142</b>
<b>L. Algoritmo: Red neuronal recurrente ELMAN</b>	<b>148</b>

Dirección General de Bibliotecas UAQ

## Índice de tablas

3.1. Problemas de los datos . . . . .	29
4.1. Descripción de datos . . . . .	40
4.2. Productos seleccionados viables . . . . .	43
4.3. Precios semanales promedio de aguacate Hass . . . . .	44
4.4. Resultados de la regresión lineal . . . . .	45
4.5. Modelos ajustados a los datos . . . . .	47
4.6. Predicción de los modelos ARIMA . . . . .	48
4.7. Parámetros de ajuste de la red neuronal . . . . .	50
4.8. Predicción de la red neuronal . . . . .	51
4.9. Precisión de los modelos . . . . .	52
4.10. Prueba de normalidad . . . . .	53
4.11. Intervalos de confianza . . . . .	54
4.12. Estadístico F . . . . .	54
4.13. Parámetros de ajuste de la red neuronal . . . . .	57
4.14. Predicción de la red neuronal . . . . .	58
4.15. Errores de la red neuronal . . . . .	59
4.16. Precisión de los modelos . . . . .	60
A.1. Productos comercializados en el mercado de Querétaro . . . . .	63
B.1. Análisis descriptivo de los datos . . . . .	67
D.1. Precios semanales del aguacate Hass . . . . .	89
D.2. Precios semanales del cacahuete . . . . .	89
D.3. Precios semanales de la calabacita italiana . . . . .	90
D.4. Precios semanales del chayote sin espinas . . . . .	90
D.5. Precios semanales del chile serrano . . . . .	90
D.6. Precios semanales de la coliflor grande . . . . .	91
D.7. Precios semanales del epazote . . . . .	91
D.8. Precios semanales de la jícama . . . . .	91
D.9. Precios semanales del kiwi . . . . .	92
D.10. Precios semanales de la lima . . . . .	92
D.11. Precios semanales del limon #5 . . . . .	92
D.12. Precios semanales de la naranja Valencia grande . . . . .	93
D.13. Precios semanales del nopal grande . . . . .	93
D.14. Precios semanales de la piña mediana . . . . .	93
D.15. Precios semanales del plátano macho . . . . .	94
D.16. Precios semanales del plátano Tabasco . . . . .	94
D.17. Precios semanales del tomate saladette . . . . .	94
D.18. Precios semanales de la toronja roja . . . . .	95

## Índice de figuras

2.1. Ruido blanco Gaussiano . . . . .	6
2.2. Red neuronal artificial . . . . .	10
2.3. Neurona artificial . . . . .	11
2.4. Red neuronal recurrente ELMAN . . . . .	19
3.1. Series de precios de ajo morado . . . . .	29
3.2. Tendencia lineal de la serie de aguacate Hass . . . . .	30
3.3. Primera diferencia de la serie de aguacate Hass . . . . .	32
4.1. Series con pocos precios semanales . . . . .	41
4.2. Series con periodos sin información . . . . .	42
4.3. Series con determinado comportamiento . . . . .	43
4.4. Intervalos de predicción para la serie de aguacate Hass . . . . .	49
4.5. Valor crítico de $F_{7,7}$ . . . . .	55
C.1. Precios del aguacate Hass . . . . .	68
C.2. Precios del ajo blanco . . . . .	68
C.3. Precios del ajo morado . . . . .	68
C.4. Precios del apio . . . . .	69
C.5. Precios del betabel . . . . .	69
C.6. Precios del brócoli . . . . .	69
C.7. Precios del cacahuete . . . . .	69
C.8. Precios de la calabacita italiana . . . . .	70
C.9. Precios de la calabaza de castilla . . . . .	70
C.10. Precios del camote . . . . .	70
C.11. Precios de la caña . . . . .	70
C.12. Precios de la cebolla bola . . . . .	71
C.13. Precios de la cebolla bola grande . . . . .	71
C.14. Precios del champiñón . . . . .	71
C.15. Precios del chayote . . . . .	71
C.16. Precios del chayote sin espinas . . . . .	72
C.17. Precios del chícharo . . . . .	72
C.18. Precios del chile ancho . . . . .	72
C.19. Precios del chile chilaca . . . . .	72
C.20. Precios del chile de árbol seco . . . . .	73
C.21. Precios del chile guajillo . . . . .	73
C.22. Precios del chile jalapeño . . . . .	73
C.23. Precios del chile pasilla . . . . .	73
C.24. Precios del chile pimiento morrón . . . . .	74
C.25. Precios del chile poblano . . . . .	74
C.26. Precios del chile puya seco . . . . .	74
C.27. Precios del chile serrano . . . . .	74
C.28. Precios del cilantro . . . . .	75

C.29. Precios de la ciruela roja . . . . .	75
C.30. Precios de la col mediana . . . . .	75
C.31. Precios de la coliflor grande . . . . .	75
C.32. Precios de la coliflor mediana . . . . .	76
C.33. Precios del durazno amarillo . . . . .	76
C.34. Precios del ejote . . . . .	76
C.35. Precios del elote . . . . .	76
C.36. Precios del elote grande . . . . .	77
C.37. Precios del epazote . . . . .	77
C.38. Precios de la espinaca . . . . .	77
C.39. Precios de la fresa . . . . .	77
C.40. Precios de la guayaba . . . . .	78
C.41. Precios de la jícama . . . . .	78
C.42. Precios del kiwi . . . . .	78
C.43. Precios de la lechuga romanita grande . . . . .	78
C.44. Precios de la lima . . . . .	79
C.45. Precios del limón con semilla #3 . . . . .	79
C.46. Precios del limón con semilla #5 . . . . .	79
C.47. Precios de la mandarina . . . . .	79
C.48. Precios del mango Ataulfo . . . . .	80
C.49. Precios del mango Haden . . . . .	80
C.50. Precios del mango manila . . . . .	80
C.51. Precios de la manzana golden delicious . . . . .	80
C.52. Precios de la manzana red delicious . . . . .	81
C.53. Precios del melón Cantaloupe #27 . . . . .	81
C.54. Precios del melón Cantaloupe sin clasificación . . . . .	81
C.55. Precios de la naranja Valencia grande . . . . .	81
C.56. Precios de la naranja Valencia mediana . . . . .	82
C.57. Precios del nopal . . . . .	82
C.58. Precios del nopal grande . . . . .	82
C.59. Precios de la papa alpha . . . . .	82
C.60. Precios de la papaya maradol . . . . .	83
C.61. Precios del pepino . . . . .	83
C.62. Precios de la Pera D´anjou #100 . . . . .	83
C.63. Precios de la piña mediana . . . . .	83
C.64. Precios del plátano dominico . . . . .	84
C.65. Precios del plátano macho . . . . .	84
C.66. Precios del plátano Tabasco . . . . .	84
C.67. Precios de la sandía rayada . . . . .	84
C.68. Precios de la sandía sangría . . . . .	85
C.69. Precios del tejocote . . . . .	85
C.70. Precios del tomate saladette . . . . .	85
C.71. Precios del tomate verde . . . . .	85
C.72. Precios de la toronja roja . . . . .	86
C.73. Precios de la tuna blanca . . . . .	86
C.74. Precios de la uva globo . . . . .	86
C.75. Precios de la uva Perlette . . . . .	86
C.76. Precios de la uva superior . . . . .	87

C.77. Precios de la uva Thompson . . . . .	87
C.78. Precios de la zanahoria leña . . . . .	87
C.79. Precios de la zanahoria mediana . . . . .	87
C.80. Precios de la zanahoria polvo . . . . .	88
E.1. Serie semanal del aguacate Hass . . . . .	96
E.2. Serie semanal del cacahuate . . . . .	96
E.3. Serie semanal de la calabacita italiana . . . . .	97
E.4. Serie semanal del chayote sin espinas . . . . .	97
E.5. Serie semanal del chile serrano . . . . .	97
E.6. Serie semanal de la coliflor grande . . . . .	98
E.7. Serie semanal del epazote . . . . .	98
E.8. Serie semanal de la jícama . . . . .	98
E.9. Serie semanal del kiwi . . . . .	99
E.10. Serie semanal de la lima . . . . .	99
E.11. Serie semanal del limón #5 . . . . .	99
E.12. Serie semanal de la naranja Valencia grande . . . . .	100
E.13. Serie semanal del nopal grande . . . . .	100
E.14. Serie semanal de la piña mediana . . . . .	100
E.15. Serie semanal del plátano macho . . . . .	101
E.16. Serie semanal del plátano Tabasco . . . . .	101
E.17. Serie semanal del tomate saladette . . . . .	101
E.18. Serie semanal de la toronja roja . . . . .	102
F.1. Serie estacionaria del aguacate Hass . . . . .	103
F.2. Serie estacionaria del cacahuate . . . . .	103
F.3. Serie estacionaria de la calabacita italiana . . . . .	104
F.4. Serie estacionaria del chayote sin espinas . . . . .	104
F.5. Serie estacionaria del chile serrano . . . . .	104
F.6. Serie estacionaria de la coliflor grande . . . . .	105
F.7. Serie estacionaria del epazote . . . . .	105
F.8. Serie estacionaria de la jícama . . . . .	105
F.9. Serie estacionaria del kiwi . . . . .	106
F.10. Serie estacionaria de la lima . . . . .	106
F.11. Serie estacionaria del limón #5 . . . . .	106
F.12. Serie estacionaria de la naranja Valencia grande . . . . .	107
F.13. Serie estacionaria del nopal grande . . . . .	107
F.14. Serie estacionaria de la piña mediana . . . . .	107
F.15. Serie estacionaria del plátano macho . . . . .	108
F.16. Serie estacionaria del plátano Tabasco . . . . .	108
F.17. Serie estacionaria del tomate saladette . . . . .	108
F.18. Serie estacionaria de la toronja roja . . . . .	109
G.1. Análisis de correlación del aguacate Hass . . . . .	110
G.2. Análisis de correlación del cacahuate . . . . .	110
G.3. Análisis de correlación de la calabacita italiana . . . . .	111
G.4. Análisis de correlación del chayote sin espinas . . . . .	111
G.5. Análisis de correlación del chile serrano . . . . .	112
G.6. Análisis de correlación de la coliflor grande . . . . .	112

G.7. Análisis de correlación del epazote . . . . .	113
G.8. Análisis de correlación de la jícama . . . . .	113
G.9. Análisis de correlación del kiwi . . . . .	114
G.10. Análisis de correlación de la lima . . . . .	114
G.11. Análisis de correlación del limón #5 . . . . .	115
G.12. Análisis de correlación de la naranja Valencia grande . . . . .	115
G.13. Análisis de correlación del nopal grande . . . . .	116
G.14. Análisis de correlación de la piña mediana . . . . .	116
G.15. Análisis de correlación del plátano macho . . . . .	117
G.16. Análisis de correlación del plátano Tabasco . . . . .	117
G.17. Análisis de correlación del tomate saladette . . . . .	118
G.18. Análisis de correlación de la toronja roja . . . . .	118
H.1. Aguacate Hass . . . . .	119
H.2. Cacahuete . . . . .	119
H.3. Calabacita italiana . . . . .	120
H.4. Chayote sin espinas . . . . .	120
H.5. Chile serrano . . . . .	121
H.6. Coliflor grande . . . . .	121
H.7. Epazote . . . . .	122
H.8. Jícama . . . . .	122
H.9. Kiwi . . . . .	123
H.10. Lima . . . . .	123
H.11. Limón con semilla #5 . . . . .	124
H.12. Naranja Valencia grande . . . . .	124
H.13. Nopal grande . . . . .	125
H.14. Piña mediana . . . . .	125
H.15. Plátano macho . . . . .	126
H.16. Plátano Tabasco . . . . .	126
H.17. Tomate saladette . . . . .	127
H.18. Toronja roja . . . . .	127
I.1. Predicciones del aguacate Hass . . . . .	128
I.2. Predicciones del cacahuete . . . . .	128
I.3. Predicciones de la calabacita italiana . . . . .	129
I.4. Predicciones del chayote sin espinas . . . . .	129
I.5. Predicciones del chile serrano . . . . .	129
I.6. Predicciones de la coliflor grande . . . . .	130
I.7. Predicciones del epazote . . . . .	130
I.8. Predicciones de la jícama . . . . .	130
I.9. Predicciones del kiwi . . . . .	131
I.10. Predicciones de la lima . . . . .	131
I.11. Predicciones del limón #5 . . . . .	131
I.12. Predicciones de la naranja Valencia grande . . . . .	132
I.13. Predicciones del nopal grande . . . . .	132
I.14. Predicciones de la piña mediana . . . . .	132
I.15. Predicciones del plátano macho . . . . .	133
I.16. Predicciones del plátano Tabasco . . . . .	133
I.17. Predicciones del tomate saladette . . . . .	133

I.18. Predicciones de la toronja roja . . . . .	134
J.1. Aguacate Hass . . . . .	135
J.2. Cacahuete . . . . .	135
J.3. Calabacita italiana . . . . .	136
J.4. Chayote sin espinas . . . . .	136
J.5. Chile serrano . . . . .	136
J.6. Coliflor grande . . . . .	137
J.7. Epazote . . . . .	137
J.8. Jícama . . . . .	137
J.9. Kiwi . . . . .	138
J.10. Lima . . . . .	138
J.11. Limón son semilla #5 . . . . .	138
J.12. Naranja Valencia grande . . . . .	139
J.13. Nopal grande . . . . .	139
J.14. Piña mediana . . . . .	139
J.15. Plátano macho . . . . .	140
J.16. Plátano Tabasco . . . . .	140
J.17. Tomate saladette . . . . .	140
J.18. Toronja roja . . . . .	141

Dirección General de Bibliotecas UAQ

# CAPÍTULO 1

## Introducción

La agricultura ha sido una actividad muy importante desde que los seres humanos empezaron a dejar de ser nómadas para convertirse en sedentarios, incluso en la actualidad continua teniendo un papel primordial dentro de la sociedad. El cultivo nos otorga una parte de los alimentos que forman parte de nuestra dieta diaria, de los cuales obtenemos nutrientes importantes. Además, la agricultura nos provee de materias primas para otros sectores, por ejemplo, el algodón, la manzanilla y la sábila. Como resultado de esto, la agricultura tiene impactos políticos y económicos.

En este sentido, en México, la agricultura en el 2017 aportó un 3.35 % del producto interno bruto (PIB) (Paredes-Garcia et al., 2019). Por otro lado, entre 2000 y 2018, las exportaciones de productos agrícolas de México a los Estados Unidos aumentaron en un 405 % (Canales et al., 2019). En el año 2018, la balanza comercial agroindustrial mexicana resultó en 34,255 millones de dólares de las exportaciones y 28,416 millones de dólares de las importaciones (Paredes-Garcia et al., 2019).

Sin embargo, existen diversas problemáticas entorno a la agricultura como reducir el impacto de la actividad en el ecosistema, el desafío de proveer de alimentos a la población mundial, los problemas del cambio climático y sus efectos en el cultivo. En particular, uno de los más grandes desafíos que enfrenta la agricultura es la volatilidad de los precios de frutas y verduras el cual potencia la precarización del sector (Xie and Wang, 2017).

No obstante, el uso de tecnología se ha convertido en una posible solución potencial de la precarización de la agricultura debido a que al aplicar los conocimientos y las herramientas desarrolladas por la civilización se han podido resolver algunos problemas y mitigar otros. Como es el caso de los modelos para la predicción de precios los cuales permiten al productor planear su producción para aumentar su ganancia.

Desde el punto de vista estadístico, los modelos para la predicción de precios pueden ser divididos en métodos de regresión lineal y métodos de series de tiempo. Sin embargo, el

rápido desarrollo del poder computacional ha permitido que los métodos no lineales lleguen a ser aplicables para el modelado y la predicción, un ejemplo de ello son las redes neuronales, estos modelos matemáticos no lineales, intentan simular la estructura y los aspectos funcionales de las redes neuronales biológicas basándose en el almacenamiento distributivo y el procesamiento paralelo de mensajes (Li et al., 2010).

Las redes neuronales han evidenciado ser un método que puede ser utilizado en la predicción de precios en distintas áreas debido a que pueden aprender patrones y tendencias, así como observar relaciones ocultas en un conjunto de datos, aún si las relaciones son desconocidas o son difíciles de identificar (Wang and Wang, 2016). Aunque en la predicción de precios de alimentos también se ha empleado el uso de redes neuronales, en la mayoría de los casos estudiados sólo se realiza el análisis para un producto, pero para que la predicción de los precios sea una estrategia factible es necesario tener la información de distintos productos simultáneamente.

En esta tesis se predice los precios de varias frutas y hortalizas del mercado de abastos del estado de Querétaro implementando dos métodos de predicción los cuales son el método de series de tiempo y una red neuronal recurrente de manera que se puede realizar una comparación por medio del error cuadrado medio de cada modelo.

En el capítulo 2 se explican los modelos para la predicción de precios y se exponen los conceptos importantes de cada modelo. Además se presenta la información básica de la metodología usada de cada método para la predicción de precios.

En el apartado del capítulo 3 se presenta a detalle la metodología empleada para la obtención de datos, así como para realizar el análisis exploratorio. Además, se explica el proceso para ajustar un modelo a las series de tiempo y entrenar el modelo de la red neuronal recurrente.

Finalmente, en el capítulo 4 se muestran los resultados obtenidos de la predicción de los precios de cada producto así como el error cuadrado medio de cada metodología. También se muestra el resultado de la comparación del error cuadrado medio del modelo de redes neuronales recurrentes con el error cuadrado medio de los modelos de series de tiempo.

## 1.1 Hipótesis y objetivos

### 1.1.1 Hipótesis

El modelo de redes neuronales recurrentes empleado para la predicción de precios semanales promedio de frutas y hortalizas en México reportados por el SNIIM, presenta un error cuadrado medio mayor que los modelos de series de tiempo basados en la metodología Box-Jenkins.

### 1.1.2 Objetivos

Objetivo general

Estimar el error cuadrado medio del modelo de redes neuronales recurrentes para la predicción de los precios semanales promedio de frutas y hortalizas en México reportados por el Sistema Nacional de Información e Integración de Mercados en México (SNIIM).

Objetivos específicos

1. Extraer los precios de las frutas y hortalizas del sitio web del SNIIM.
2. Determinar los productos viables para la implementación del análisis.
3. Realizar el análisis estadístico exploratorio de los datos.
4. Entrenar y validar el modelo de redes neuronales recurrentes para cada producto determinado como viable.
5. Comparar el error cuadrado medio del modelo de redes neuronales recurrentes con el error cuadrado medio de los modelos de series de tiempo basados en la metodología Box-Jenkins.

## CAPÍTULO 2

### Marco teórico

#### 2.1 Web scraping

*Web scraping*, también conocido como *screen scraping*, *data mining* o *web harvesting* es una técnica para extraer datos de la *World Wide Web (WWW)* y guardarlos en un archivo o base de datos para su posterior recuperación o análisis. Esto se logra más comúnmente escribiendo un programa automatizado que consulta un servidor web, solicita datos (generalmente en forma de HTML y otros archivos que componen las páginas web) y luego analiza esos datos para extraer la información necesaria. El método de *web scraping* abarca una amplia variedad de técnicas y tecnologías de programación, como análisis de datos, análisis de lenguaje natural y seguridad de la información (Mitchell, 2018).

Si la única forma de acceder a Internet es a través de un navegador, se está perdiendo una gran variedad de posibilidades. Aunque los navegadores son útiles para ejecutar JavaScript, mostrar imágenes y organizar objetos en un formato más legible para el ser humano, el método de *Web scraping* es excelente para recopilar y procesar grandes cantidades de datos rápidamente debido a que en lugar de ver una página a la vez a través de la ventana estrecha de un monitor, puede ver bases de datos que abarcan miles o incluso millones de páginas a la vez (Mitchell, 2018). La técnica de *Web scraping* se puede utilizar para una amplia variedad de escenarios, como obtener contactos, monitorear y comparar cambios de precios, recopilar reseñas de productos, recopilar listados de bienes raíces, monitorear datos meteorológicos, detectar cambios en el sitio web e integrar de datos web (Zhao, 2017).

El proceso de extracción de datos de Internet se puede dividir en dos pasos secuenciales; el primero es adquirir los recursos web y luego extraer la información deseada a partir de los datos adquiridos. Específicamente, un programa de *web scraping* comienza componiendo una solicitud HTTP para adquirir recursos de un sitio web específico. Una vez que la solicitud

es recibida y procesada con éxito por el sitio web de destino, el recurso solicitado se recuperará del sitio web y luego se enviará de vuelta a el programa de *web scraping* dado. Una vez que se descargan los datos web, el proceso de extracción pasa a analizar, reformar y organizar los datos de forma estructurada (Zhao, 2017).

## 2.2 Métodos de predicción

### 2.2.1 Regresión lineal

El método de regresión lineal existe desde hace mucho tiempo, sin embargo, sigue siendo un método de aprendizaje estadístico útil y ampliamente utilizado, particularmente la regresión lineal es una herramienta útil para predecir una respuesta cuantitativa. En este método el enfoque es predecir una respuesta cuantitativa con base en la información de las variables de explicación, suponiendo que hay aproximadamente una relación lineal entre las variables de explicación y la variable respuesta. Matemáticamente, podemos escribir esta relación lineal como

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon \quad (2.1)$$

donde  $Y$  es la variable respuesta,  $X_j$  son las variables de explicación,  $\beta_0$  es el término constante,  $\beta_j (j > 0)$  son los coeficientes de la relación entre la variable  $X_j$  y la respuesta,  $n$  es el número de parámetros independientes a tener en cuenta en la regresión y  $\epsilon$  es un término aleatorio de media cero (James et al., 2013).

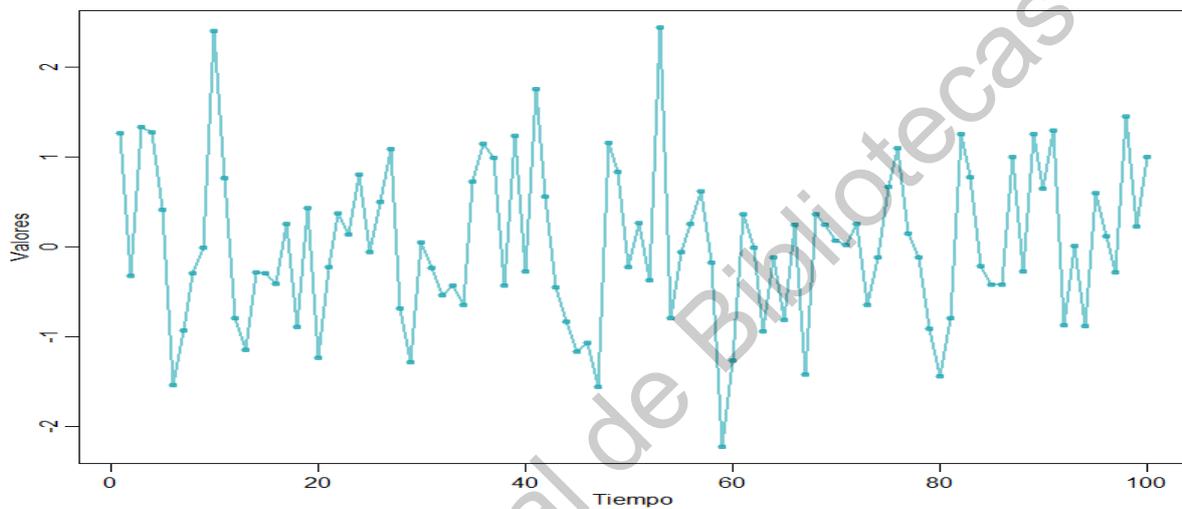
### 2.2.2 Series de tiempo

El análisis de una muestra correlacionada que ha sido observada en diferentes instantes en el tiempo conduce a problemas nuevos y únicos en el modelado, por consiguiente, se ha creado un nuevo enfoque sistemático llamado análisis de series de tiempo que nos permite responder las respuestas matemáticas y estadísticas planteadas por la correlación de los datos.

Con el fin de proporcionar un entorno estadístico para describir el carácter de los datos, asumimos que una serie de tiempo se puede definir como una colección de variables

aleatorias indexadas por el tiempo y se denota como  $\{X_t\}_{t \in \mathbb{N}}$  donde  $X_t$  es una variable aleatoria en el tiempo  $t$  (Shumway and Stoffer, 2017).

Al graficar una serie de tiempo, los valores de las variables aleatorias se establecen en el eje vertical, u ordenadas, con la escala de tiempo como la abscisa (Figura 2.1). Además, es común conectar los valores en períodos de tiempo adyacentes para reconstruir visualmente una serie de tiempo continua hipotética que podría haber producido esos valores (Shumway and Stoffer, 2017).



*Figura 2.1: Ruido blanco Gaussiano*

Una descripción completa de una serie de tiempo vista como una colección de  $n$  variables aleatorias en los tiempos arbitrarios  $t_1, t_2, \dots, t_n$ , para cualquier entero positivo  $n$  es dada por la función de distribución conjunta evaluada como la probabilidad de que los valores de la serie sean menores a las constantes  $c_1, c_2, \dots, c_n$ , es decir,

$$F(c_1, c_2, \dots, c_n) = P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_n} \leq c_n); \quad (2.2)$$

sin embargo, la función de distribución conjunta es una herramienta difícil de manejar para mostrar y analizar datos de series de tiempo, por lo cual se utilizan medidas descriptivas marginales como la función de distribución marginal o la función de la media (Shumway and

Stoffer, 2017). La función de la media en el instante  $t$  se define como

$$\mu_t := E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx. \quad (2.3)$$

### Autocorrelación

Por otra parte, la falta de independencia entre dos valores adyacentes  $x_s$  y  $x_t$  puede ser evaluada numéricamente, como en la estadística clásica, utilizando las nociones de covarianza y correlación. Suponiendo que la varianza de  $x_t$  es finita, tenemos que la función de autocovarianza se define como

$$\gamma(s, t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)], \quad (2.4)$$

para todo  $s$  y  $t$ . La autocovarianza mide la dependencia lineal entre dos puntos de la misma serie observada en diferentes momentos. Las series muy suaves exhiben funciones de autocovarianza que permanecen grandes incluso cuando la  $t$  y  $s$  están muy separadas (Shumway and Stoffer, 2017). A raíz de esto, es conveniente trabajar con una medida de dependencia acotada como la función de autocorrelación. La función de autocorrelación (**ACF**) es definida como

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}. \quad (2.5)$$

Sin embargo, la función de autocorrelación no es capaz de evaluar ciertas características de la serie de tiempo, por lo cual es necesario utilizar otras medidas de dependencia como la función de autocorrelación parcial. La función de autocorrelación parcial (**PACF**) de un proceso estacionario,  $\{x_t\}$ , denotada  $\Phi_{hh}$  con  $h = 1, 2, 3, \dots$  es definida como

$$\Phi_{11} = \text{corr}(x_1, x_0) = \rho(1) \quad (2.6)$$

y

$$\Phi_{hh} = \text{corr}(x_h - \hat{x}_h, x_0 - \hat{x}_0) = \rho(1), h \geq 2, \quad (2.7)$$

donde  $\hat{x}_h$  es la regresión de  $x_h$  sobre  $\{x_1, x_2, \dots, x_{h-1}\}$  y  $\hat{x}_0$  es la regresión de  $x_0$  sobre

$\{x_1, x_2, \dots, x_{h-1}\}$ . Así, debido a la estacionariedad,  $\Phi_{hh}$ , es la correlación entre  $x_{t+h}$  y  $x_t$  con la dependencia lineal de  $\{x_1, x_2, \dots, x_{h-1}\}$  eliminada (Shumway and Stoffer, 2017).

## Estacionariedad

Una serie de tiempo es estrictamente estacionaria si el comportamiento probabilístico de cada colección de valores  $\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$  es idéntico a la del conjunto  $\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$ , Esto es, una serie de tiempo es estrictamente estacionaria si

$$P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_k} \leq c_k) = P(x_{t_1+h} \leq c_1, x_{t_2+h} \leq c_2, \dots, x_{t_k+h} \leq c_k),$$

para todo  $k = 1, 2, \dots$ , para todo  $h = 0, \pm 1, \pm 2, \dots$ , para todos los tiempos  $t_1, t_2, \dots, t_k$  y para todo los números  $c_1, c_2, \dots, c_k$ . No obstante, la estacionariedad estricta es demasiado fuerte para la mayoría de las aplicaciones y es difícil de evaluar a partir de un único conjunto de datos, por lo cual se define una versión más suave la estacionariedad (Shumway and Stoffer, 2017).

Una serie de tiempo débilmente estacionaria es un proceso de varianza finita tal que

- (i) el valor de la función de la media,  $\mu_t$ , definido en (2.3) es constante y no depende del tiempo  $t$
- (ii) la función de autocovarianza,  $\gamma(s, t)$ , definido en (2.4) depende de  $s$  y  $t$  sólo a través de su diferencia  $|s - t|$ .

### 2.2.3 Redes neuronales

Una de las ramas más destacadas del campo científico de la inteligencia artificial es la que corresponde a las redes neuronales artificiales (**RNAs**), las cuales se definen como modelos computacionales que surgieron como un intento de conseguir formalizaciones matemáticas acerca de la estructura del cerebro. Las RNAs imitan la estructura del sistema nervioso, centrándose en el funcionamiento del cerebro humano, basado en el aprendizaje a través de la experiencia, con la consiguiente extracción de conocimiento a partir de la misma. El fin perseguido por una RNA es la emulación del sistema nervioso central biológico a través de

procesadores artificiales, que incluso permitan evitar fallos o errores humanos (López et al., 2008).

Las redes neuronales artificiales son herramientas simples, pero potentes y flexibles para pronosticar siempre que haya suficientes datos para el entrenamiento, una selección adecuada de las muestras de entrada y salida, un número apropiado de unidades ocultas y suficientes recursos computacionales disponibles (Catalão et al., 2007). Estos modelos también ofrecen numerosas ventajas como el manejo de datos, la asociación de memoria, la auto organización y la auto adaptación (Wang and Wang, 2016).

La adaptación es la capacidad de aprender a llevar a cabo ciertas tareas mediante un entrenamiento o en una experiencia inicial y la auto adaptación, es la capacidad de crear una organización propia de la información que recibe mediante una etapa de aprendizaje. Mientras que la memoria asociativa, se refiere a que cuando se aplica un estímulo (dato de entrada) la red responde con una salida asociada a dicha información de entrada (Matich, 2001).

## Arquitectura

Una red neuronal puede definirse como un grafo dirigido, con las siguientes propiedades:

- A cada nodo  $j$  se le asocia una variable de estado  $x_j$
- A cada conexión  $(i, j)$  de los nodos  $i$  y  $j$  se le asocia un peso  $w_{ij} \in \mathbb{R}$
- A cada nodo  $j$  se le asocia un umbral  $\theta_j$
- Para cada nodo  $j$  se define una función  $f_j(x_j, w_{ij}, \theta_j)$  que depende de los pesos de sus conexiones, del umbral y de los estados de los nodos  $i$  a él conectados.

En general, una red neuronal es un conjunto de unidades interconectadas donde cada conexión tiene un peso asociado. A cada una de esas unidades se les llama neuronas, cada neurona está conectada con otras a través de unos enlaces, estas conexiones son direccionales. Además, las neuronas se suelen agrupar en unidades estructurales que denominaremos capas (Larranaga et al., 1997), cuya agrupación conforma el sistema neuronal completo (Figura 2.2).

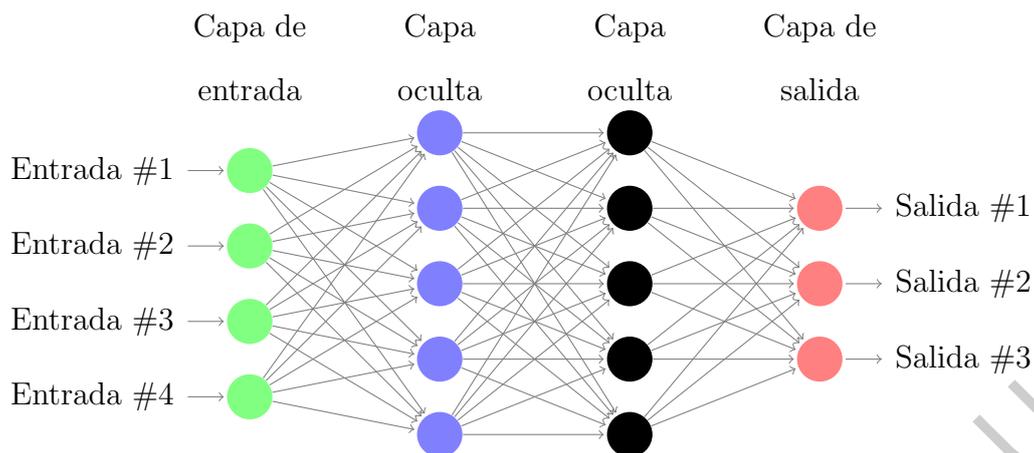


Figura 2.2: Red neuronal artificial

En una red neuronal artificial se distinguen tres tipos de capas: de entrada, de salida y ocultas. Una capa de entrada, también denominada sensorial, está compuesta por neuronas que reciben datos o señales procedentes del entorno. Una capa de salida se compone de neuronas que proporcionan la respuesta de la red neuronal (Larranaga et al., 1997). Las capas ocultas son internas a la red y no tienen contacto directo con el entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distintas maneras (Matich, 2001).

### Neuronas artificiales

La neurona artificial (Figura 2.3) trata a muchos valores de entrada como si fueran uno solo haciendo uso de la función de entrada la cual es la regla que establece el procedimiento a seguir para combinar los valores de entrada y los pesos de las conexiones que llegan a esa unidad. De esta forma, la función de entrada permite obtener el valor del potencial post-sináptico o entrada total al momento  $t$  como:

$$Net(t) = \sigma(W_j x(t)) + \theta_j(t), \quad (2.8)$$

donde  $W_j$  es un vector de pesos,  $x(t)$  son las entradas en el tiempo  $t$  y  $\sigma_j$  es una función como la suma o el máximo. Los valores de entrada se multiplican por los pesos anteriormente ingresados a la neurona. Por consiguiente, los pesos que generalmente no están restringidos

cambian la medida de influencia que tienen los valores de entrada (López et al., 2008).

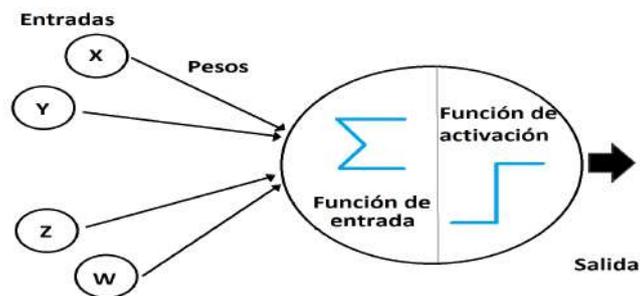


Figura 2.3: Neurona artificial

Una neurona biológica puede estar activa o inactiva; es decir, que tiene un “estado de activación”. Las neuronas artificiales (Figura 2.3) también tienen diferentes estados de activación (Matich, 2001). La función de activación calcula el estado de actividad de una neurona combinando la entrada total a la  $j$ -ésima neurona con el estado inicial de la neurona para producir un nuevo estado de activación acorde con la información recibida, aunque en muchos modelos de RNAs se considera que el estado actual de la neurona no depende de su estado previo, por lo cual la expresión se simplifica como:

$$a(t) = f(Net(t)), \quad (2.9)$$

donde  $a(t)$  es el estado de activación en el tiempo  $t$ ,  $Net$  es la entrada total en el tiempo  $t$  y  $f$  es una función determinista, y en la mayor parte de los modelos es monótona creciente y continua como la sigmodial o la identidad (López et al., 2008).

El último componente que una neurona necesita es la función de salida. El valor resultante de esta función es la salida de la neurona; por ende, la función de salida determina qué valor se transfiere a las neuronas vinculadas. Dos de las funciones de salida más comunes son la función identidad y binaria (Matich, 2001).

## Aprendizaje

La predicción con redes neuronales involucra dos etapas: entrenamiento y evaluación. Se supone que hay disponible un conjunto de entrenamiento, dado por los datos históricos, que contiene tanto las entradas como las salidas deseadas correspondientes. La selección

adecuada de entradas para el entrenamiento de la red neuronal es muy influyente para el éxito del entrenamiento (Catalão et al., 2007).

Una red neuronal debe aprender a calcular la salida correcta para cada vector de entrada en el conjunto de ejemplos. A este aprendizaje se le denomina proceso de entrenamiento o acondicionamiento, debido a que es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada (López et al., 2008).

Existen dos métodos de aprendizaje importantes que pueden distinguirse: el aprendizaje supervisado y el aprendizaje no supervisado. El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor controla la salida de la red y en caso de que ésta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, con el fin de conseguir que la salida obtenida se aproxime a la deseada (Matich, 2001).

Las redes con aprendizaje no supervisado no requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta. Estas redes deben encontrar las características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presenten en su entrada (Matich, 2001).

El método de entrenamiento más utilizado es el método del gradiente descendente en el cual se define una función que proporciona el error que comete la red en función del conjunto de pesos sinápticos. El objetivo de este método es buscar los pesos que correspondan al mínimo global de la función de error utilizando la ecuación iterativa:

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta E(w_{ij}(t)), \quad (2.10)$$

donde  $w_{ij}(t)$  es el peso que conecta la neurona  $i$  con la neurona  $j$  en el instante  $t$  y  $\Delta E(w_{ij}(t))$  es la variación del error generada por la regla de aprendizaje en el instante  $t$  (López et al., 2008).

Los cambios que se producen durante el entrenamiento se reducen a la destrucción,

modificación y creación de conexiones entre las neuronas. En los modelos de redes neuronales artificiales, la creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero. De la misma manera, una conexión se destruye cuando su peso pasa a ser cero (Matich, 2001).

Para determinar cuándo se detendrá el proceso de aprendizaje, es necesario establecer una condición de detención. El entrenamiento se detiene cuando el error observado está por debajo de un determinado umbral o puede ser cuando un cierto número de ciclos de entrenamiento hayan sido completamente corridos. Luego de alcanzarse la condición de detención, los pesos no se volverán a cambiar. Entonces podemos decir que la transformación de los datos de entrada a los de salida está resuelta. Esto se puede interpretar como una función oculta en el conjunto de la red neuronal. Esta función es exactamente la instrucción de cómo la salida será calculada a partir de un vector de entrada (Matich, 2001).

Después del proceso de entrenamiento, el paso siguiente es comprobar si la red neuronal puede resolver nuevos problemas. Por lo tanto, con el propósito de validar la red neuronal se requiere de otro conjunto de datos, denominado conjunto de prueba. Cada ejemplo del conjunto de prueba contiene los valores de las variables de entrada, con su correspondiente solución tomada; pero ahora esta solución no se le es otorgada a la red neuronal (Matich, 2001).

Un problema central en el aprendizaje automático es cómo hacer un algoritmo que funcione bien no solo en los datos de entrenamiento, sino también en las nuevas entradas. Muchas estrategias utilizadas en el aprendizaje automático están diseñadas explícitamente para reducir el error de prueba, posiblemente a expensas de un mayor error de entrenamiento. Estas estrategias se conocen colectivamente como regularización (Goodfellow et al., 2016).

La estrategia conocida como early stopping es probablemente la forma de regularización más utilizada en el aprendizaje profundo. Su popularidad se debe tanto a su eficacia como a su sencillez. El algoritmo de la early stopping está especificado en el siguiente conjunto de pasos:

Se define la paciencia  $p$  como el número de veces que se observará el error de validación antes de terminar el algoritmo.

Se inicializan los parámetros de la red neuronal  $\theta_0$ .

$$\theta, \theta^* \leftarrow \theta_0$$

$$i, j \leftarrow 0$$

$$i^* \leftarrow i$$

$$\nu \leftarrow \infty$$

**Mientras** ( $j < p$ ) **Entonces**

$$\theta \leftarrow \text{AlgoritmoEntrenamiento}(\theta)$$

$$i \leftarrow i + 1$$

$$\nu' \leftarrow \text{ErrorValidación}(\theta)$$

**Si** ( $\nu' < \nu$ ) **Entonces**

$$j \leftarrow 0$$

$$\theta^* \leftarrow \theta$$

$$i^* \leftarrow i$$

$$\nu \leftarrow \nu'$$

**Si no**

$$j \leftarrow j + 1$$

**Fin Si**

**Fin Mientras**

Este algoritmo es una estrategia general que funciona bien con una variedad de algoritmos de entrenamiento y formas de cuantificar el error en el conjunto de validación (Goodfellow et al., 2016).

## 2.3 Modelos

### 2.3.1 Modelo autorregresivo integrado de media móvil

El objetivo principal del análisis de series de tiempo es desarrollar modelos matemáticos que proporcionen descripciones plausibles para los datos en la muestra. Los modelos autorregresivos se basan en la idea de que el valor actual de la serie puede explicarse en función de  $p$  valores pasados. Matemáticamente, un modelo autorregresivo de orden  $p$ , abreviado **AR(p)** es de la forma

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \quad (2.11)$$

donde  $x_t$  es una serie de tiempo estacionaria,  $\alpha$  es una constante,  $\phi_1, \dots, \phi_p$  son los parámetros del modelo con  $\phi_p \neq 0$  y  $w_t \sim wn(0, \sigma_w^2)$  (Shumway and Stoffer, 2017).

Como alternativa a la representación autorregresiva en la que se supone que el lado izquierdo de la ecuación es una combinación lineal, el modelo de media móvil asume que el ruido blanco  $w_t$  en el lado derecho de la ecuación definida se combina linealmente para formar los datos observados. El modelo de media móvil de orden  $q$ , abreviado **MA(q)**, se define como

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}, \quad (2.12)$$

donde  $\theta_1, \dots, \theta_q$  son los parámetros del modelo con  $\theta_q \neq 0$  y  $w_t \sim wn(0, \sigma_w^2)$  (Shumway and Stoffer, 2017).

Para lograr una mayor flexibilidad en el ajuste de la serie de tiempo se incluyen los términos del modelo autorregresivo y del modelo de media móvil en un único modelo. La mezcla de ambos modelos crea un nuevo modelo conocido como modelo autorregresivo de media móvil (**ARMA**). Una serie de tiempo  $\{X_t; t \in \mathbb{Z}\}$  es **ARMA(p,q)** si es estacionaria y

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \quad (2.13)$$

donde  $\theta_q \neq 0$ ,  $\phi_p \neq 0$  y  $\sigma_w^2 > 0$ . Los parámetros  $p$  y  $q$  son llamados orden de la parte

autorregresiva y del promedio móvil, respectivamente (Box et al., 2015). La ecuación 2.13 supone que  $x_t$  tiene media cero.

Actualmente muchas series de tiempo encontradas en la industria o en los negocios exhiben un comportamiento no estacionario y no varían sobre una media fija. Sin embargo, tales series pueden exhibir un comportamiento homogéneo a lo largo del tiempo (Box et al., 2015). Como resultado se creó el modelo autorregresivo integrado de media móvil (**ARIMA**), el cual es una ampliación de la clase de modelos ARMA que incluye diferenciación. Una serie de tiempo  $\{X_t; t \in \mathbb{Z}\}$  es **ARIMA**(p,d,q) si

$$\nabla^d x_t = (1 - B)^d x_t \quad (2.14)$$

es **ARMA**(p,q). El operador de retraso o retroceso ( $B$ ) opera en un elemento de una serie de tiempo para producir el elemento anterior, es decir,  $Bx_t = x_{t-1}$ .

#### Método estacional

Normalmente, la dependencia en el pasado tiende a ocurrir con más fuerza en múltiplos de algún rezago estacional subyacente, por consiguiente, se realizaron varias modificaciones al modelo ARIMA para tener en cuenta el comportamiento estacional, creando el modelo multiplicativo autorregresivo estacional integrado de media móvil o modelo (**SARIMA**) viene dado por

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \delta + \Theta_Q(B^s)\theta(B)w_t \quad (2.15)$$

donde  $w_t \sim wn(0, \sigma_w^2)$  y  $s$  es el período estacional. El modelo es denotado como **ARIMA**(p, d, q)  $\times$  (P, D, Q)<sub>s</sub>. Los componentes autorregresivos y de media móvil no estacionales están representados por polinomios  $\phi(B)$  y  $\theta(B)$  de órdenes  $p$  y  $q$ , respectivamente, y los componentes autoregresivos y de media móvil estacionales están dados por  $\Phi_P(B^s)$  y  $\Theta_Q(B^s)$  de órdenes  $P$  y  $Q$ . La diferencia no estacional está representada por  $\nabla^d = (1 - B)^d$  mientras que la diferencia estacional está representada por  $\nabla_s^D = (1 - B^s)^D$  (Shumway and Stoffer, 2017).

## Metodología Box-Jenkins

La metodología Box-Jenkins es un procedimiento de análisis estadístico para ajustar a una serie de tiempo un modelo autorregresivo integrado de media móvil. El método original utiliza un enfoque de modelado iterativo de tres etapas, las tres etapas del modelado iterativo son las siguientes:

1. Identificación de un modelo: Emplear datos y conocimientos del sistema para sugerir un proceso apropiado.
2. Estimación de parámetros: El modelo propuesto es ajustado a los datos y los parámetros se estiman.
3. Validación del modelo: El diagnóstico se aplica con el objetivo de descubrir una posible falta de ajuste y diagnosticar la causa.

Si la estimación es inadecuada, el ciclo iterativo de identificación, estimación y validación se repite hasta que una se encuentre una representación adecuada (Box et al., 2015).

### 2.3.2 Red neuronal recurrente ELMAN

Las redes neuronales recurrentes (**RNR**) son sistemas dinámicos que no tienen una estructura de capas, sino que permiten conexiones arbitrarias entre las neuronas, incluso pudiendo crear ciclos (Cruz et al., 2007). Las redes neuronales recurrentes usan conexiones de retroalimentación para modelar dependencias espaciales y temporales entre las entradas y salidas para hacer que los estados iniciales y los estados pasados de las neuronas tomen parte de una serie de procesamientos (Wang and Wang, 2016).

Existen tres tipos de tareas esenciales que se pueden realizar con este tipo de redes, la primera es el reconocimiento de secuencias, es decir, se produce un patrón de salida particular cuando se especifica una secuencia de entrada; la segunda es la reproducción de secuencias, esto es que la red debe ser capaz de generar el resto de una secuencia cuando ve parte de ella y por último, la asociación temporal, las redes recurrentes tienen la posibilidad de formar una memoria a corto plazo y pueden aprender a comprimir toda la historia en un espacio de baja

dimensión (Cruz et al., 2007). Esta habilidad permite aplicar este tipo de redes a la predicción de series de tiempo con resultados satisfactorios.

Como una red neuronal recurrente especial, la red neuronal recurrente tipo ELMAN (**ERNN**) es un sistema de control predictivo variable en el tiempo que se desarrolla con la capacidad de mantener la memoria de eventos recientes para predecir la salida futura. En esta red, a las salidas de la capa oculta se les permite retroalimentarse sobre sí mismas a través de una capa de amortiguación, llamada capa recurrente. Esta retroalimentación permite a la ERNN aprender, reconocer y generar patrones temporales y espaciales. Cada neurona en la capa oculta está conectada a una sola de neurona de la capa recurrente a través de un peso constante de valor uno. De ahí la capa recurrente virtualmente constituye una copia del estado de la capa oculta un instante anterior (Wang and Wang, 2016).

#### Arquitectura

En la Figura 2.4 se muestra la red neuronal recurrente ELMAN donde el número de neuronas en la capa de entrada es  $m$ , en la capa oculta es  $n$  y en la capa de salida sólo hay una neurona. Sea  $x_i$  el vector de entrada de la neurona  $i$  de la capa de entrada,  $z_j$  la salida de neurona  $j$  de la capa oculta,  $u_j$  la neurona  $j$  de la capa recurrente y  $y_1$  la salida de la red. Además,  $w_{ij}$  es el peso que conecta el nodo  $i$  de la capa de entrada con el nodo  $j$  de la capa oculta. Mientras que  $c_j$  es el peso que conectan el nodo  $j$  de la capa oculta con su nodo de la capa recurrente y  $v_j$  es el peso que conectan el nodo  $j$  de la capa oculta a el nodo de la capa de salida.

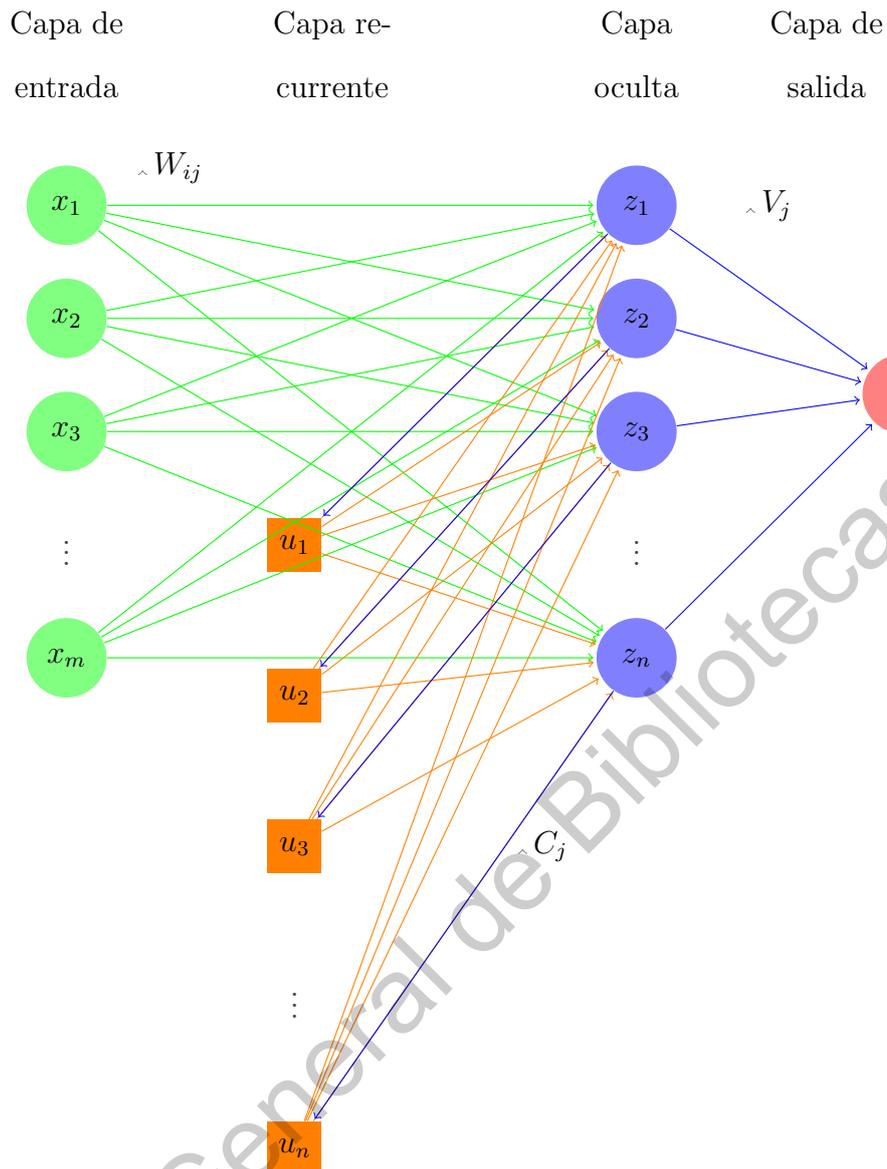


Figura 2.4: Red neuronal recurrente ELMAN

### Aprendizaje

Para realizar el proceso de entrenamiento de la red neuronal recurrente ELMAN se emplea un algoritmo de aprendizaje supervisado llamado propagación hacia atrás el cual minimiza el error global usando el método de cálculo del gradiente desde la capa de salida hasta la capa de entrada, esto permite que los pesos sobre las conexiones de las neuronas ubicadas en las capas ocultas cambien durante el entrenamiento (Rojas, 1996). Como la propagación hacia atrás es un algoritmo de aprendizaje supervisado se necesita un conjunto de datos de entrenamiento de la forma  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$  en donde  $X_k$  es un vector con los datos

de entrada de la  $k$ -ésima iteración y  $Y_k$  es un vector con los resultados deseados de la  $k$ -ésima iteración.

El algoritmo de la propagación hacia atrás esencialmente se resume en el siguiente conjunto de pasos:

1. Calcular la salida de la red del vector de entradas.
2. Calcular el valor de la función de pérdida.
3. Calcular las derivadas parciales del error con respecto a los pesos.
4. Ajustar los pesos de cada neurona para reducir el error.
5. Repetir el proceso varias veces.

En la red neuronal recurrente de Elman (Figura 2.4) para calcular la salida de la red se necesita pasar por tres etapas. En primer lugar, se necesita calcular el valor de las neuronas recurrentes definido por la ecuación (2.16) como el valor de la neurona oculta en la iteración anterior.

$$u_j(k) = z_j(k-1), \quad j = 1, 2, \dots, m; \quad \forall k \in \mathbb{N} \quad (2.16)$$

En segunda instancia, se calculan las salidas de las neuronas ocultas en la iteración  $k$  definidas como

$$z_j(k) = f_H\left(s_H(w_{ij} x_i(k)) + s_H(c_j u_j(k))\right), \quad j = 1, 2, \dots, m; \quad \forall k \in \mathbb{N}. \quad (2.17)$$

donde  $f_H$  es la función de activación de la capa oculta y  $s_H$  es la función de entrada de la capa oculta.

Por último, se establece el valor de la neurona de salida en la  $k$ -ésima iteración como

$$y_1 = f_O\left(s_O(v_j z_j(k))\right), \quad j = 1, 2, \dots, m; \quad \forall k \in \mathbb{N}; \quad (2.18)$$

donde  $f_O$  es la función de activación de la capa de salida y  $s_O$  es la función de entrada de la capa de salida.

El objetivo principal del algoritmo de aprendizaje es minimizar el valor de la función de pérdida hasta que alcance el valor mínimo preestablecido por el proceso de aprendizaje. La función de pérdida de una muestra del conjunto de entrenamiento en la red **ERNN** se define como

$$E(n) = \frac{1}{2}\phi(n)(d_n - y_n)^2, \quad (2.19)$$

donde  $n$  es el tiempo de la muestra ( $n = 1, \dots, N$ ),  $y_n$  es el valor real,  $d_n$  es la salida de la red y  $\phi(n)$  es la función estocástica efectiva del tiempo. La función estocástica efectiva del tiempo provee cada dato con un peso dependiendo del momento en que ocurrió y está definida como

$$\phi(n) = \frac{1}{\beta} \exp \int_{n_0}^n \mu(t) dt + \int_{n_0}^n \sigma(t) dB(t), \quad (2.20)$$

donde  $\beta$  ( $\beta > 0$ ) es el coeficiente de resistencia del tiempo,  $n_0$  es el tiempo más antiguo de los datos en el conjunto de entrenamiento y  $n$  es un tiempo arbitrario en el conjunto de entrenamiento de datos,  $\mu(t)$  es la función de deriva,  $\sigma(t)$  es la función de volatilidad y  $B(t)$  es el movimiento browniano estándar (Wang and Wang, 2016). Además, el error global por cada iteración del conjunto de entrenamiento en la red está dado por la función de pérdida  $E$  definida como

$$E = \frac{1}{N} \sum_{n=1}^N E(n), \quad (2.21)$$

la cual es una especie de error cuadrático ponderado por los pesos  $\phi(n)$ .

El algoritmo de propagación hacia atrás busca el mínimo de la función de error en espacio de los pesos utilizando el método de descenso de gradiente. La combinación de pesos que minimizan la función de error se considera una solución del problema del aprendizaje. Dado que este método requiere el cálculo del gradiente de la función de error en cada paso de iteración, es necesario que la función de error sea diferenciable respecto a los parámetros de los pesos de la red (Rojas, 1996).

Para los pesos en la capa de entrada, el gradiente viene dado por:

$$\Delta W_{ij}(k) = -\eta \frac{\partial E(t_n)}{\partial W_{ij}} = \eta (d_{t_n} - y_{t_n}) v_j \phi(t_n) f'_H \left( s_H(w_{ij} x_i(k)) + \text{Net}(c_j u_j(k)) \right) x_{it_n}, \quad (2.22)$$

para los pesos de los nodos en la capa recurrente, el gradiente está dado por:

$$\Delta C_j(k) = -\eta \frac{\partial E(t_n)}{\partial C_j} = (d_{t_n} - y_{t_n}) v_j \phi(t_n) f'_H \left( s_H(w_{ij} x_i(k)) + \text{Net}(c_j u_j(k)) \right) u_{jt_n} \quad (2.23)$$

y para los pesos de los nodos en la capa oculta, el gradiente está dado por:

$$\Delta V_j(k) = -\eta \frac{\partial E(t_n)}{\partial V_j} = \eta (d_{t_n} - y_{t_n}) \phi(t_n) f_H \left( s_H(w_{ij} x_i(k)) + \text{Net}(c_j u_j(k)) \right) \quad (2.24)$$

donde  $0 < \eta < 1$  es un parámetro conocido como la tasa de aprendizaje (Wang and Wang, 2016).

## 2.4 Simulación numérica

La función estocástica efectiva del tiempo es una característica importante en este modelo debido a que pondera el peso que tiene cada uno de los errores de la salida respecto al tiempo, ocasionando que los datos más actuales aporten más al error global. Sin embargo, para que esta función sea viable en la aplicación del modelo de la red neuronal recurrente ELMAN es necesario recurrir a la simulación numérica.

Considerando la ecuación 2.20 de la función estocástica efectiva del tiempo se encuentran tres componentes importantes a calcular: la función de deriva  $\mu(t)$ , la función de volatilidad  $\sigma(t)$  y el movimiento browniano estándar  $B(t)$ . La función de deriva se utiliza para modelar tendencias deterministas mientras que la función de volatilidad se utiliza a menudo para modelar un conjunto de eventos impredecibles que ocurren durante este movimiento (Wang and Wang, 2016).

Así mismo, un movimiento browniano estándar escalar o proceso Wiener estándar, sobre  $[0, T]$  es un proceso estocástico a tiempo continuo que cumple las siguientes tres condiciones:

1.  $W(0) = 0$  con probabilidad 1.
2. Para  $0 < s < t < T$ , la variable aleatoria dada por el incremento  $W(t) - W(s)$  se

distribuye normal con media cero y varianza  $t - s$ ; de manera equivalente,

$$W(t) - W(s) \sim \sqrt{t - s}N(0, 1)$$

donde  $N(0, 1)$  denota una distribución aleatoria normal con media cero y varianza unitaria.

3. Para  $0 < s < t < u < v < T$  los incrementos  $W(t) - W(s)$  y  $W(v) - W(u)$  son independientes.

Para propósitos computacionales, es útil considerar el movimiento browniano discreto, donde  $W(t)$  se especifica en valores discretos de  $t$ . Por lo tanto, establecemos  $\delta t = \frac{T}{N}$  para algún entero positivo  $N$  y denotamos a  $W_j$  como  $W(t_j)$  donde  $t_j = j\delta t$ . La condición 1 establece que  $W_0 = 0$  con probabilidad 1, y las condiciones 2 y 3 nos dicen que

$$W_j = W_{j-1} + dW_j, j = 1, 2, \dots, N,$$

donde cada  $dW_j$  es una variable aleatoria independiente de la forma  $\sqrt{\delta t}N(0, 1)$  (Higham, 2001).

Otro rasgo importante en función estocástica efectiva del tiempo son las integrales estocásticas que forman parte de la función. En una integral estocástica los integrandos y los integradores son procesos estocásticos como en la siguiente integral

$$\int_{t_0}^{t_n} \sigma(t) dW(t).$$

El método para calcular numéricamente una integral estocástica es análogo al proceso de calcular una integral usando sumas de Riemann. Dada una función adecuada  $h$ , la integral  $\int_0^T h(t) dt$  puede ser aproximado por la suma de Riemann

$$\sum_{j=0}^N h(t_j)(t_{j+1} - t_j), \quad (2.25)$$

donde  $t_j = j\delta t$ . De manera similar, podemos considerar una suma de la forma

$$\sum_{j=0}^N h(t_j) \left( W(t_{j+1}) - W(t_j) \right), \quad (2.26)$$

que, por analogía con 2.25, puede considerarse como una aproximación a una integral estocástica de la forma  $\int_0^T h(t) dW(t)$ . Aquí, estamos integrando  $h$  con respecto al movimiento browniano (Higham, 2001).

## 2.5 Método de remuestreo

Uno de los aspectos más importantes en la construcción y desarrollo de las redes neuronales artificiales es la capacidad de la red para generalizar a partir de ejemplos, evitando la simple memorización de patrones de aprendizaje. De manera que, interesa más una buena generalización de la red que la consecución de un error muy reducido en la submuestra de entrenamiento, ya que ello será indicativo de que el sistema ha capturado correctamente las relaciones subyacentes de los datos (López et al., 2008).

Al ajuste excesivo de la red por los patrones empleados en el entrenamiento se le conoce como sobreaprendizaje. El sobreaprendizaje puede evitarse utilizando procesos de validación cruzada (López et al., 2008). La validación cruzada es un método de remuestreo para ampliar nuestros recursos al momento del proceso del entrenamiento. Los métodos de remuestreo implican extraer repetidamente muestras de un conjunto de entrenamiento y reajustar un modelo de interés en cada muestra para obtener información adicional sobre el modelo ajustado (James et al., 2013).

Dentro de la validación cruzada se presenta el enfoque del conjunto de validación el cual es una estrategia muy simple para estimar el error de prueba asociado con el ajuste de un método de aprendizaje estadístico particular en un conjunto de muestras. Este método implica dividir aleatoriamente el conjunto de muestras disponible en dos partes, un conjunto de entrenamiento y un conjunto de validación o conjunto de reserva. El modelo se ajusta al conjunto de entrenamiento y el modelo ajustado se utiliza para predecir las respuestas de las observaciones en el conjunto de validación (James et al., 2013).

Una alternativa al enfoque anterior es la validación cruzada de  $k$  grupos. Este enfoque implica aleatoriamente dividir el conjunto de observaciones en  $k$  grupos, de aproximadamente igual tamaño. El primer grupo se trata como un conjunto de validación y el método se ajusta en los  $k - 1$  grupos restantes. Este procedimiento es repetido  $k$  veces; cada vez, se trata un grupo diferente de observaciones como un conjunto de validación (James et al., 2013).

Por otra parte, el algoritmo bootstrap es un método de remuestreo para estimar errores estándar y calcular intervalos de confianza. A menudo este método se usa cuando no se conoce la distribución de la población objetivo, de modo que la muestra es la única información disponible. Si  $T_n = g(X_1, \dots, X_n)$  es un estadístico de interés entonces para realizar la estimación de la varianza de  $T_n$  aplicando Bootstrap se deben seguir los siguientes pasos:

1. Extraer una muestra  $X_1^*, \dots, X_n^* \sim \hat{F}_n$ .
2. Calcular  $T_n^* = g(X_1^*, \dots, X_n^*)$ .
3. Repetir los pasos 1 y 2,  $B$  veces, para obtener  $T_{n,1}^*, \dots, T_{n,B}^*$ .
4. Definimos  $V_F(T_n) = \frac{1}{B-1} \sum_{b=1}^B \left[ T_{n,b}^* - \frac{1}{B} \sum_{r=1}^B T_{n,r}^* \right]^2$ .

Observe que la función  $\hat{F}_n$  pone masa de  $\frac{1}{n}$  en cada punto de datos  $X_1, \dots, X_n$ . Por lo tanto, extraer una observación de  $\hat{F}_n$  equivale a extraer un punto al azar del conjunto de datos original (Wasserman, 2004).

## 2.6 Prueba F de Fisher

La distribución F de Fisher-Snedecor con  $m$  y  $n$  grados de libertad donde  $m, n \in \mathbb{Z}$  denotada  $F_{m,n}$  es una distribución de probabilidad que se define como

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1 + \frac{mx}{n}\right)^{\frac{m+n}{2}}}$$

para  $x > 0$ . Para cada combinación de estos grados de libertad hay una distribución F diferente.

En la estadística, la distribución F tiene diversas aplicaciones entre las cuales se encuentra el análisis de varianzas. Una forma de realizar el análisis de varianza es mediante la prueba F que se usa con mayor frecuencia cuando se comparan modelos estadísticos que se han ajustado a un conjunto de datos, para identificar el modelo que mejor se adapta a la población de la que se tomaron muestras de los datos. La mayoría de las pruebas F surgen al considerar una descomposición de la variabilidad en una colección de datos en términos de sumas de los cuadrados de los errores (Michalowicz et al., 1999).

Las hipótesis que se contrastaron en la prueba F son las siguientes:

- $H_0 : \sigma_1^2 = \sigma_2^2$
- $H_A : \sigma_1^2 > \sigma_2^2$

donde  $\sigma_1^2$ ,  $\sigma_2^2$  son las varianzas de los dos modelos ajustados. El estadístico de la prueba F generalmente es un cociente de dos cantidades que se espera que sean aproximadamente iguales bajo la hipótesis nula.

El estadístico de la prueba F está dado por:

$$F = \frac{s_1^2}{s_2^2} \quad (2.27)$$

siendo  $s_1^2$  y  $s_2^2$  las varianzas muestrales de los modelos con  $s_1 > s_2$ . Considerando que las poblaciones de donde provienen las varianzas siguen una distribución normal, el estadístico F sigue una distribución  $F_{m,n}$  con  $m$  y  $n$  grados de libertad donde los grados de libertad del numerador se calculan como el tamaño de la muestra del grupo con mayor varianza muestral menos uno y los grados de libertad del denominador que se calculan como el tamaño muestral del grupo con menor varianza muestral menos uno.

Para determinar qué tan consistentes son nuestros resultados con la hipótesis nula se compara el valor del estadístico F con el valor crítico de la distribución F, es decir, la hipótesis nula se rechaza a nivel  $\alpha$  si  $F > F_{m,n}(\alpha)$  donde  $F_{m,n}(\alpha)$  es el cuantil  $1 - \alpha$  de la distribución F. En el caso de no poder rechazar la hipótesis nula se considera que las dos varianzas son iguales (Kiernan, 2014).

## CAPÍTULO 3

### Metodología

#### 3.1 Extracción de datos

Los datos de estudio de este trabajo se obtuvieron del sitio web del Sistema Nacional de Información e Integración de Mercados en México (SNIIM) el cual es un servicio de la Secretaría de Economía que tiene el propósito de ofrecer información sobre el comportamiento de los precios al por mayor de productos agrícolas, pecuarios y pesqueros que se comercializan en los mercados nacionales e internacionales.

La metodología empleada por el SNIIM para el levantamiento y la captura de los datos se basa en realizar una encuesta a los vendedores de los productos. Al terminar de hacer las encuestas a los vendedores los encuestadores se reúnen en el salón de cómputo para vaciar en la computadora la información obtenida.

Para la generación de la base de datos se aplicó la técnica de web scraping empleando como caso de estudio a las frutas y hortalizas que tuvieron como destino el mercado de abasto de Querétaro en el periodo de tiempo del 1 de enero de 2009 al 28 de febrero del 2019. En el apéndice K se muestra el algoritmo empleado para la generación de la base de datos.

Para efectuar el *web scraping* se usó el paquete *rves* disponible en el software estadístico R. El paquete *rves* es una subrutina que facilita la descarga y la manipulación de archivos HTML y XML (Wickham and Wickham, 2016). En este paquete, se usó de la instrucción *read\_html()* para separar los componentes de la página HTML, luego se usó *read\_node()* para seleccionar la tabla con los datos deseados y por último se usó *html\_table()* para convertir la tabla HTML previamente seleccionada en una estructura de datos de dos dimensiones que puede ser manipulada por el software R.

Para gestionar la base de datos, en primer lugar debido a que los productos no tiene una única presentación se decidió calcular el precio por kilogramo (para el apio, la caña, la col

mediana y grande, la coliflor mediana y grande, el elote, el elote grande y la lechuga romanita grande se calculó el precio por pieza mientras que para el epazote y la espinaca se usó el precio por manojo) de cada registro usando su precio frecuente para que el análisis posterior fuera coherente. Después, se dividió el origen de los productos en dos tipos: origen nacional o importado. Además, es necesario recalcar que todos los registros de la base de datos tienen como atributo ser de primera calidad.

Por último, para obtener registros de precios espaciados a intervalos iguales se utilizó el precio por kilogramo de los registros en donde el origen es de tipo nacional (en el caso del kiwi, la fresa, la pera D´anjou #100, la manzana red delicious, la manzana golden delicious, la uva superior, la uva globo y la uva Thompson se usaron los registros de origen internacional) para calcular el precio semanal promedio de cada una de las semanas entre el año 2009 y el año 2019 de cada producto de la base de datos.

Una vez que los precios semanales promedio de los productos fueron calculados se realizó un análisis de los datos para empezar a comprender la información obtenida utilizando la estadística descriptiva. En el proceso se buscan características significativas del conjunto de datos a estudiar como el precio máximo y mínimo que alcanzó cada producto en el periodo de observación, así como también nos interesa conocer las medidas de tendencia central de los datos como la media o calcular medidas de dispersión.

Prosiguiendo con el análisis exploratorio, se graficaron las series de tiempo de los precios semanales promedio de cada producto. Como se observa en la Figura 3.1, construir el gráfico es de gran utilidad para conocer el comportamiento global de la serie, observar la volatilidad de los precios a lo largo del tiempo y reconocer los lapsos de tiempo en donde hay falta de información.

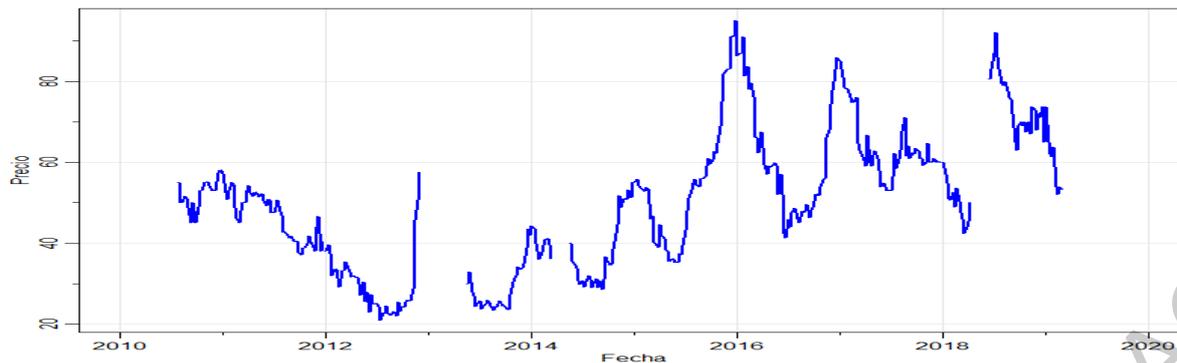


Figura 3.1: Series de precios de ajo morado

### 3.2 Elección de productos

Con respecto a los datos, se descubrió que algunos productos presentaron una ausencia de información significativa por lo cual fue necesario filtrar los datos para determinar cuál de estos productos tenían suficientes datos para ser estudiados. En ese sentido, se ha decidido que sólo 18 productos serán usados para el ajuste de los modelos. En la Tabla 3.1 se presentan los tres tipos de problemas observados en los datos, además se presentan los detalles de algunos de los productos dentro de cada problemática.

Problema	Producto	Núm. de registros	Primer registro	Último registro
Falta de información	Acelga	1	Semana 16-2019	Semana 16-2019
Producto discontinuado	Ajo blanco	106	Semana 1-2009	Semana 31-2010
Producto nuevo	Champiñon	159	Semana 22-2017	Semana 9-2019
Falta de información	Uva blanca	5	Semana 27-2012	Semana 31-2011

Tabla 3.1: Problemas de los datos

Los 18 productos que seleccionados fueron elegidos viables para aplicar la metodología Box-Jenkins por Paredes-Garcia et al. (2019), basándose en que los productos no exhibieran ambigüedad en su presentación, los precios faltantes en los últimos cuatro años no exceden el 30% y la familia a la cual pertenecen tiene sólo un miembro en la base de datos o contienen al menos dos miembros pero los datos no se superponen en el tiempo. También se tomó en consideración que al menos por tres semanas se tenga una gran oportunidad de

comercialización, es decir, el precio durante esa semana fue mayor que el cuartil del 75 % de los precios semanales durante al menos siete años.

Sin embargo, el número de observaciones difiere entre los productos seleccionados como viables por lo cual se seleccionó el periodo de tiempo que abarca de la primer semana del 2015 hasta la semana 8 del 2019 en donde la cantidad de registros es similar para los 18 productos. Además, se definieron los valores de los precios faltantes como el promedio de los precios de 3 semanas anteriores y posteriores. De esta manera se logró obtener 18 series de tiempo semanales sobre el mismo periodo de tiempo sin ningún dato faltante las cuales conformaron el conjunto de datos que se utilizó en la construcción de los modelos ARIMA usando la metodología Box-Jenkins y para entrenar y validar la red neuronal recurrente ELMAN.

### 3.3 Análisis exploratorio de los datos

Las series de tiempo de los datos a estudiar presentan características indeseadas para el análisis. Como primer paso en el análisis exploratorio de los datos se usó el método de regresión lineal para ver el comportamiento global de los precios y estimar la tendencia de la serie. La Figura 3.2 muestra que entre las características de las series de tiempo de los precios semanales promedio sin datos faltantes se observa una media variable en el tiempo, así como incrementos o decrementos abruptos en distintos lapsos del tiempo y una variabilidad considerable dependiente del tiempo.

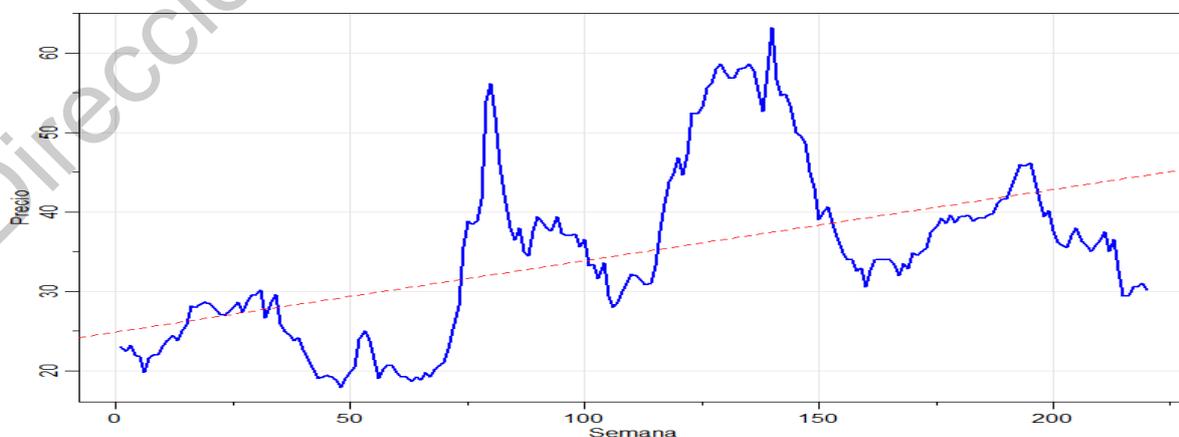


Figura 3.2: Tendencia lineal de la serie de aguacate Hass

Mediante el enfoque de series de tiempo, la relación de una variable con sus valores pasados es un factor importante para la construcción del modelo. No obstante, es difícil poder estimar autocorrelaciones con precisión si la estructura de dependencia no es regular o cambia en cada punto de tiempo. Por lo tanto, para lograr cualquier análisis estadístico significativo de series de tiempo datos, será crucial que, al menos, la media y la autocovarianza las funciones satisfacen las condiciones de estacionariedad (Shumway and Stoffer, 2017). Sin embargo, la volatilidad presente en los precios de las frutas y hortalizas ocasiona que las series de tiempo de estudio no satisfagan las condiciones de estacionariedad. Por esta razón, la última etapa del análisis estadístico exploratorio se centró en transformar las series de estudio en series estacionarias.

A causa de que los datos de las frutas y hortalizas presentan grandes fluctuaciones en distintos lapsos, a la serie de tiempo de estudio se les aplicó la transformación

$$y_t = \log(x_t)$$

para igualar la variabilidad a lo largo de la longitud de la serie. Después, dado que se desea tener una serie sin tendencia, a cada una de las series de tiempo se les aplicó el proceso de diferenciación definido por

$$z_t := \nabla y_t = y_t - y_{t-1}.$$

La diferenciación de la serie es una operación muy importante en las series de tiempo debido a que busca eliminar la tendencia tomando diferencias de términos sucesivos. La primera diferencia puede eliminar una tendencia lineal, mientras que una segunda diferencia, puede eliminar una tendencia cuadrática (Shumway and Stoffer, 2017). Este proceso de diferenciar el logaritmo de la serie define la serie de tiempo de la proporción de cambio del precio semanal promedio del producto la cual tiene con un comportamiento menos errático y sin tendencia como se puede observar en la Figura 3.3.

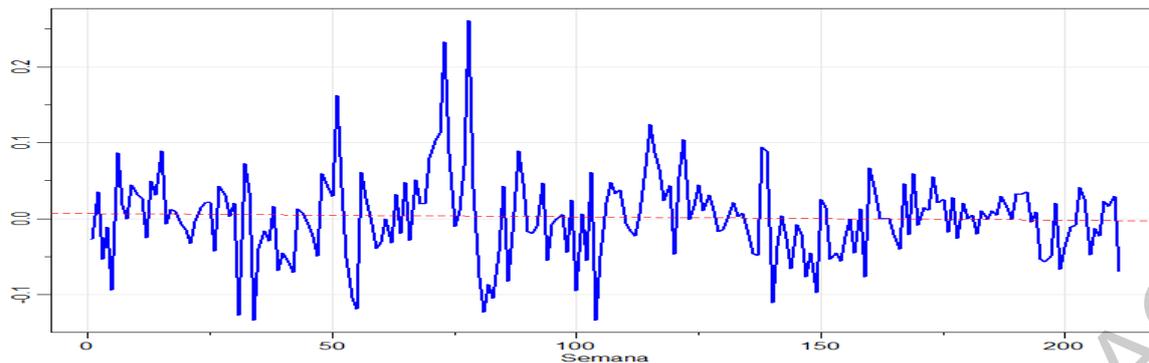


Figura 3.3: Primera diferencia de la serie de aguacate Hass

### 3.4 Estimación de modelos

#### 3.4.1 Modelos ARIMA

Para construir el modelo ARIMA de la serie de tiempo de la proporción de cambio del precio semanal promedio construida a partir de los precios semanales de las frutas y verduras se empleó la metodología Box-Jenkins haciendo uso del paquete *astsa* disponible en el software estadístico R que contiene un conjunto de instrucciones útiles para el estudio de series de tiempo (Stoffer and Stoffer, 2020). Los datos del 1 de enero de 2009 al 31 de diciembre de 2018 se utilizaron para ajustar el modelo y los datos restantes se utilizaron como conjunto de datos de prueba.

En primer lugar, para cada una de las series estacionarias se procedió a identificar un modelo que se ajuste a los datos calculando el orden del modelo autorregresivo y del modelo de media móvil por medio de la función de autocorrelación parcial (**PACF**) y la función de autocorrelación (**ACF**). La ACF Y PACF permiten identificar el orden del modelo autorregresivo y del modelo de media móvil de los datos ya que la ACF de un modelo  $MA(q)$  es cero para valores mayores que  $q$  mientras que la PACF de un modelo  $AR(p)$  es cero para valores mayores que  $p$  (Shumway and Stoffer, 2017). Por lo tanto, se graficó la función de autocorrelación y de autocorrelación parcial de la muestra para después analizar a la función teórica de que modelo se asemeja observando a partir de que rezago los valores de las funciones son cero. Para graficar la función de autocorrelación de los datos se usó la instrucción `acf2()` que también realiza la gráfica de la función de autocorrelación parcial.

En segundo lugar se realizó la estimación de los parámetros del modelo ARIMA usando el método de estimación de máxima verosimilitud (MLE), que determina los valores de los parámetros que son más probables que hayan producido las observaciones (Shumway and Stoffer, 2017). En esta etapa se suele recurrir a paquetes informáticos que realicen el cálculo de estimación, en este trabajo se utilizó la instrucción *sarima()* del software R para estimar los parámetros del modelo.

En último lugar realizamos la validación del modelo con el objetivo de descubrir una posible falta de ajuste y decidir si el modelo es estadísticamente adecuado a través del análisis de los residuales. Los residuales de un modelo ARIMA definidos como  $e_t = y_t - \hat{y}_t$  son una aproximación del componente aleatorio presente en una serie de tiempo, por lo que si los residuales están correlacionados entre sí, estos no son ruido blanco y significa que hay un patrón que no ha sido tomado en cuenta por el modelo ARIMA propuesto (Shumway and Stoffer, 2017).

Se establece que un modelo se ajusta bien a los datos, si los residuales estandarizados se comportan como una secuencia independiente e idénticamente distribuida con media cero y varianza uno, por lo cual se grafica la serie tiempo de los errores para detectar cualquier desviación obvia de esta suposición. Para probar la aleatoriedad de los residuos se grafica  $\hat{\rho}_e(h)$  contra  $h$  en busca de patrones o valores grandes y se grafica el estadístico Q de Ljung-Box Pierce dado por

$$Q = n(n+2) \sum_{h=1}^H \frac{\hat{\rho}_e^2(h)}{n-h}$$

el cual considera las magnitudes de  $\hat{\rho}_e(h)$  como grupo (Shumway and Stoffer, 2017).

A menos que la serie de tiempo sea gaussiana, no es suficiente que los residuos no estén correlacionados. En el caso no gaussiano, es posible tener un proceso no correlacionado para el cual los valores contiguos son altamente dependientes. En la investigación de la normalidad marginal, una gráfica de probabilidad normal o una gráfica Q-Q puede ayudar a identificar desviaciones de la normalidad (Shumway and Stoffer, 2017). Para validar el modelo propuesto con los parámetros estimados no se necesitó de ninguna instrucción extra debido a que la instrucción *sarima()* realiza el análisis de los residuos obtenidos después de estimar los parámetros, graficando la serie de tiempo de los residuos, la función de autocorrelación,

la gráfica Q-Q y el estadístico Q de Ljung-Box Pierce.

A fin de concluir el análisis con el modelo ARIMA se realizó la predicción de los valores futuros de la serie de tiempo. Una vez se validó que el modelo se ajusta bien a la serie de tiempo, se utilizó la instrucción *sarima.for()* para estimar el valor de los precios de las frutas y hortalizas de las próximas 8 semanas.

Después de haber realizado la predicción, se analizó el desempeño del modelo ARIMA comparando la estimación con los valores reales de la serie de tiempo. Los valores reales se encuentran en el conjunto de prueba previamente seleccionado que contiene los precios semanales promedio de la semana 1 a la 8 del año 2019. Para analizar el desempeño de la predicción se utilizan los siguientes criterios de evaluación de los errores: el error absoluto medio (**MAE**), la raíz del error cuadrático medio (**RMSE**) y el error porcentual absoluto medio (**MAPE**), definidos como

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t|,$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2},$$

$$MAPE = 100 \times \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right|,$$

donde  $y_t$  es el valor real,  $\hat{y}_t$  es el valor de la predicción en el tiempo  $t$  y  $N$  es el número total de datos a comparar (Wang and Wang, 2016).

### 3.4.2 Red neuronal recurrente ELMAN

A continuación, se procedió con el estudio construyendo una red neuronal recurrente ELMAN (**ERNN**) para predecir los valores de la serie de tiempo por medio de una red neuronal. Para construir, entrenar y validar la red neuronal artificial se desarrolló un algoritmo en el lenguaje de programación R. Este algoritmo puede encontrarse en el apéndice L el cual empleo como datos de entrada la serie de tiempo semanal de los precios frecuentes sin valores faltantes de cada uno de los 18 productos seleccionados, es decir, los datos producidos en el

apéndice K.

En este trabajo, se utilizó una **ERNN** con la siguiente arquitectura:  $n$  neuronas en la capa de entrada,  $2(n+1)$  neuronas en la capa oculta,  $2(n+1)$  neuronas en la capa recurrente y una neurona en la capa de salida. En esta red se estableció como  $x_i$  la salida de la neurona  $i$  de la capa de entrada,  $z_j$  la salida de neurona  $j$  de la capa oculta,  $u_j$  la salida de la neurona  $j$  de la capa recurrente y  $y_1$  la salida de la capa oculta. También, se definió como  $w_{ij}$  el peso de la conexión de la neurona  $i$  de la capa de entrada con la neurona  $j$  de la capa oculta,  $c_j$  de la neurona  $j$  de la capa recurrente con cada neurona de la capa oculta,  $v_j$  de la neurona  $j$  de la capa oculta a la neurona de la capa de salida y el peso de la conexión de la neurona  $j$  de la capa oculta con la neurona  $j$  de la capa recurrente es igual a 1 durante todo el proceso de aprendizaje. Además, en esta **ERNN** para la capa oculta se establece a la suma como función de entrada y a la función sigmoide como función de activación mientras que en la capa de salida se establece a la suma como función de entrada y a la función identidad como función de activación.

Como resultado de la arquitectura anterior, el valor de  $u_j$  se define como

$$u_j(k) = z_j(k-1)$$

donde  $k$  es el número de iteración, mientras que el valor de salida de la neurona  $j$  en la capa oculta en la  $k$ -ésima iteración dado por

$$z_j(k) = \frac{1}{1 + e^{-\left(\sum_{i=1}^n (w_{ij}x_i(k)) + \sum_{j=1}^{2(n+1)} (c_j u_j(k))\right)}}$$

y el valor de salida la capa de salida en la  $k$ -ésima iteración dado por

$$y_1 = \sum_{j=1}^{2(n+1)} v_j z_j(k).$$

Antes de comenzar el proceso de entrenamiento se dividió la serie de tiempo de cada producto en tres conjuntos de datos: el conjunto de entrenamiento, el conjunto de validación y el conjunto de prueba. Los datos del 1 de enero de 2015 al 31 de diciembre de 2017 se

utilizaron para entrenar la red neuronal, mientras que para la validación se usaron los datos del 1 de enero de 2018 al 31 de diciembre de 2018 y los datos del 1 de enero de 2019 al 28 de febrero de 2019 conformaron el conjunto de prueba de la red.

Los datos de entrada de una red neuronal artificial no necesitan sufrir alguna modificación o transformación, sin embargo, para reducir el impacto del ruido en los datos y finalmente conducir a una mejor predicción, los datos recopilados deben ajustarse y normalizarse adecuadamente al comienzo del modelado (Wang and Wang, 2016). Para este trabajo los datos se normalizaron siguiendo la siguiente formula

$$S'(t) = \frac{S(t) - \min S(t)}{\max S(t) - \min S(t)}$$

donde  $S(t)$  es el valor de la serie de tiempo en el tiempo  $t$  y los valores mínimo y máximo se obtienen del conjunto de entrenamiento durante el proceso de entrenamiento.

Después de normalizar los datos se realizó el proceso de entrenamiento de la red. Para empezar a entrenar la red, en primer lugar se establecieron ciertos parámetros como la tasa de aprendizaje, el número de iteraciones y el número de entradas del modelo. Además, se inicializaron los pesos conectivos  $w_{ij}$ ,  $b_j$  y  $c_j$  simulando valores de una distribución uniforme de  $(-1, 1)$ .

Mas aún, se establecieron los componentes de la función estocástica efectiva del tiempo de la ecuación 2.20, definiendo

$$\mu(t) = \frac{1}{(N - t)^2}$$

como la función de deriva y

$$\sigma(t) = \left[ \frac{1}{N - 1} \sum_{i=1}^N (x_i - \bar{x})^2 \right]^{\frac{1}{2}}$$

como la función de volatilidad en donde  $N$  es número de datos del conjunto de entrenamiento, también, se simulo un movimiento Browniano para estimar el valor de la integral y se definió el coeficiente de resistencia del tiempo como  $\beta = 1$ .

Una vez que se definieron los parámetros del proceso de entrenamiento, se inició con el algoritmo de propagación hacia atrás para entrenar la red. En primer lugar, se emplearon los precios promedio semanales de las primeras  $n$  semanas (la primer muestra) para predecir el precio de la semana  $n + 1$ . Luego, se calculó la salida de la red y la función de pérdida de esa muestra.

En segundo lugar, se calcularon las derivadas parciales de la función de pérdida con respecto a los pesos de la red por medio del algoritmo de propagación hacia atrás. Después, se ingresaron los datos de la segunda muestra, la cual está conformada por los precios promedio semanales de la semana 2 a la semana  $n+1$  y se repitió el proceso de ingresar las entradas, calcular la salida, calcular la función de pérdida y sus derivadas respecto a los pesos. A continuación se aplicó el procedimiento en la siguiente muestra y así sucesivamente hasta ingresar todo el conjunto de entrenamiento una vez en la red.

En último lugar, se obtuvo el error global de la red de esta iteración y se calculó el error global del conjunto de validación con los pesos actuales de la red. A continuación, se aplicó el algoritmo de propagación hacia atrás junto al algoritmo del gradiente descendiente para modificar los pesos de la red. Después, una vez más se ingresaron los datos del conjunto de entrenamiento a la red repitiendo el proceso anteriormente descrito con los nuevos pesos calculados. Esto se realizó una y otra vez empleando el algoritmo de *early stopping* para evitar un sobreajuste de la red.

Prosiguiendo el estudio, posteriormente de haber entrenado exitosamente la red neuronal recurrente Elman se procedió a realizar la evaluación de la red con nuestro conjunto de prueba. En el proceso de evaluación, la red se entrenó con los datos de entrenamiento (conjunto de entrenamiento y validación) el mismo número de veces que en la etapa anterior.

Al terminar este proceso iterativo, se ingresaron los valores de las últimas  $n$  semanas del 2018 y usando los pesos conectivos establecidos por el proceso de entrenamiento se calculó el precio semanal promedio de la primer semana del 2019. Después, se ingresaron los valores de las últimas  $n - 1$  semanas del 2018 y el precio estimado de la primer semana del 2019 para calcular el precio de la segunda semana del 2019. Este proceso se repitió hasta ingresar el valor de la última semana del 2018 y los valores estimados de las primeras siete semanas

del 2019 para calcular el valor de la octava semana del 2019.

Para analizar el desempeño de la predicción, al igual que para las predicciones de los modelos ARIMA se utilizan los siguientes criterios de evaluación: el error absoluto medio, la raíz del error cuadrático medio y el error porcentual absoluto medio.

### **3.5 Comparación de modelos**

Para finalizar este trabajo se comparó el error cuadrado medio de la red neuronal recurrente de EIMAN con el error cuadrado medio del modelo ARMA. La primera etapa de la comparación estuvo constituida en analizar la normalidad de los errores de las predicciones de cada uno de los productos. Después se utilizó la prueba F sobre el cociente de los MSE para analizar el supuesto de homogeneidad de varianzas. Por último, se interpretaron los resultados de la prueba F para determinar cuál de los dos modelos realiza una mejor predicción de los precios de las frutas y hortalizas estudiadas.

## CAPÍTULO 4

### Resultados

#### 4.1 Extracción de la base de datos

La información extraída del sitio web del SNIIM contiene los datos de 91 frutas y hortalizas comercializados en el mercado de abastos de Querétaro, entre las cuales encontramos la acelga, el aguacate Hass, el champiñón, el chile puya seco, el chile serrano, la jícama, el limón con semilla #5, el mango Ataulfo, el tomate saladette, la naranja Valencia grande, la uva Thompson y otros más. La lista completa de los 91 productos que conformaron el estudio se encuentra en el apéndice A.

La información extraída se encuentra distribuida en 137244 registros con los siguientes atributos:

- Fecha: Día, mes y año de cuando se realizó el estudio.
- Producto: Nombre de la fruta u hortaliza observada.
- Calidad: Describe si la calidad es de primera o de exportación.
- Presentación: Presentación del producto, por ejemplo, por kilogramo o por pieza.
- Origen: Lugar de donde proviene el producto.
- Precio mínimo: El valor más bajo de la cotización dentro de una muestra.
- Precio máximo: El valor más alto de la cotización dentro de una muestra.
- Precio frecuente: Es el dato que más se repite en la muestra.

## 4.2 Análisis de datos

Para realizar el primer análisis de los datos se calculó el número de registros que corresponden a los precios de cada producto así como el precio máximo y el precio mínimo registrado a lo largo del tiempo. La Tabla 4.1 muestra los resultados de algunos productos presentes en la base de datos y en el apéndice B se muestra el resultado del análisis de cada uno de los 91 productos.

Nombre	Núm. datos	Precio mínimo	Precio máximo
Acelga	1	3	3
Aguacate Hass	583	13	63.2
Champiñón	159	47	58
Chile serrano	583	3.73	83.4
Chile puya seco	477	40.8	130
Jícama	583	2.5	13.4
Limón c/semilla #5	424	3.05	28.4
Mango Ataulfo	390	6.36	33.33
Naranja Valencia grande	424	2.05	13.8
Tomate saladette	583	3.65	28.53
Uva blanca	5	18.98	20.91

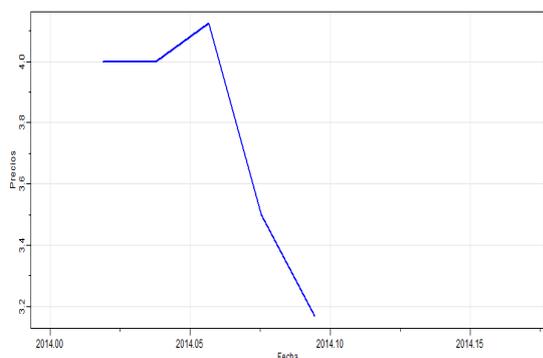
*Tabla 4.1: Descripción de datos*

En la Tabla 4.1 se puede observar que no todos los productos tiene el mismo número de datos e incluso hay productos que tiene una cantidad de datos insignificantes como la acelga que sólo tiene un registro en la base de datos.

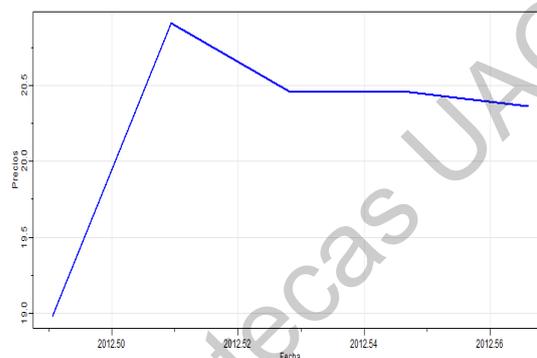
## 4.3 Gráficas de las series

La gráfica de la series de la acelga, la cereza, la col grande, la granada roja, la mandarina Mónica, el melón cantaloupe #12, el plátano Chiapas, el rábano, el tamarindo,

la toronja roja mediana y la uva blanca no se analizaron debido a que se tenía muy poca información de estos productos como se muestra en la Figura 4.1. La granada roja y la toronja roja mediana son los productos con más información de los antes mencionados con sólo los precios de 10 semanas.



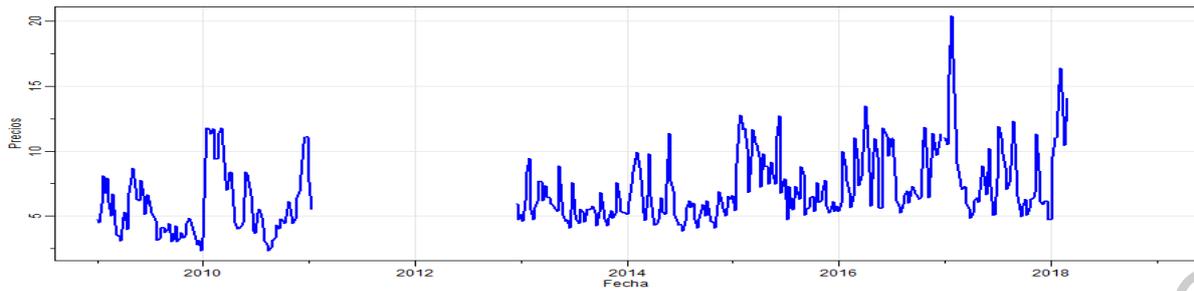
(a) Serie de toronja roja mediana



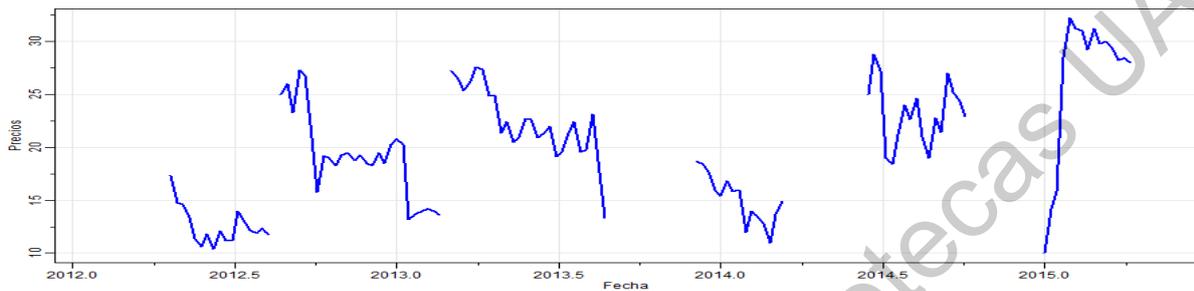
(b) Serie de uva blanca

Figura 4.1: Series con pocos precios semanales

Por otra parte, en la serie de tiempo de algunos productos como se muestra en la Figura 4.2a se presenta una discontinuidad que proviene de un periodo de ausencia de información de productos como el chícharo, el chayote sin espinas, la cebolla bola y el chile chilaca. Otros productos presentaron más de un periodo de pérdida de información a lo largo de tiempo como la ciruela roja, el chile árbol seco, el mango manila, la tuna blanca y la uva superior como se muestra en la Figura 4.2b.



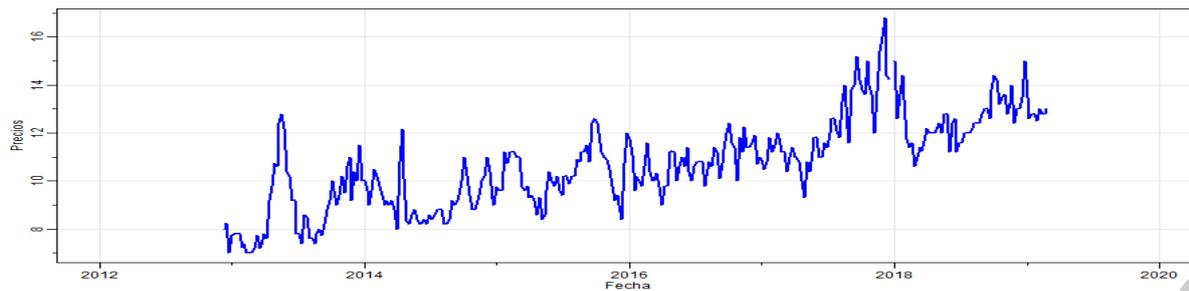
(a) Serie de chayote sin espinas



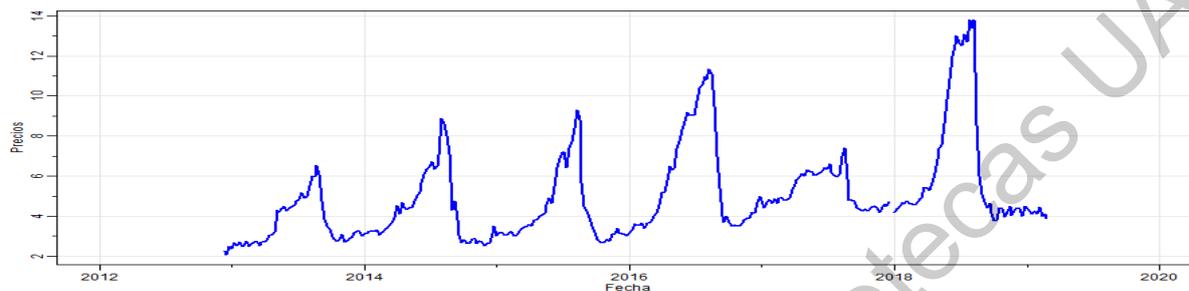
(b) Serie de ciruela roja

Figura 4.2: Series con periodos sin información

A su vez, al graficar las series de tiempo se observó que algunos productos tiene un comportamiento similar. Así, por ejemplo, la serie de tiempo del cacahuate y la coliflor grande presentan una tendencia lineal. En la Figura 4.3a se puede observar que la serie presenta una tendencia ascendente y que varía en una banda respecto a una línea recta. Mientras que otro grupo de productos entre los cuales se encuentra la jícama y la naranja Valencia grande muestran evidencia de tener una serie de tiempo estacional. En la Figura 4.3b se puede observar un patrón que se repite año tras año con pequeños cambios.



(a) Serie de coliflor grande



(b) Serie de naranja Valencia grande

Figura 4.3: Series con determinado comportamiento

En apéndice C se muestran las gráficas de la serie de tiempo semanales de las frutas y hortalizas de las cuales se tiene registros de por lo menos un periodo de tiempo mayor a medio año.

#### 4.4 Selección de frutas y hortalizas

En base en el estudio realizado por Paredes-García et al. (2019), los 18 productos mostrados en la Tabla 4.2 tienen ciertas características que hacen viable el estudio de sus series de tiempo.

Aguacate Hass	Epazote	Nopal grande
Cacahuete	Jícama	Piña mediana
Calabacita Italiana	Kiwi	Plátano macho
Chayote sin espinas	Lima	Plátano Tabasco
Chile Serrano	Limón con semilla #5	Tomate saladette
Coliflor grande	Naranja Valencia grande	Toronja roja

Tabla 4.2: Productos seleccionados viables

Para cada uno de estos 18 productos se obtuvo como resultado una serie de tiempo de los precios promedios semanales sin valores faltantes para el periodo de tiempo comprendido entre la primer semana del 2015 y la 8 semana del 2019 como se muestra en la Tabla 4.3 (los precios mostrados en la tabla son los precios reales redondeados). En el apéndice D se encuentran las tablas con los precios promedio semanales redondeados de los 18 productos.

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	23	22.4	23.2	22	21.75	19.8	21.6	22	22	23	23.75	24.4	23.8	25	25.8	28.2	28	28.33	28.6	28.4	28	27.1	27	27.4	28	28.6	27.4
2016	23.75	21.4	19	20.2	20.75	20.6	19.8	19.2	19.2	18.6	19.2	18.83	19.75	19.2	20.2	20.6	21	22.75	25.2	28.2	35.6	38.8	38.4	38.8	41.6	54	56.2
2017	28.6	30	31	32.2	32	31.5	30.8	31	33.2	37.6	41	43.75	44.8	46.8	44.67	47.2	52.4	52.33	53.2	55.6	56.2	58	58.6	57.6	56.8	56.8	58
2018	30.5	32.6	34	34	34	34	33.3	32	33.5	32.8	34.8	34.5	35	35.4	37.4	38.2	39.2	38.5	39.6	38.6	39.4	39.4	39.6	38.8	39.2	39.2	39.6
2019	36.5	32.6	29.4	29.5	30.5	30.6	31	30.2																			

Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53
2015	28.6	29.5	29.6	30.2	26.6	28.6	29.6	25.9	24.9	24.5	23.8	24.2	22.6	21.6	20.4	19	19.25	19.4	19.25	18.8	17.9	19	19.8	20.4	24	25
2016	52	46	42.2	38	36.4	38	35	34.4	37.6	39.4	38.75	38	37.6	39.4	37.3	37	37	37.2	35.6	36.5	33.2	33.4	31.6	33.6	29.4	28
2017	58.2	58.6	57.8	55.2	52.6	57.8	63.2	56.6	54.6	54.8	53.4	50	49.6	48.6	45	43	39	40	40.6	38.5	36.8	34.8	34	34	32.5	32.93
2018	39.8	41	41.6	41.6	43	44.4	46	45.8	46.2	43.8	41.4	39.4	40.2	37.6	36.2	35.8	35.5	37	38	36.25	35.8	35	35.8	36.4	37.5	35
2019																										

Tabla 4.3: Precios semanales promedio de aguacate Hass

## 4.5 Transformación de los datos

El método de regresión lineal evidenció que los productos presentan una tendencia global creciente pero en la mayoría de los casos esta tendencia no es lineal. Únicamente el cacahuate y la toronja roja presentan un buen ajuste de la tendencia a la serie de tiempo. En la Tabla 4.4 se muestran los coeficientes obtenidos por la regresión lineal ( $x_t = \beta_1 + \beta_2 t + \varepsilon$ ) junto a sus errores estándar, el estimador de la desviación estándar de los residuos obtenidos y el coeficiente de determinación ( $R^2$ ).

La Tabla 4.4 no muestra el valor del estadístico t de los parámetros debido a que en cada uno de los 18 productos cada coeficiente rechaza la hipótesis nula de que el coeficiente sea cero. Además, todos los productos a excepción del cacahuate reportaron un coeficiente de determinación muy bajo, por lo cual, el modelo de regresión lineal resulta ser un mal modelo para realizar predicciones de nuestros datos. En el apéndice E se muestra la línea estimada por la regresión lineal junto a la serie de tiempo de cada uno de los productos.

Producto	$\beta_1$	$\beta_2$	$\sigma_\varepsilon$	$R^2$
Aguacate Hass	24.86 <sub>(1.235)</sub>	0.089 <sub>(0.0096)</sub>	9.133	0.279
Cacahuate	26.30 <sub>(0.622)</sub>	0.177 <sub>(0.0048)</sub>	4.604	0.857
Calabacita italiana	7.94 <sub>(0.503)</sub>	0.014 <sub>(0.0039)</sub>	3.72	0.054
Chayote sin espinas	6.42 <sub>(0.352)</sub>	0.012 <sub>(0.0027)</sub>	2.609	0.075
Chile serrano	10.51 <sub>(1.29)</sub>	0.059 <sub>(0.0101)</sub>	9.6	0.129
Coliflor grande	9.69 <sub>(0.146)</sub>	0.016 <sub>(0.0011)</sub>	1.08	0.494
Epazote	49.3 <sub>(1.218)</sub>	0.042 <sub>(0.0095)</sub>	9.005	0.078
Jícama	5.07 <sub>(0.229)</sub>	0.011 <sub>(0.0018)</sub>	1.695	0.144
Kiwi	43.32 <sub>(0.967)</sub>	0.051 <sub>(0.0075)</sub>	7.151	0.171
Lima	8.52 <sub>(0.454)</sub>	0.026 <sub>(0.0035)</sub>	3.359	0.198
Limón c/semilla #5	7.11 <sub>(0.528)</sub>	0.017 <sub>(0.0041)</sub>	3.903	0.070
Naranja Valencia grande	4.47 <sub>(0.326)</sub>	0.009 <sub>(0.0025)</sub>	2.411	0.052
Nopal grande	9.91 <sub>(0.582)</sub>	0.025 <sub>(0.0045)</sub>	4.308	0.121
Piña mediana	8.44 <sub>(0.242)</sub>	0.006 <sub>(0.0019)</sub>	1.796	0.051
Plátano macho	9.84 <sub>(0.174)</sub>	0.016 <sub>(0.0013)</sub>	1.287	0.393
Plátano Tabasco	5.71 <sub>(0.3)</sub>	0.011 <sub>(0.0023)</sub>	2.22	0.099
Tomate saladette	10.96 <sub>(0.587)</sub>	0.014 <sub>(0.0046)</sub>	4.346	0.036
Toronja roja	4.39 <sub>(0.239)</sub>	0.032 <sub>(0.0018)</sub>	1.767	0.57

Tabla 4.4: Resultados de la regresión lineal

El proceso de transformación de las 18 series semanales de frutas y hortalizas nos dio como resultado las gráficas del apéndice F. Al calcular el logaritmo de los datos se logró tener una varianza más estable alrededor de la función media y con la primera diferencia de la serie se obtuvo una serie con un mejor comportamiento, con una tendencia lineal similar a una función media constante.

Sin embargo, aún se observan picos o cambios abruptos a lo largo de la serie. Las 18 series obtenidas al realizar la primera diferencia del logaritmo de la serie original mostraron evidencia estadística de ser estacionarias.

## 4.6 ARIMA

Empleando la metodología Box-Jenkins a las series de tiempo estacionarias se ajustaron diferentes modelos a los datos y con base en el criterio de información de Akaike se determinó el modelo a utilizar en el estudio obteniendo los modelos presentados en la Tabla 4.5.

El análisis de la función de autocorrelación parcial muestral y autocorrelación muestral de las series de tiempo nos dio como resultado que a las series de tiempo de la jícama, el limón con semilla #5 y la naranja Valencia grande se les debe ajustar un modelo  $SARIMA(p, d, q) \times (P, D, Q)_s$  mientras que el resto de los productos se les ajusto un modelo  $ARIMA(p, d, q)$ . En el apéndice G se agregaron las gráficas de la ACF y la PACF muestral de los 18 productos.

Producto	Modelo	Parámetros significativos	$\hat{\sigma}_w^2$
Aguacate Hass	$ARIMA(1, 0, 0)$	$\phi_1 = 0.2784_{(0.066)}$	0.002819
Cacahuate	$ARIMA(0, 0, 0)$		0.001323
Calabacita italiana	$ARIMA(0, 0, 2)$	$\theta_2 = -0.2372_{(0.0693)}$	0.04687
Chayote sin espinas	$ARIMA(0, 0, 2)$	$\theta_2 = -0.4472_{(0.0642)}$	0.04448
Chile serrano	$ARIMA(1, 0, 0)$	$\phi_1 = 0.2505_{(0.0687)}$	0.01774
Coliflor grande	$ARIMA(1, 0, 1)$	$\phi_1 = 0.6482_{(0.1688)}, \theta_1 = -0.8926_{(0.1188)}$	0.003978
Epazote	$ARIMA(1, 0, 1)$	$\phi_1 = 0.8477_{(0.0419)}, \theta_1 = -1_{(0.0315)}$	0.008644
Jícama	$ARIMA(0, 0, 0) \times (1, 0, 0)_{53}$	$\Phi_1 = 0.2049_{(0.0687)}$	0.008494
Kiwi	$ARIMA(9, 0, 0)$	$\phi_1 = 0.3540_{(0.0679)}, \phi_9 = -0.1697_{(0.068)}$	0.001649
Lima	$ARIMA(0, 0, 4)$	$\theta_4 = 0.2247_{(0.0741)}$	0.007272

Limón #5	$ARIMA(1, 0, 0) \times (1, 0, 0)_{53}$	$\phi_1 = 0.2243_{(0.0019)}, \Phi_1 = 0.2569_{(0.0019)}$	0.0159
Naranja Valencia grande	$ARIMA(1, 0, 0) \times (1, 0, 0)_{53}$	$\phi_1 = 0.3977_{(0.0663)}, \Phi_1 = 0.2583_{(0.0874)}$	0.00618
Nopal grande	$ARIMA(0, 0, 1)$	$\theta_1 = 0.2051_{(0.0737)}$	0.01516
Piña mediana	$ARIMA(1, 0, 0)$	$\phi_1 = 0.171_{(0.0678)}$	0.005027
Plátano macho	$ARIMA(0, 0, 1)$	$\theta_1 = -0.3811_{(0.0657)}$	0.003573
Plátano Tabasco	$ARIMA(1, 0, 0)$	$\phi_1 = 0.3542_{(0.0652)}$	0.007286
Tomate saladette	$ARIMA(1, 0, 2)$	$\phi_1 = 0.8295_{(0.0428)}, \theta_1 = -0.6194_{(0.0668)}, \theta_2 = -0.3806_{(0.0647)}$	0.02088
Toronja roja	$ARIMA(1, 0, 0)$	$\phi_1 = -0.1697_{(0.0705)}$	0.006974

Tabla 4.5: Modelos ajustados a los datos

Los modelos de la Tabla 4.5 presentaron evidencia estadística de ser adecuados a través del análisis de los residuales. La evaluación de cada uno de los modelos se presenta en el apéndice H. La predicción usando los modelos de la Tabla 4.5 produjo como resultado los valores de la Tabla 4.6 que representan los precios semanales promedio estimados de cada producto 8 semanas posteriores a la última semana usada para ajustar el modelo.

Producto	S1	S2	S3	S4	S5	S6	S7	S8
Aguacate Hass	34.42	34.26	34.24	34.27	34.32	34.38	34.44	34.50
Cacahuate	66.26	6.53	66.79	67.06	67.33	67.59	67.86	68.14
Calabacita italiana	12.02	12.05	12.07	12.10	12.12	12.15	12.17	12.20
Chayote sin espinas	4.74	4.73	4.73	4.73	4.73	4.73	4.72	4.72
Chile serrano	73.25	76.07	77.38	78.29	79.10	79.89	80.69	81.49
Coliflor grande	14.61	14.33	14.13	13.99	13.90	13.83	13.79	13.76
Epazote	83.78	78.91	75.06	72.00	69.53	67.55	65.93	64.62
Jícama	8.08	8.09	8.02	8.29	8.30	8.33	8.39	8.38
Kiwi	48.59	48.27	47.99	47.61	47.52	48.03	48.34	48.27
Lima	10.02	10.04	10.06	10.07	10.09	10.11	10.13	10.15
Limón c/semilla #5	7.62	7.20	7.39	7.44	7.52	7.45	7.74	8.00
Naranja Valencia grande	4.70	4.84	4.91	4.97	4.97	5.02	4.99	5.00
Nopal grande	22.02	22.04	22.06	22.08	22.10	22.12	22.14	22.16
Piña mediana	9.09	9.11	9.12	9.12	9.13	9.13	9.14	9.14
Plátano macho	15.03	15.06	15.09	15.12	15.15	15.18	15.22	15.25
Plátano Tabasco	6.00	6.00	6.01	6.02	6.03	6.04	6.04	6.05
Tomate saladette	18.50	18.50	18.54	18.58	18.62	18.66	18.70	18.74
Toronja roja	8.44	8.52	8.55	8.59	8.63	8.67	8.71	8.75

Tabla 4.6: Predicción de los modelos ARIMA

Una característica muy importante de los modelos ARIMA es la posibilidad de construir intervalos de predicción, los cuales generalmente se calculan junto con los pronósticos para evaluar la precisión del pronóstico. En general, los intervalos de predicción  $(1 - \alpha)$  tienen la forma

$$x_{n+m}^n \pm c_{\frac{\alpha}{2}} \sqrt{P_{n+m}^n}$$

donde  $x_{n+m}^n = E[x_{n+m}|x_1, \dots, x_n]$ ,  $P_{n+m}^n = E[x_{n+m} - x_{n+m}^n]^2$  y  $c_{\frac{\alpha}{2}}$  se selecciona para obtener el grado de confianza deseado (Shumway and Stoffer, 2017).

En la Figura 4.4 se graficaron los valores de predicción del aguacate Hass, así mismo se muestran dos intervalos de confianza para nuestras predicciones, en color verde claro se muestra un intervalo de la forma  $\hat{x}_t \pm \sqrt{P_t^{212}}$  con  $t = 213, \dots, 220$  mientras que en color verde fuerte se muestra un intervalo de la forma  $\hat{x}_t \pm z_{0.975} * \sqrt{P_t^{212}}$  con  $t = 213, \dots, 220$ . En el apéndice I se muestran las gráficas de las predicciones de los 18 productos junto a los dos intervalos de predicción anteriormente mencionados.

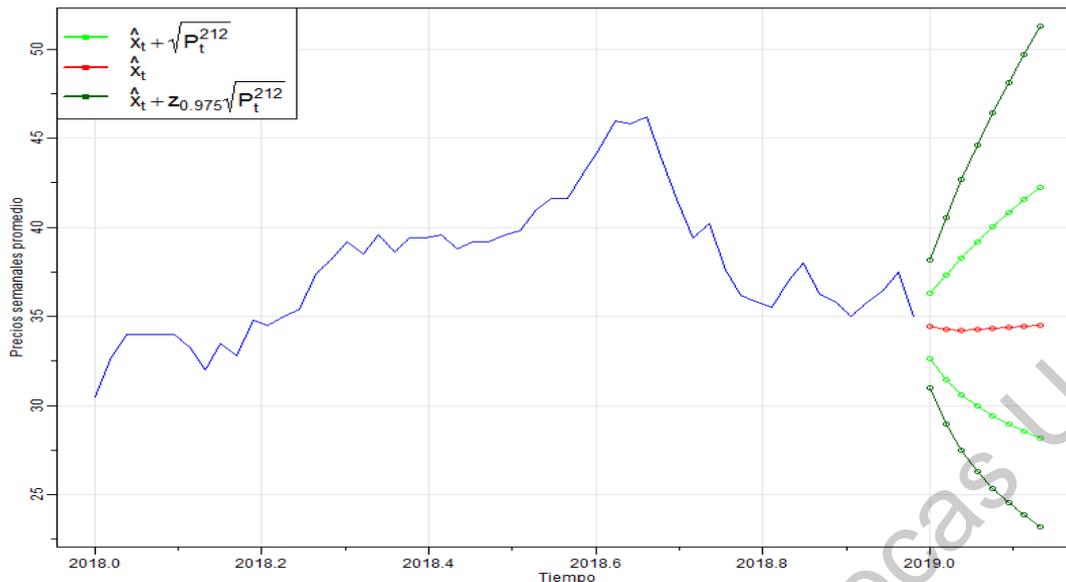


Figura 4.4: Intervalos de predicción para la serie de aguacate Hass

Los intervalos de predicción nos representan una estimación del rango en el cual se situarán los valores futuros con determinada probabilidad dado lo que ya se ha observado. De manera que intervalos de predicción de la Figura 4.4 nos permiten conocer que tanto pueden variar los precios semanales promedio de las primeras 8 semanas del 2019 de los productos con cierto grado de confianza. Una característica importante de los intervalos de predicción es que aumentan de longitud a medida que aumenta el tiempo del pronóstico por lo cual entre más alejado del presente se realice la predicción más incertidumbre se asociará con el pronóstico.

## 4.7 ERNN

Para la construcción y entrenamiento de la red neuronal recurrente, se necesitó especificar un conjunto de características para la aplicación del algoritmo. Sin embargo, para cada producto se asignaron distintas características como se muestra en la Tabla 4.7.

El número de neuronas en la capa de entrada se calculó con base en el número de valores significativos que se muestran en la función de autocorrelación parcial muestral de la serie de tiempo mientras que el valor de la tasa de aprendizaje se estableció en 0.5 para que el aprendizaje fuera más rápido. Por otra parte, se utilizaron el doble del número de neuronas de entrada más dos neuronas en la capa oculta y en la capa recurrente debido a que si el

número de neuronas en la capa de entrada es  $N$ , el número de neuronas en la capa oculta está previsto que sea cercano a  $2N + 1$  (Wang and Wang, 2016).

Además, el número de iteraciones se calculó por medio del algoritmo de *early stopping* con un parámetro de tolerancia igual a 20, este último parámetro sólo se empleó en el entrenamiento y el número de iteraciones sólo se empleó a la etapa de la evaluación.

Producto	Neuronas de entrada	Neuronas ocultas	Tasa de aprendizaje	Núm. de iteraciones	Tolerancia del error
Aguacate Hass	2	6	0.5	135	20
Cacahuete	1	4	0.5	500	20
Calabacita italiana	1	4	0.5	500	20
Chayote sin espinas	3	8	0.5	500	20
Chile serrano	1	4	0.5	500	20
Coliflor grande	1	4	0.5	500	20
Epazote	1	4	0.5	500	20
Jícama	1	4	0.5	500	20
Kiwi	2	6	0.5	500	20
Lima	2	6	0.5	500	20
Limón #5	2	6	0.5	500	20
Naranja Valencia grande	2	6	0.5	500	20
Nopal grande	2	6	0.5	500	20
Piña mediana	2	6	0.5	500	20
Plátano macho	2	6	0.5	500	20
Plátano Tabasco	2	6	0.5	500	20
Tomate saladette	2	6	0.5	500	20
Toronja roja	2	6	0.5	500	20

Tabla 4.7: Parámetros de ajuste de la red neuronal

La predicción usando la red neuronal recurrente Elman entrenada con los parámetros de la Tabla 4.7 estimó que los valores de los precios futuros de las frutas y hortalizas serán los mostrados en la Tabla 4.8. En el apéndice J se muestran las gráficas de la validación de la red neuronal para cada uno de las 18 frutas y hortalizas.

Producto	S1	S2	S3	S4	S5	S6	S7	S8
Aguacate Hass	35.85	36.13	36.38	36.61	36.79	36.94	37.06	37.17
Cacahuate	65.36	64.84	64.40	64.03	63.72	63.45	63.22	63.03
Calabacita italiana	11.15	10.62	10.28	10.06	9.91	9.82	9.75	9.71
Chayote sin espinas	6.15	6.38	6.99	7.17	7.39	7.46	7.54	7.57
Chile serrano	23.20	16.56	15.74	15.67	15.67	15.67	15.67	15.67
Coliflor grande	13.74	12.97	12.49	12.18	11.99	11.86	11.77	11.71
Epazote	72.69	63.99	59.36	56.82	55.41	54.61	54.16	53.91
Jícama	7.55	7.23	6.99	6.82	6.69	6.59	6.52	6.47
Kiwi	50.59	50.75	50.92	51.08	51.23	51.37	51.50	51.61
Lima	10.07	10.23	10.38	10.54	10.69	10.86	11.02	11.19
Limón c/semilla #5	7.75	7.92	8.10	8.28	8.46	8.63	8.81	8.98
Naranja Valencia grande	4.50	4.57	4.64	4.72	4.79	4.87	4.95	5.04
Nopal grande	21.21	20.54	19.96	19.47	19.05	18.69	18.37	18.09
Piña mediana	9.02	9.10	9.16	9.23	9.29	9.34	9.39	9.44
Plátano macho	14.45	14.25	14.03	13.86	13.72	13.60	13.50	13.41
Plátano Tabasco	6.11	6.21	6.32	6.42	6.53	6.64	6.75	6.86
Tomate saladette	16.15	15.76	15.49	15.26	15.06	14.89	14.75	14.63
Toronja roja	8.54	8.67	8.79	8.90	9.01	9.11	9.22	9.31

Tabla 4.8: Predicción de la red neuronal

## 4.8 Errores cuadrados medios

Una vez estimados los valores futuros de las series de tiempo a través de ambos métodos se procedió a calcular el error absoluto medio, la raíz del error cuadrático medio y el error porcentual absoluto medio de las predicciones. Los resultados de los criterios anteriores se muestran en la Tabla 4.9. Analizando la Tabla 4.9 se puede observar que para algunos productos como la coliflor grande y la naranja Valencia grande el modelo de la red neuronal reduce el valor de los criterios empleados, sin embargo, para algunos productos la predicción es mejor con los modelos ARIMA.

Producto	ARIMA			ERNN		
	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)
Aguacate Hass	3.59	3.75	11.71	5.49	5.88	18.05
Cacahuate	1.16	1.27	1.76	2.15	2.52	3.23
Calabacita italiana	2.82	3.07	25.96	2.10	2.98	15.82
Chayote sin espinas	6.66	7.46	52.82	4.66	5.23	38.64
Chile serrano	29.11	33.73	76.05	35.27	40.20	63.86
Coliflor grande	1.19	1.22	9.30	0.66	0.75	5.18
Epazote	4.30	5.33	6.23	9.06	9.75	13.45
Jícama	0.57	0.78	6.34	1.69	2.00	18.81
Kiwi	7.22	8.22	12.63	4.60	5.39	8.02
Lima	0.44	0.50	4.52	0.89	1.03	9.29
Limón c/semilla #5	1.13	1.56	12.15	1.00	1.23	12.04
Naranja Valencia grande	0.71	0.74	16.98	0.54	0.61	13.09
Nopal grande	8.63	9.37	75.16	5.97	6.52	52.08
Piña mediana	0.38	0.42	4.29	0.40	0.48	4.52
Plátano macho	0.96	1.06	6.89	0.45	0.53	3.17
Plátano Tabasco	4.24	4.52	39.17	3.78	4.05	34.89
Tomate saladette	5.68	6.53	51.64	3.79	4.23	31.43
Toronja roja	0.35	0.40	4.21	0.61	0.68	7.37

Tabla 4.9: Precisión de los modelos

Para determinar si hay una diferencia significativa entre la predicción de ambos modelos se utilizó la prueba F. Para realizar la prueba F, se usaron los errores cuadrados medios definidos como

$$MSE = \frac{1}{N-1} \sum_{t=1}^N (\hat{y}_t - y_t)^2,$$

debido a que el error cuadrado medio es un estimador insesgado de la varianza. Sin embargo, la prueba F requiere que las varianzas muestrales independientes sean tomadas de poblaciones normales, por lo que se aplicó la prueba de Anderson-Darling a los residuos  $e_t = \hat{y}_t - y_t$  de cada modelo para analizar si seguían una distribución normal. Los p valores que resultaron de la prueba de Anderson-Darling se muestran en la Tabla 4.10.

Producto	ARIMA	ERNN
Aguacate Hass	0.01582	0.00382
Cacahuate	0.4343	0.8069
Calabacita italiana	0.0357	0.0338
Chayote sin espinas	0.5341	0.5148
Chile serrano	0.1706	0.1012
Coliflor grande	0.7799	0.3827
Epazote	0.04575	0.01755
Jícama	0.1769	0.5809
Kiwi	0.1278	0.1396
Lima	0.3817	0.7135
Limón c/semilla #5	0.255	0.4335
Naranja Valencia grande	0.7582	0.7231
Nopal grande	0.3591	0.4212
Piña mediana	0.582	0.6524
Plátano macho	0.7366	0.2837
Plátano Tabasco	0.0001	0.0001
Tomate saladette	0.2608	0.2858
Toronja roja	0.2759	0.3568

Tabla 4.10: Prueba de normalidad

La Tabla 4.10 muestra que en la mayoría de los casos la hipótesis de normalidad ( $H_0$ ) no se rechaza para un nivel de significancia de 0.05 a excepción de los residuos del aguacate Hass, la calabacita italiana, el epazote y el plátano Tabasco. Para estos casos, se utilizó el método de remuestreo *bootstrap* para crear 1000 muestras aleatorias bootstrap de los errores obtenidos, calcular el error cuadrado medio de cada una de las muestras y después calcular el estadístico  $\hat{F}^{(b)} = \frac{MSE_1^{(b)}}{MSE_2^{(b)}}$  para cada una de las 1000 muestras aleatorias bootstrap.

Por último, se realizó la estimación del error estándar del estimador F por medio de la desviación estándar muestral de los  $\hat{F}^{(1)}, \dots, \hat{F}^{(1000)}$  y se construyeron los intervalos de confianza del estimador F de la forma  $\hat{F} \pm z_{0.025} \sqrt{V[\hat{F}]}$ . En la Tabla 4.11 se observa que para los intervalos de confianza de la calabacita italiana, el epazote y el plátano Tabasco se contienen al valor 1 concluyendo que la varianza de los residuos de ambos métodos es igual en estos 3 productos mientras que el intervalo de confianza del aguacate Hass no contiene al valor 1.

Producto	$\hat{F} - z_{0.025}\sqrt{V[\hat{F}]}$	$\hat{F} = \frac{MSE_1}{MSE_2}$	$\hat{F} + z_{0.025}\sqrt{V[\hat{F}]}$
Aguacate Hass	1.12	2.46	3.8
Calabacita italiana	-3.79	1.06	5.91
Epazote	-3.79	3.34	10.48
Plátano Tabasco	0.76	1.25	1.73

Tabla 4.11: Intervalos de confianza

Por otra parte, para cada uno de los 14 productos cuyos residuos no rechazaron la hipótesis de normalidad se utilizó el error cuadrado medio de los errores de las predicciones del modelo ARIMA y de la red Elman para realizar la prueba F. Las hipótesis que se contrastaron en la prueba F fueron las siguientes:

- $H_0 : \sigma_1^2 = \sigma_2^2$
- $H_1 : \sigma_1^2 > \sigma_2^2$

donde  $\sigma_1^2$ ,  $\sigma_2^2$  son las varianzas de los errores de predicción de las dos metodologías. Los estimadores de  $\sigma_1^2$  y  $\sigma_2^2$  de los productos se muestran en la Tabla 4.12, así como el valor del estadístico F calculado como se describe en la ecuación 2.27.

Producto	$MSE_{ARIMA}$	$MSE_{ERNN}$	Estimador F
Aguacate Hass	16.16	39.68	2.46
Cacahuate	1.85	7.25	3.92
Calabacita italiana	10.52	10	1.05
Chayote sin espinas	63.59	31.32	2.03
Chile serrano	1300.28	1847.07	1.42
Coliflor grande	1.71	0.64	2.67
Epazote	32.41	109.28	3.37
Jícama	0.70	4.56	6.50
Kiwi	77.15	33.24	2.32
Lima	0.29	1.21	4.19
Limón c/semilla #5	2.79	1.74	1.60
Naranja Valencia grande	0.62	0.43	1.47
Nopal grande	100.32	48.59	2.06
Piña mediana	0.21	0.26	1.26
Plátano macho	1.29	0.33	3.95
Plátano Tabasco	23.3	18.7	1.25
Tomate saladette	48.70	20.42	2.38
Toronja roja	0.18	0.53	2.89

Tabla 4.12: Estadístico F

Para obtener el valor crítico de  $F_{7,7}$ , se usó un nivel de significancia de  $\alpha = 0.05$  obteniendo como resultado el valor  $F_{7,7,0.95} = 3.787044$ . Por consiguiente la región de rechazo esta comprendida por el intervalo  $[3.787044, \infty)$  como se muestra en la Figura 4.5.

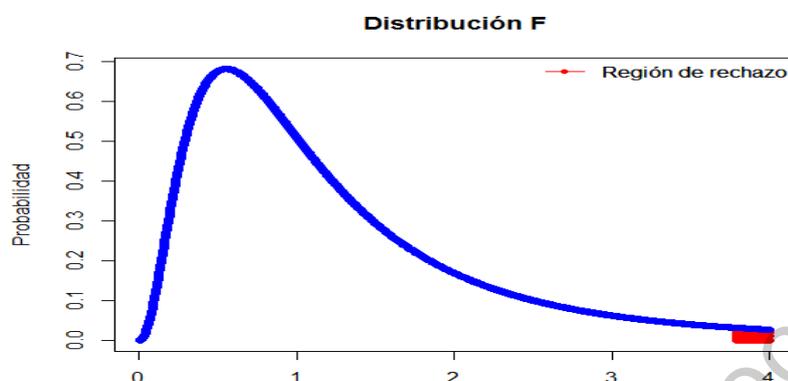


Figura 4.5: Valor crítico de  $F_{7,7}$

Como resultado de los valores del estadístico F de la Tabla 4.9 únicamente 4 productos rechazan la hipótesis nula. El estadístico F del cacahuate, la jícama, la lima y el plátano macho se encuentran en la región de rechazo, por lo cual se rechaza la hipótesis nula de igualdad de varianzas. De manera que la varianza de los errores de las predicciones de la red neuronal es menor que la varianza de los errores de las predicciones de los modelos ARIMA para el plátano macho mientras que para el cacahuate, la jícama y la lima sucede lo contrario.

Para el resto de los 10 productos estudiados no existe evidencia estadística para rechazar la hipótesis nula por lo cual se concluye que la varianza de los errores de las predicciones de la red neuronal recurrente Elman y los modelos autorregresivos integrados de media móvil son iguales.

## 4.9 Red neuronal recurrente Elman

Para la obtención de los resultados anteriores, las entradas de la red neuronal recurrente Elman estuvieron conformadas por los precios semanales promedio de las  $n$  semanas anteriores a la semana a predecir su precio. Sin embargo, existe otro enfoque para realizar la predicción de los precios semanales por medio de la red neuronal recurrente Elman el cual se basa en establecer las entradas de la red como los precios diarios del producto de la semana anterior a estimar.

Empleando la metodología antes descrita se obtuvieron los precios diarios por kilogramo de cada uno de los 18 productos de estudio entre el 1 de enero del 2009 y el 25 de febrero del 2019 excluyendo las fechas correspondientes a los días sábados y domingos en los cuales el SNIIM no tiene registros. Los precios diarios obtenidos contenían algunos valores faltantes los cuales fueron sustituidos por el precio semanal promedio de la semana y año correspondiente al dato faltante.

Para la construcción y entrenamiento de la red neuronal recurrente Elman mediante este nuevo enfoque se utilizó la misma arquitectura y parámetros definidos anteriormente. Las únicas características modificadas para la aplicación del algoritmo usando este nuevo conjunto de datos fue el número de neuronas en la capa de entrada establecido en 5 para todos los productos por los 5 días que conforman cada una de las semana mientras que el número de iteraciones se eligió evitando un sobreajuste en los datos. Los datos finales se muestran en la Tabla 4.13.

Producto	Neuronas de entrada	Neuronas ocultas	Tasa de aprendizaje	Núm. de iteraciones	Tolerancia
Aguacate Hass	5	12	0.5	7	20
Cacahuate	5	12	0.5	2	20
Calabacita italiana	5	12	0.5	7	20
Chayote sin espinas	5	12	0.5	500	20
Chile serrano	5	12	0.5	4	20

Coliflor grande	5	12	0.5	500	20
Epazote	5	12	0.5	500	20
Jícama	5	12	0.5	6	20
Kiwi	5	12	0.5	7	20
Lima	5	12	0.5	500	20
Limón #5	5	12	0.5	500	20
Naranja Valencia grande	5	21	0.5	500	20
Nopal grande	5	12	0.5	500	20
Piña mediana	5	12	0.5	500	20
Plátano macho	5	12	0.5	4	20
Plátano Tabasco	5	12	0.5	500	20
Tomate saladette	5	12	0.5	7	20
Toronja roja	5	12	0.5	6	20

*Tabla 4.13: Parámetros de ajuste de la red neuronal*

La predicción usando la red neuronal recurrente Elman con los precios diarios y entrenados con los parámetros de la Tabla 4.13 estimó que los valores de los precios futuros de las frutas y hortalizas serán los mostrados en la Tabla 4.14.

Producto	S1	S2	S3	S4	S5	S6	S7	S8
Aguacate Hass	36.09	38.30	34.65	29.63	30.89	30.91	31.05	31.26
Cacahuate	52.95	51.78	52.78	52.78	54.18	52.96	52.06	52.74
Calabacita italiana	11.98	8.99	15.37	15.94	13.75	10.34	10.94	9.85
Chayote sin espinas	5.24	5.24	9.89	10.05	13.76	12.29	14.71	9.66
Chile serrano	44.01	38.58	50.69	47.46	44.90	27.77	25.24	29.12
Coliflor grande	14.36	13.57	12.82	12.83	12.73	12.53	12.95	12.96
Epazote	78.54	65.19	69.70	64.94	65.82	65.47	64.28	63.32
Jícama	7.76	7.70	7.48	7.78	9.67	9.86	7.87	8.37
Kiwi	50.69	50.68	50.54	55.33	55.97	57.79	53.29	54.73
Lima	10.17	11.57	9.80	9.66	9.55	9.05	10.07	9.25
Limón c/semilla #5	7.68	8.59	6.40	7.27	7.24	8.06	8.58	9.84
Naranja Valencia grande	4.55	4.58	4.17	4.43	4.52	3.99	4.50	4.08
Nopal grande	21.25	18.94	16.21	15.35	12.80	11.88	12.17	10.30
Piña mediana	9.08	8.60	8.49	9.16	9.67	9.42	9.16	9.23
Plátano macho	14.01	13.59	13.55	13.34	13.54	14.00	13.12	13.94
Plátano Tabasco	6.10	6.15	10.50	10.60	10.70	10.65	10.82	10.80
Tomate saladette	16.29	15.05	19.74	18.97	11.77	9.38	11.15	11.50
Toronja roja	8.63	8.19	8.65	8.76	9.65	8.23	9.00	8.20

*Tabla 4.14: Predicción de la red neuronal*

Una vez estimados los valores futuros de las series de tiempo se procedió a calcular el error absoluto medio, la raíz del error cuadrático medio y el error porcentual absoluto medio de las predicciones. Los resultados de los criterios anteriores se muestran en la Tabla 4.15.

Producto	MAE	RMSE	MAPE
Aguacate Hass	1.66	2.78	5.36
Cacahuate	13.38	13.42	20.21
Calabacita italiana	2.79	3.23	24.77
Chayote sin espinas	1.91	2.37	16.57
Chile serrano	16.63	21.05	29.19
Coliflor grande	0.36	0.50	2.79
Epazote	3.05	4.55	4.25
Jícama	0.52	0.62	6.00
Kiwi	2.41	2.71	4.26
Lima	0.43	0.67	4.38
Limón c/semilla #5	1.02	1.25	12.67
Naranja Valencia grande	0.26	0.31	6.07
Nopal grande	1.48	1.78	12.13
Piña mediana	0.31	0.40	3.55
Plátano macho	0.54	0.69	3.73
Plátano Tabasco	0.75	1.48	7.18
Tomate saladette	2.92	3.92	22.34
Toronja roja	0.51	0.65	6.14

*Tabla 4.15: Errores de la red neuronal*

Para determinar si hay una mejora con este nuevo enfoque de utilizar los precios diarios como datos de entrada de la red se utilizó la prueba F para comparar el error cuadrado medio de este modelo con los dos modelos estudiados anteriormente.

En la Tabla 4.16 se puede observar que para la mayoría de las frutas y hortalizas estudiadas el error cuadrado medio disminuye al utilizar la red neuronal recurrente Elman con los precios diarios como entradas. Además, en ambas comparaciones se obtuvo que 6 productos rechazaron la hipótesis nula de igualdad de varianzas en la prueba F realizadas y 5 de estos 6 productos rechazaron la hipótesis nula con la conclusión de que la ERNN usando

los precios diarios como entradas tiene una varianza menor. Sin embargo, este nuevo enfoque presenta un problema en la predicción el cual es que únicamente se puede predecir un paso hacia adelante lo cual no sirve para realizar una planeación a futuro.

Producto	MSE ARIMA	MSE ERNN por día	Estimador F	Producto	MSE ERNN por semana	MSE ERNN por día	Estimador F
Cacahuete	1.85	205.93	111.33	Cacahuete	7.25	205.93	28.41
Chayote sin espinas	63.59	6.40	9.93	Chayote sin espinas	31.32	6.40	4.89
Chile serrano	1300.28	506.36	2.57	Chile serrano	1847.07	506.36	3.65
Coliflor grande	1.71	0.29	5.93	Coliflor grande	0.64	0.29	2.23
Jícama	0.70	0.43	1.62	Jícama	4.56	0.43	10.52
Kiwi	77.15	8.41	9.17	Kiwi	33.24	8.41	3.95
Limón c/semilla #5	2.79	1.79	1.55	Limón c/semilla #5	1.74	1.79	1.03
Naranja Valencia grande	0.62	0.11	5.7	Naranja Valencia grande	0.43	0.11	3.92
Nopal grande	100.32	3.61	27.79	Nopal grande	48.59	3.61	13.46
Piña mediana	0.21	0.18	1.13	Piña mediana	0.26	0.18	1.43
Tomate saladette	48.70	17.59	2.77	Tomate saladette	20.42	17.59	1.16
Toronja roja	0.18	0.48	2.63	Toronja roja	0.53	0.48	1.10

Tabla 4.16: Precisión de los modelos

## CAPÍTULO 5

### Discusión y conclusiones

#### 5.1 Discusión

Los datos de este trabajo indican que para la mayoría de los productos el error cuadrado medio de la red neuronal recurrente de Elman y los modelos Arima no son diferentes. Sin embargo, por cuestiones de tiempo y limitaciones en la capacidad de cómputo no se analizaron diversas posibilidades de las metodologías que podrían obtener predicciones distintas a las presentadas en este trabajo.

En futuras investigaciones se podría llegar a considerar buscar una configuración óptima de la arquitectura de la red que disminuya el valor del MSE de la predicción. En el presente trabajo se usaron el doble de neuronas de la capa de entrada más dos como número de neuronas ocultas, sin embargo, tener más neuronas en la capa de entrada y en la capa oculta puede mejorar la salida de red. Además se podría modificar la estructura de la red neuronal añadiendo un parámetro de sesgo en la entrada total de cada una de las neuronas de la red debido a la red neuronal recurrente Elman establece el sesgo igual a cero para todas las neuronas.

En el desarrollo del algoritmo de la red neuronal recurrente Elman se hicieron supuestos sobre algunos de los valores usados, los cuales representan una gran oportunidad de mejora para la predicción. Entre los parámetros supuestos se encuentra el valor del coeficiente de resistencia del tiempo presente de la función estocástica efectiva del tiempo empleada en la red neuronal recurrente Elman, el tamaño del conjunto de entrenamiento y validación, el parámetro de tolerancia y el número de iteraciones de entrenamiento.

## 5.2 Conclusiones

La predicción es un tema de gran importancia para diversas áreas como la economía y la agricultura, de manera que las series de tiempo han sido estudiadas desde hace muchos años. Como resultado, diversas metodologías han sido inventadas o ajustadas para realizar predicciones, disminuir el error de las estimaciones.

En este trabajo se trabajó con las series de tiempo de los precios semanales promedio de las frutas y hortalizas del mercado de abastos de Querétaro y se emplearon las metodologías del modelo autorregresivo integrado de media móvil y la red neuronal recurrente Elman, las cuales fueron aplicadas y se comparó su desempeño.

Los resultados de este trabajo sugieren que la red neuronal recurrente Elman no realiza una mejor predicción que los modelos autorregresivos integrados de media móvil, presentando una varianza de error de predicción igual en la mayoría de los casos de estudio.

Además, la predicción utilizando la red neuronal de Elman es más sencilla de usar debido a que esta sólo se basa en la definición de un conjunto de parámetros o características mientras que los modelos autorregresivos integrados de media móvil necesitan un conocimiento matemático más profundo y un análisis más complejo para la determinación de parámetros.

Se debe agregar que la red neuronal empleada no es la red más sofisticada en cuestión de arquitectura y la configuración de los parámetros no es óptima. Por esta razón se puede aún reducir el error de predicción de la serie de tiempo, ampliar el tiempo de predicción, entre otras cosas que abren el panorama de diversos caminos para alcanzar una mejora en la predicción.

Este trabajo me hizo entender las dificultades que se tiene al trabajar con datos de un problema real, la importancia del preprocesamiento de los datos y la diversidad de enfoques y alternativas con las cuales se puede enfrentar un mismo problema. Además, a lo largo de este proyecto adquirí nuevos conocimientos, experiencia trabajando con base de datos y habilidades de programación.

## APÉNDICE A

### Productos de la base de datos

Hortalizas			
Acelga	Champiñón	Coliflor grande	Nopal
Ajo blanco	Chayote	Coliflor mediana	Nopal grande
Ajo morado	Chayote sin espinas	Ejote	Papa alpha
Apio	Chícharo	Elote	Pepino
Betabel	Chile chilaca	Elote grande	Rábano
Brócoli	Chile jalapeño	Epazote	Sandía rayada
Cacahuate	Chile pimiento morrón	Espinaca	Sandía sangría
Calabacita italiana	Chile poblano	Jícama	Tomate saladette
Calabaza de castilla	Chile serrano	Lechuga romanita grande	Tomate verde
Camote	Cilantro	Melón Cantaloupe #12	Zanahoria leña
Cebolla bola	Col grande	Melón Cantaloupe #27	Zanahoria mediana
Cebolla bola grande	Col mediana	Melón Cantaloupe sin clasificación	Zanahoria polvo
Chiles secos			
Chile ancho	Chile pasilla	Chile puya seco	Chile de árbol seco
Chile guajillo			
Frutas			
Aguacate Hass	Manzana golden delicious	Papaya maradol	Plátano macho
Fresa	Manzana red delicious	Piña mediana	Plátano Tabasco
Limón c/semilla #3	Naranja Valencia grande	Plátano Chiapas	Toronja roja
Limón c/semilla #5	Naranja Valencia mediana	Plátano dominico	Toronja roja mediana
Frutas de temporada			
Caña	Kiwi	Mango manila	Uva blanca
Cereza	Lima	Pera D'anjou #100	Uva globo
Ciruela roja	Mandarina	Tamarindo	Uva Perlette
Durazno amarillo	Mandarina Mónica	Tejocote	Uva superior
Granada roja	Mango Ataulfo	Tuna blanca	Uva Thompson
Guayaba	Mango Haden		

Tabla A.1: Productos comercializados en el mercado de Querétaro

## APÉNDICE B

### Análisis descriptivo

Nombre	Núm. datos	Precio mínimo	Precio máximo	IQR	Media	Varianza
Acelga	1	3	3	0	3	NA
Aguacate Hass	583	13	63.2	14.5	28.65	105.03
Ajo blanco	106	9.84	72.5	12.48	25.14	181.31
Ajo morado	530	21	95	22.6	50.03	280
Apio	371	9.4	21	2.86	12.31	4.92
Betabel	265	2.2	8.42	1.03	3.75	1.59
Brócoli	530	4.3	15	1.1	6.67	2.26
Cacahuate	530	23	73.8	11.8	40.73	124.31
Calabacita italiana	583	3.28	24	3.96	8.09	11.64
Calabaza de castilla	192	4	14.33	2.2	7.55	4.39
Camote	583	5.2	19.6	5.06	11.51	8.8
Caña	162	4.3	11.5	1.44	5.71	1.67
Cebolla bola	530	1.47	23.33	4.4	7	19.02
Cebolla bola grande	212	3.3	24.7	5.4	8.68	24.56
Cereza	1	12	12	0	12	NA
Champiñón	159	47	58	5.2	52.65	9.5
Chayote	159	3.5	14.11	2.08	6.22	3.98
Chayote sin espinas	530	2.3	20.42	3.3	6.9	7.21
Chícharo	265	5	34.2	12.75	18.98	58.52
Chile ancho	583	45	140.2	35.25	86.27	608.33
Chile chilaca	318	8.8	30.8	6.2	16.36	18.18

Chile de árbol seco	583	36.8	104	14.22	68.29	295.13
Chile guajillo	583	34.67	110	28.4	63.3	312.38
Chile jalapeño	583	4	45.4	4.25	10.81	27.69
Chile pasilla	583	47	136.25	23.75	76.16	455.06
Chile pimiento morrón	583	6.67	37.8	7.52	20.71	33.04
Chile poblano	583	4.62	31	6.45	13.55	21.97
Chile puya seco	477	40.8	130	25.8	70.87	533.24
Chile serrano	583	3.73	83.4	7.5	13.47	62.88
Cilantro	159	2.83	16	3.81	7.41	8.34
Ciruella roja	174	10	32.2	10	19.91	33.64
Col grande	1	10	10	0	10	NA
Col mediana	583	5.17	20.6	3.38	9.3	7.65
Coliflor grande	424	7	16.8	2.6	10.71	3.45
Coliflor mediana	212	6	16.12	5.55	10.58	9.18
Durazno amarillo	76	10.45	48	2.75	17.21	145.93
Ejote	583	6	33	5.84	12.86	21.1
Elote	104	1.97	5	0.92	2.79	0.48
Elote grande	424	2	4.3	1	3.06	0.3
Epazote	318	40	90	12.1	53.89	82.84
Espinaca	371	6	43	9.4	25.34	32.82
Fresa	583	9.51	52.5	12.5	31.71	89.41
Granada roja	10	4.88	5.6	0.23	5.17	0.07
Guayaba	583	4.62	17.69	2	8.81	3.05
Jícama	583	2.5	13.4	1.9	5.43	2.74
Kiwi	583	21.6	69	16	40.8	97.42
Lechuga romanita grande	583	2	11.6	1.9	5.03	2.92
Lima	371	5.6	21	5.4	10.51	13.79
Limón c/semilla # 3	212	2.21	17.6	2.3	5.1	5.45
Limón c/semilla # 5	424	3.05	28.4	4.38	8.59	18.9

Mandarina	407	1.78	14.25	3.64	6.4	5.15
Mandarina Mónica	2	7.8	8	0.1	7.9	0.02
Mango Ataulfo	390	6.36	33.33	7.55	13.05	27.05
Mango Haden	111	4.87	17.67	6.98	9.16	13.94
Mango manila	429	6.36	28.33	4.97	11.11	12.31
Manzana golden delicious	583	14	53.5	14.65	29.56	88.21
Manzana red delicious	583	15.4	39.62	8.98	24.69	33.88
Melón Cantaloupe #12	8	6.5	7.5	0.44	7.06	0.18
Melón Cantaloupe #27	212	4.11	10.86	2.12	7.32	2.23
Melón Cantaloupe sin clasificación	424	4.56	19.8	3.15	9.79	6.43
Naranja Valencia grande	424	2.05	13.8	2.58	4.97	5.36
Naranja Valencia mediana	212	1.8	9.66	2.14	3.8	2.16
Nopal	104	6.9	14.8	4.2	10.75	5.68
Nopal grande	424	6.9	31	4.13	11.96	17.36
Papa alpha	583	6.12	23	3.6	11.06	10.36
Papaya maradol	583	5.88	23	4.6	11.03	10.55
Pepino	583	1.48	18.64	3.84	8.03	8.31
Pera D´anjou #100	583	16	52.2	11.27	27.91	49.13
Piña mediana	477	3.5	16	3.1	8.32	5.11
Plátano Chiapas	1	5	5	0	5	NA
Plátano dominico	583	6.28	20.62	2.98	11.12	6.25
Plátano macho	583	5.56	15.75	2.9	10.24	4.64
Plátano Tabasco	583	2.45	12.5	2.43	5.86	4.33
Rábano	1	12	12	0	12	NA
Sandía rayada	424	3.4	9.5	1.6	6.02	1.61
Sandía sangría	212	3.5	8	0.57	4.66	0.55
Tamarindo	2	12.33	14.67	1.17	13.5	2.72
Tejocote	28	3.12	12.22	6.56	7.44	13.13
Tomate saladette	583	3.65	28.53	5	10.09	17

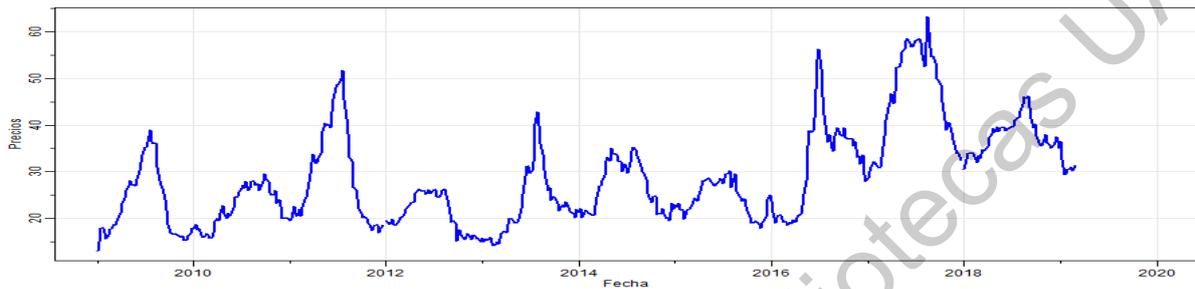
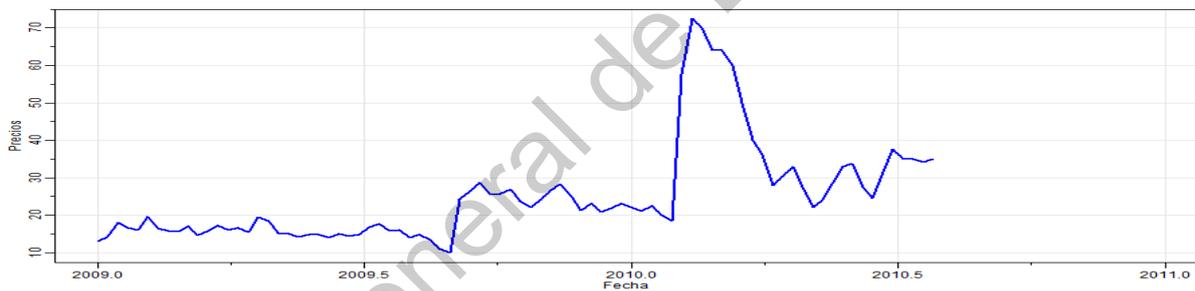
Tomate verde	583	2.78	32	4.49	8.64	18.4
Toronja roja	583	2.52	15.2	3.2	5.78	6.83
Toronja roja mediana	10	3.17	4.12	0.5	3.76	0.17
Tuna blanca	147	3.1	11.76	2.94	7.45	4.3
Uva blanca	5	18.98	20.91	0.09	20.23	0.54
Uva globo	583	20.43	68.75	16.41	46.6	121.64
Uva Perlette	65	33.44	97.5	12.92	56.67	183.45
Uva superior	352	31.39	112.5	19.56	58.3	230.81
Uva Thompson	530	25.47	103.5	22.44	55.8	225.18
Zanahoria leña	159	1.12	8.79	1.58	3.18	2.34
Zanahoria mediana	583	2.21	9.14	1.2	4.19	1.66
Zanahoria polvo	159	1.61	7.43	0.81	3.11	1.33

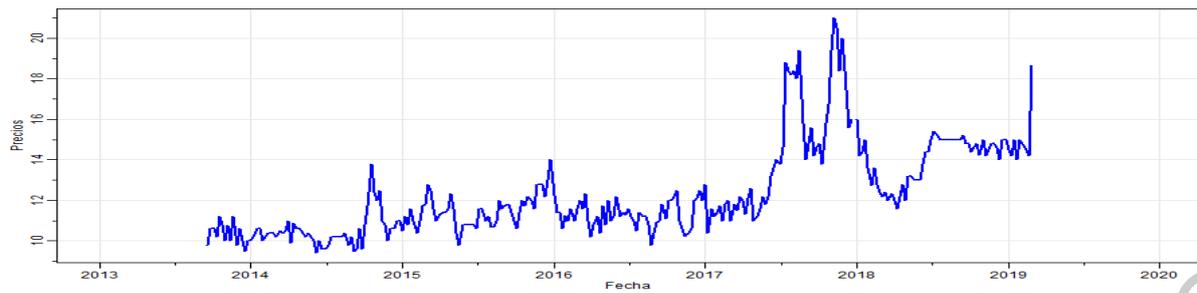
*Tabla B.1: Análisis descriptivo de los datos*

Dirección General de Bibliotecas UQO

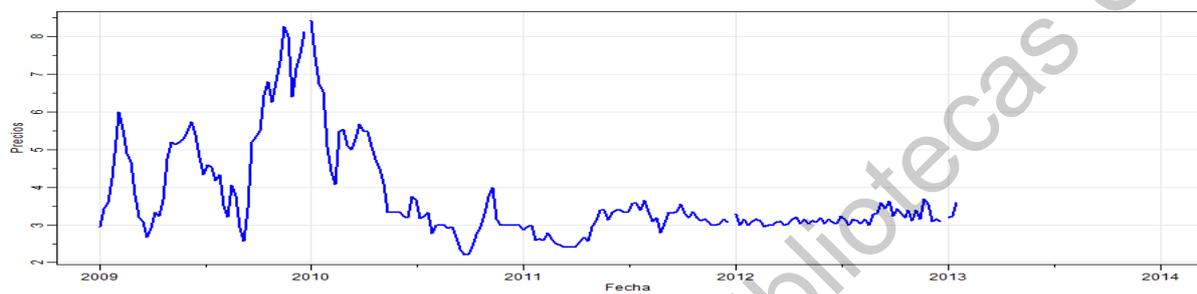
## APÉNDICE C

## Gráficas: Series de tiempo

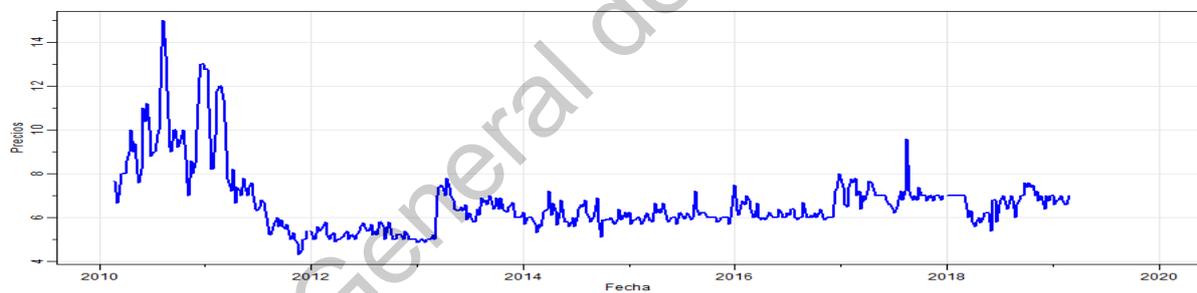
*Figura C.1: Precios del aguacate Hass**Figura C.2: Precios del ajo blanco**Figura C.3: Precios del ajo morado*



*Figura C.4: Precios del apio*



*Figura C.5: Precios del betabel*



*Figura C.6: Precios del brócoli*



*Figura C.7: Precios del cacahuate*

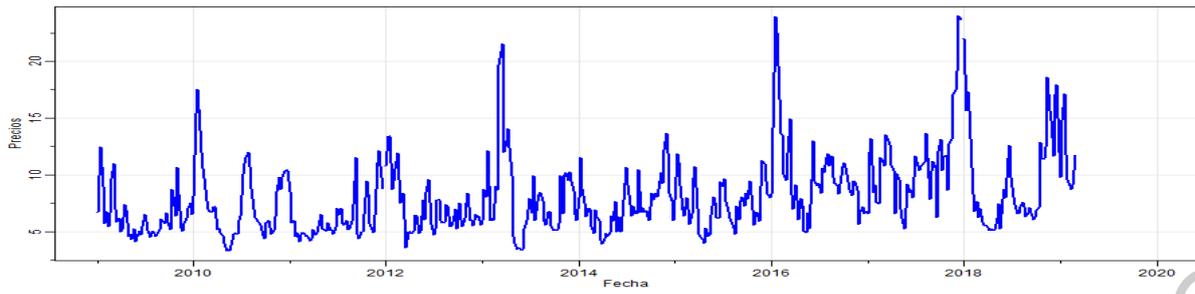


Figura C.8: Precios de la calabacita italiana

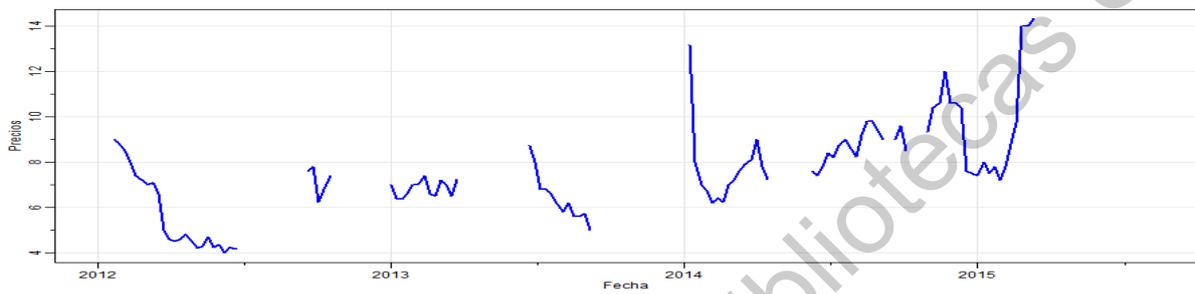


Figura C.9: Precios de la calabaza de castilla

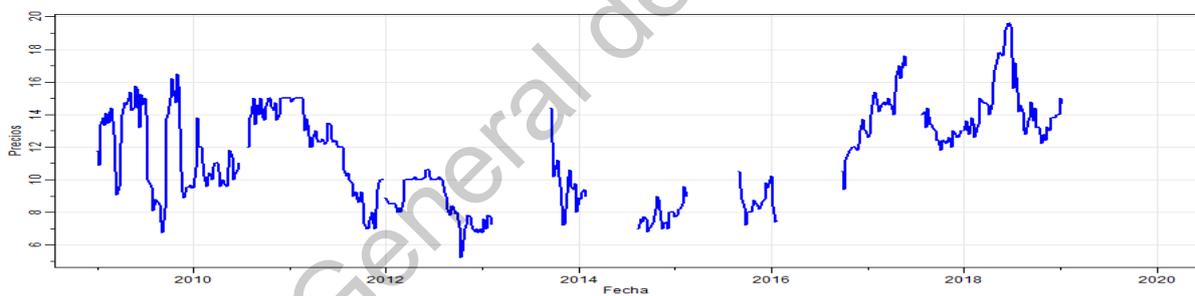


Figura C.10: Precios del camote

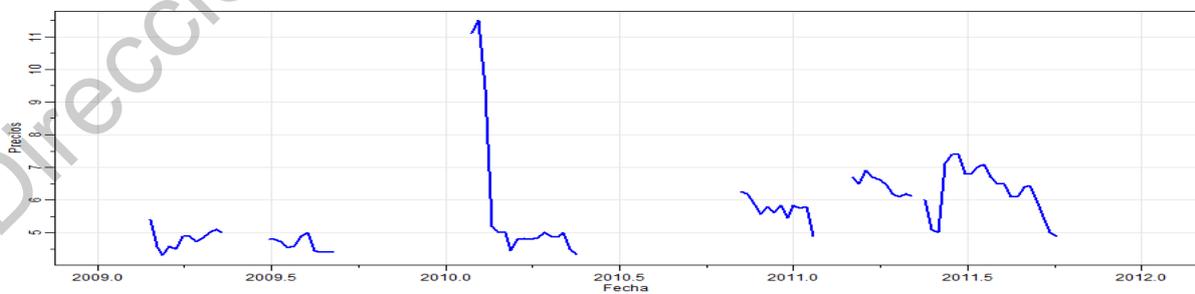
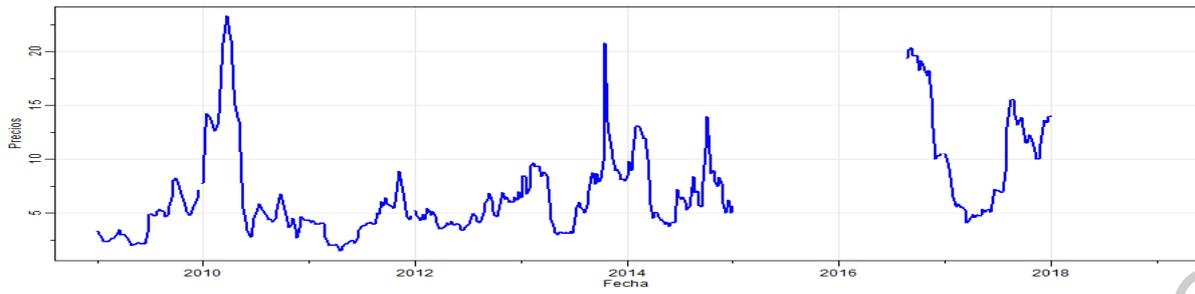
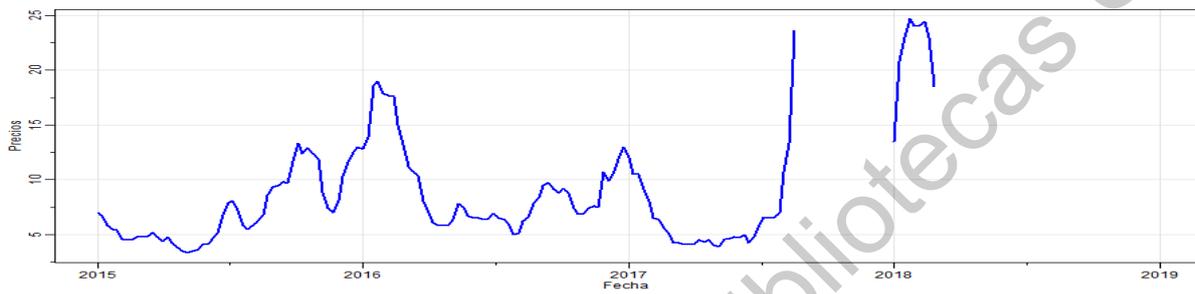


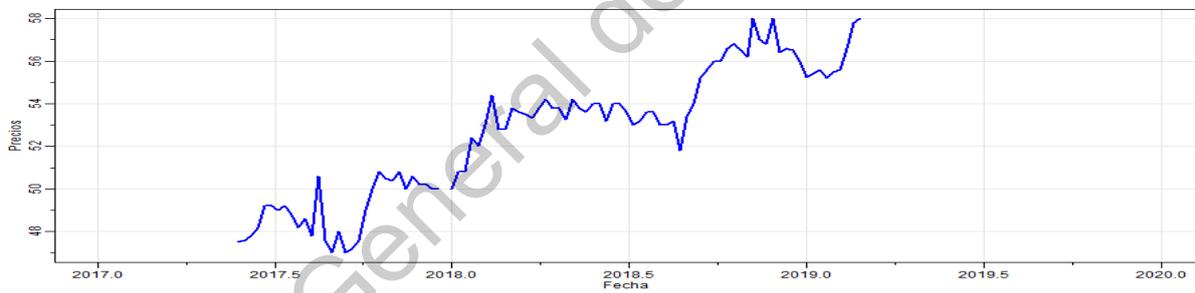
Figura C.11: Precios de la caña



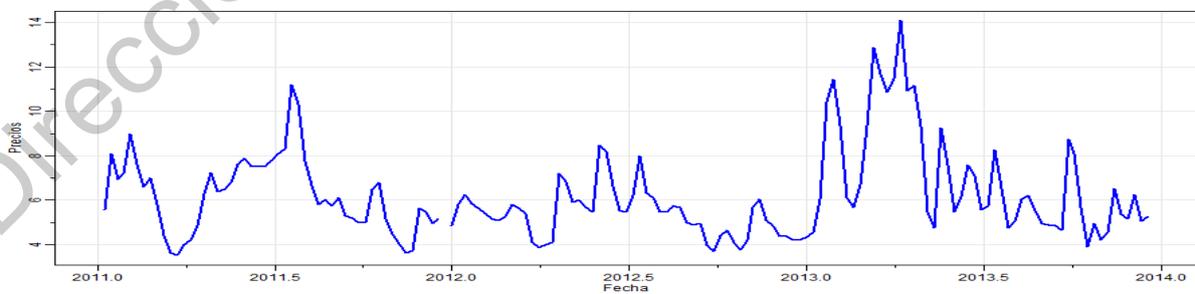
*Figura C.12: Precios de la cebolla bola*



*Figura C.13: Precios de la cebolla bola grande*



*Figura C.14: Precios del champiñón*



*Figura C.15: Precios del chayote*

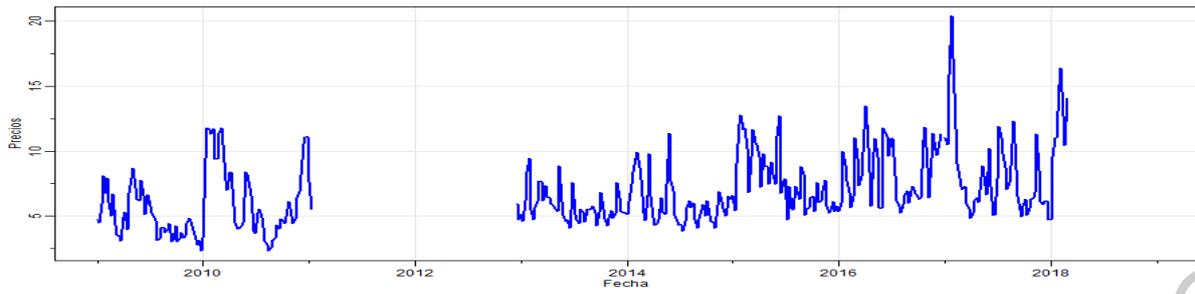


Figura C.16: Precios del chayote sin espinas

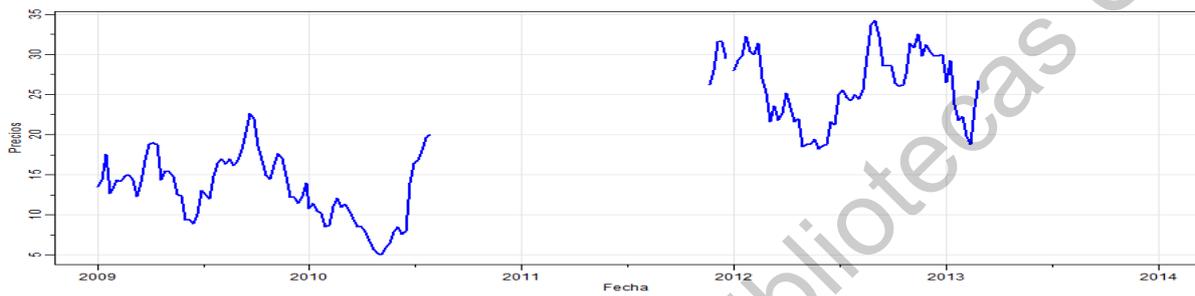


Figura C.17: Precios del chicharo



Figura C.18: Precios del chile ancho

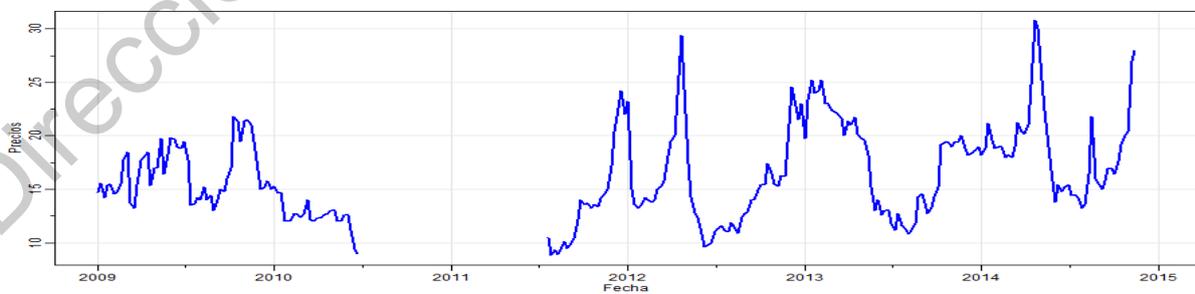
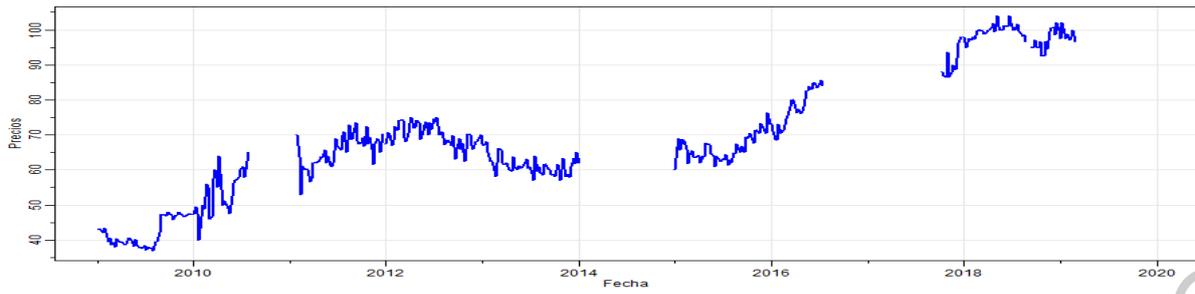


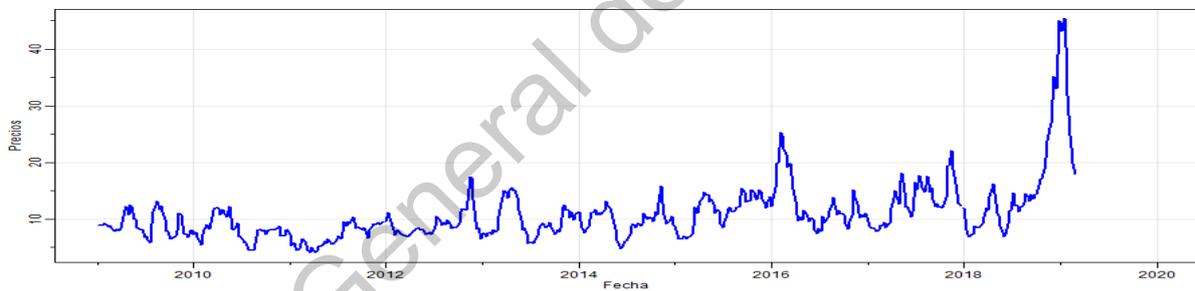
Figura C.19: Precios del chile chilaca



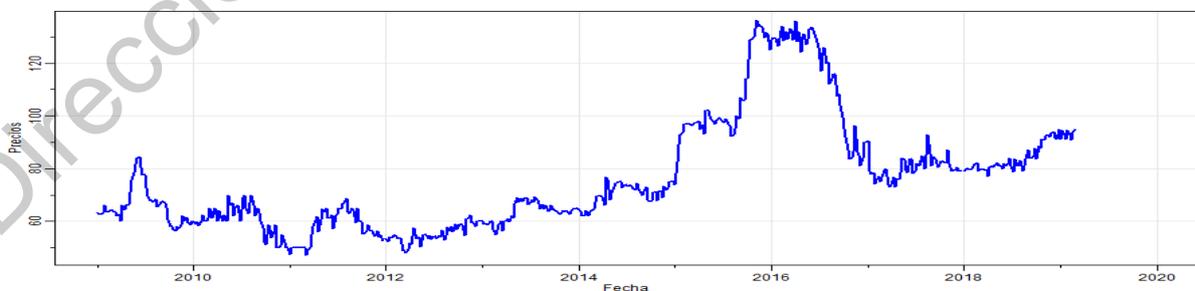
*Figura C.20: Precios del chile de árbol seco*



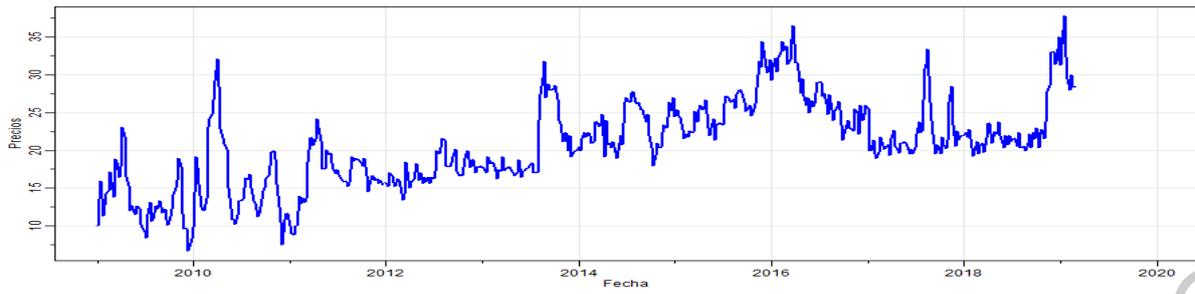
*Figura C.21: Precios del chile guajillo*



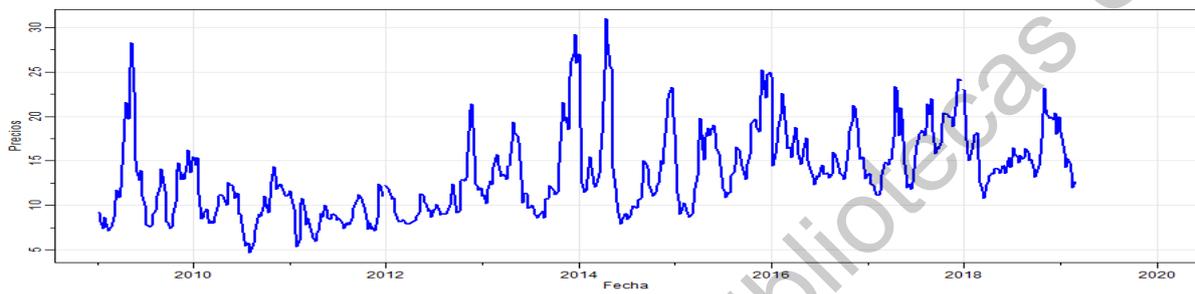
*Figura C.22: Precios del chile jalapeño*



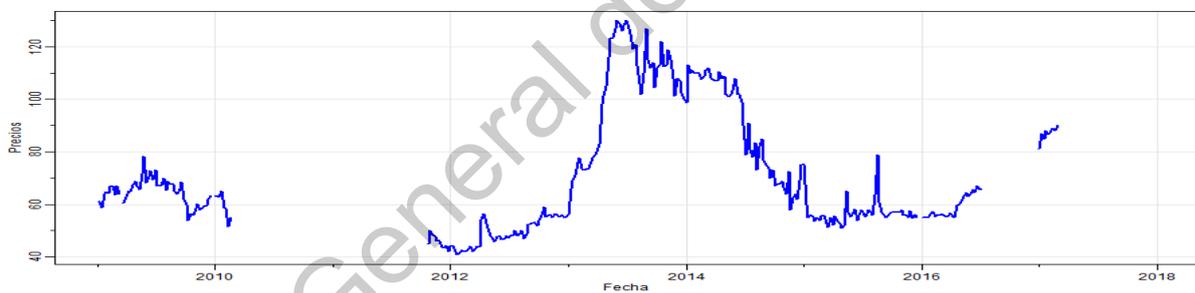
*Figura C.23: Precios del chile pasilla*



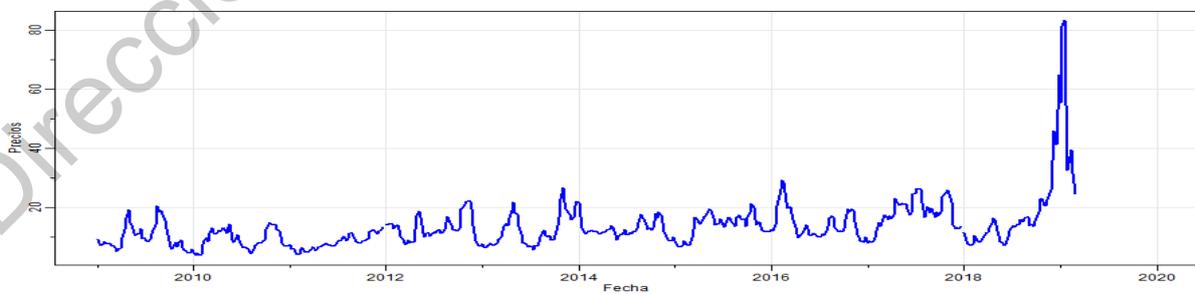
*Figura C.24: Precios del chile pimienta morrón*



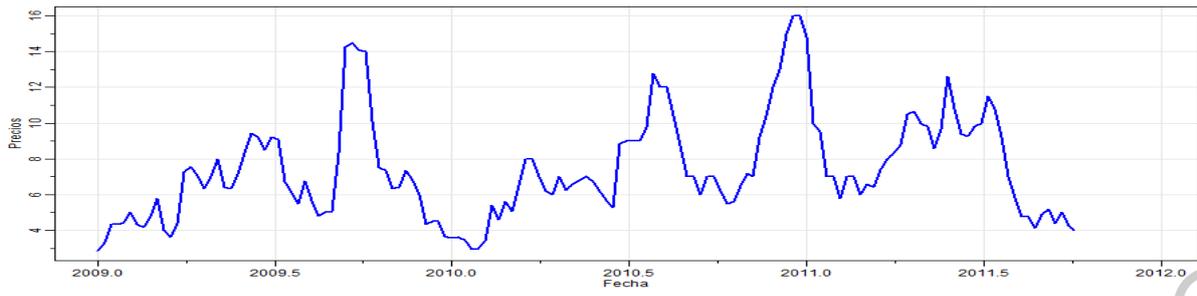
*Figura C.25: Precios del chile poblano*



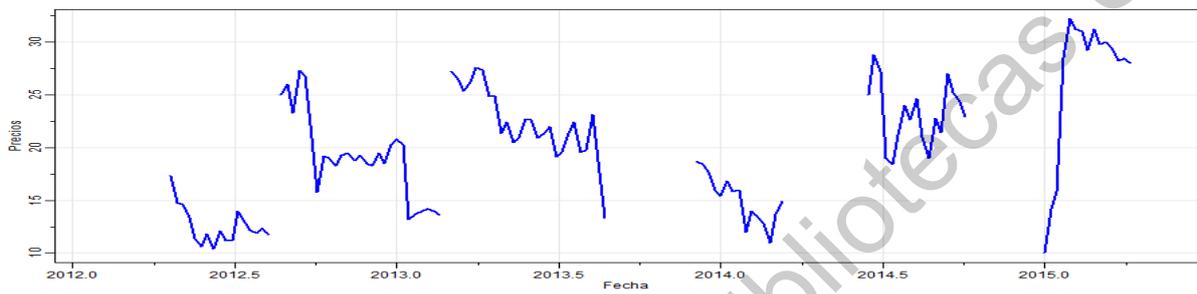
*Figura C.26: Precios del chile puya seco*



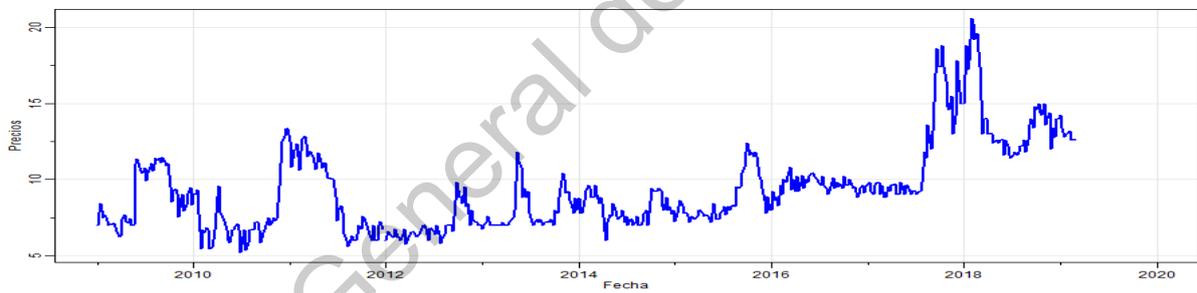
*Figura C.27: Precios del chile serrano*



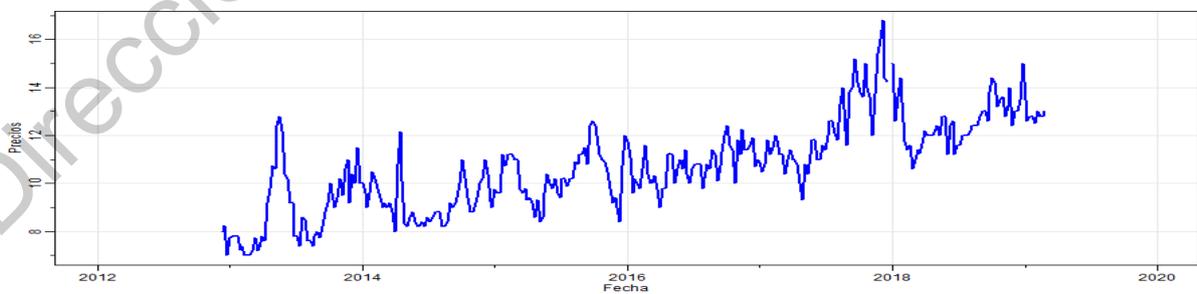
*Figura C.28: Precios del cilantro*



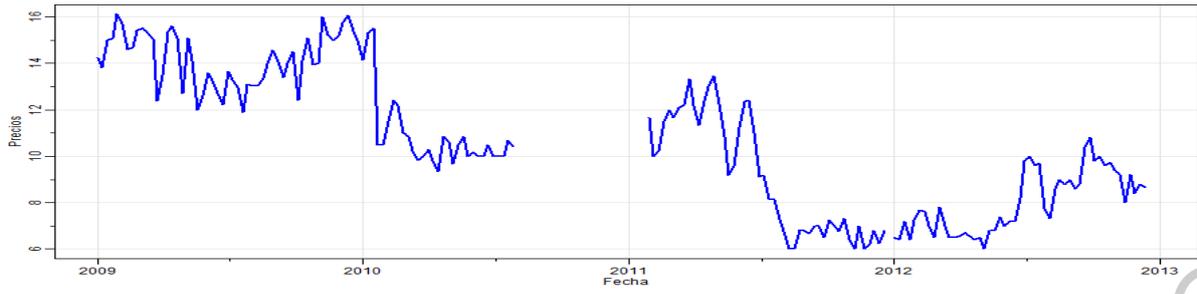
*Figura C.29: Precios de la ciruela roja*



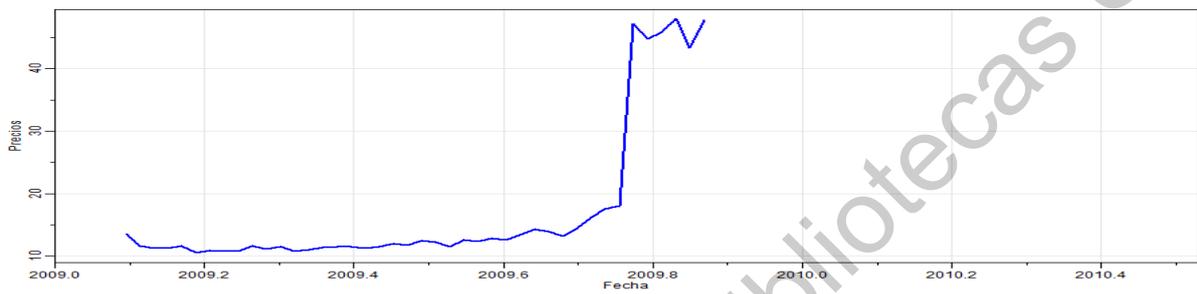
*Figura C.30: Precios de la col mediana*



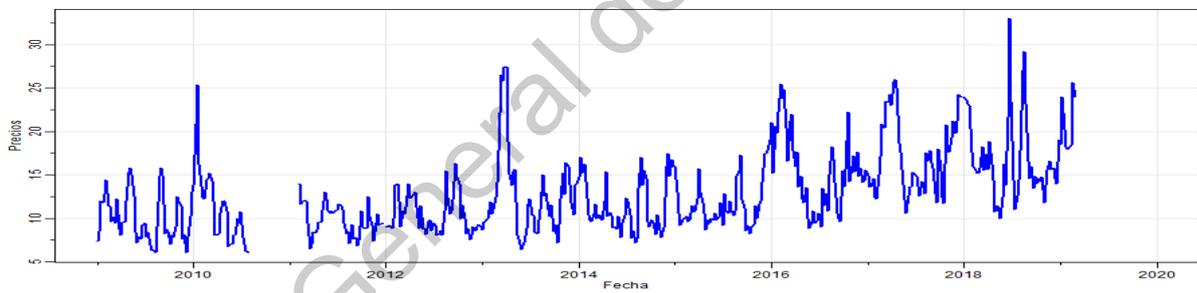
*Figura C.31: Precios de la coliflor grande*



*Figura C.32: Precios de la coliflor mediana*



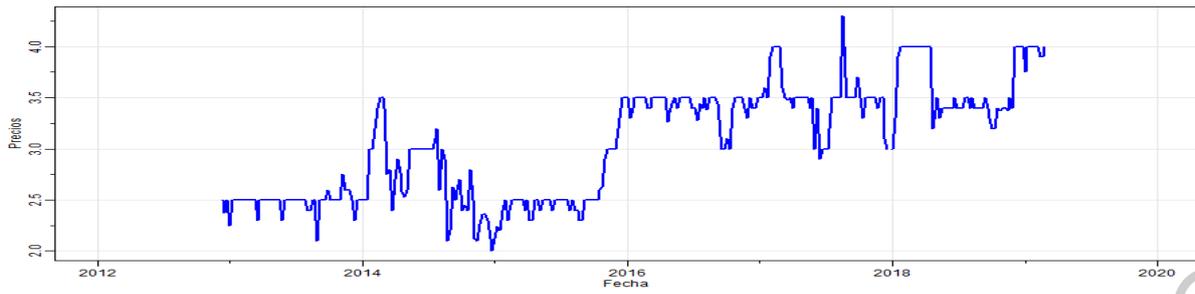
*Figura C.33: Precios del durazno amarillo*



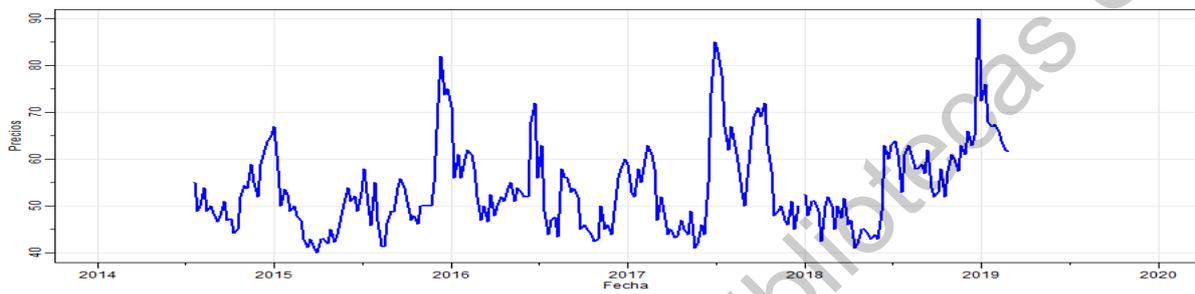
*Figura C.34: Precios del ejote*



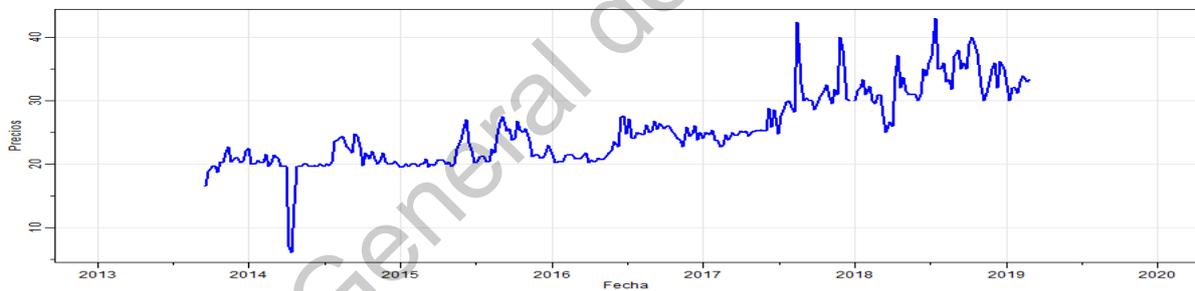
*Figura C.35: Precios del elote*



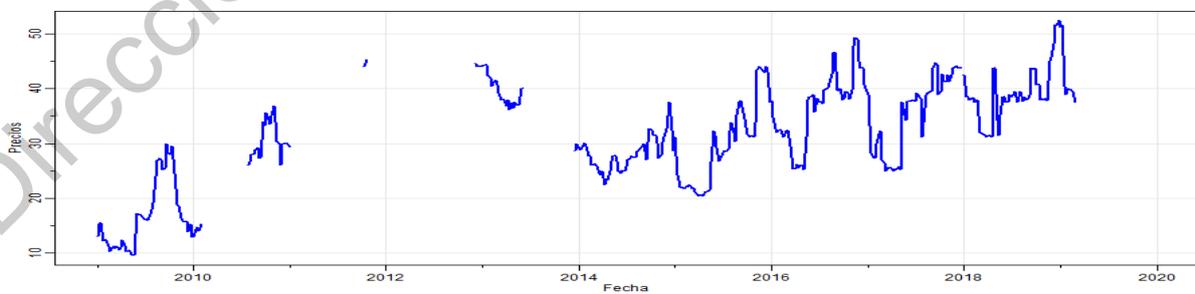
*Figura C.36: Precios del elote grande*



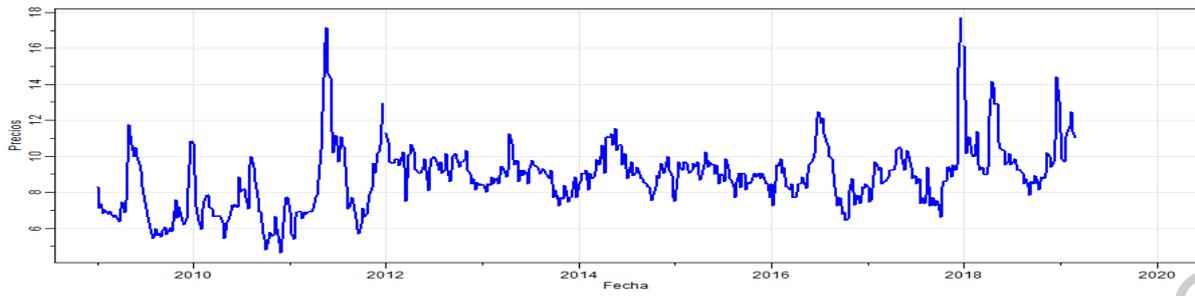
*Figura C.37: Precios del epazote*



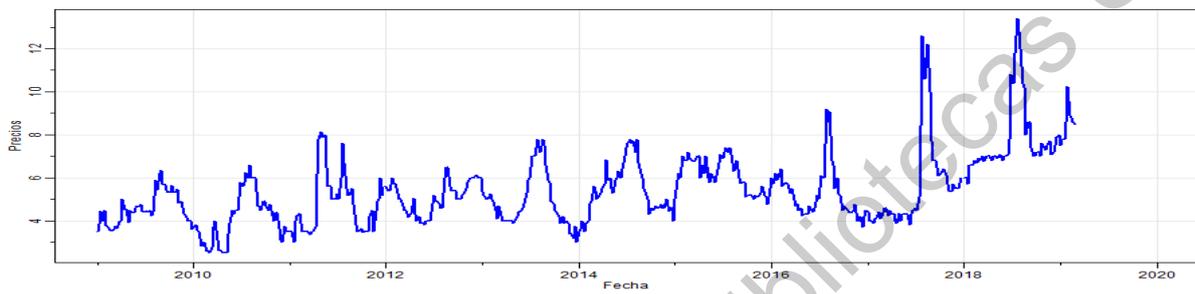
*Figura C.38: Precios de la espinaca*



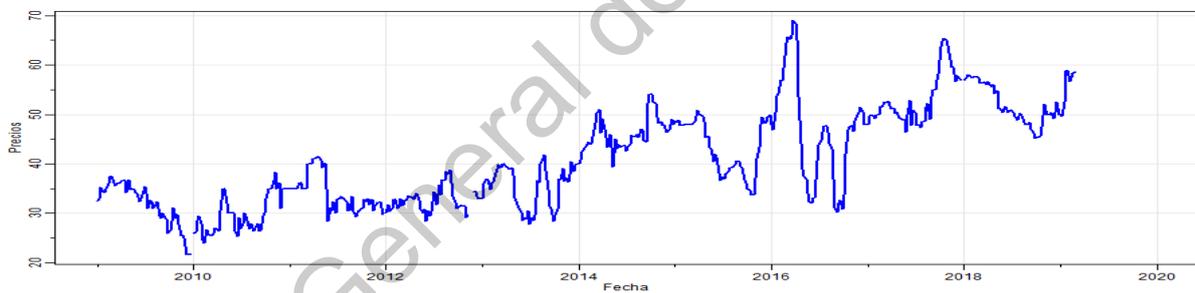
*Figura C.39: Precios de la fresa*



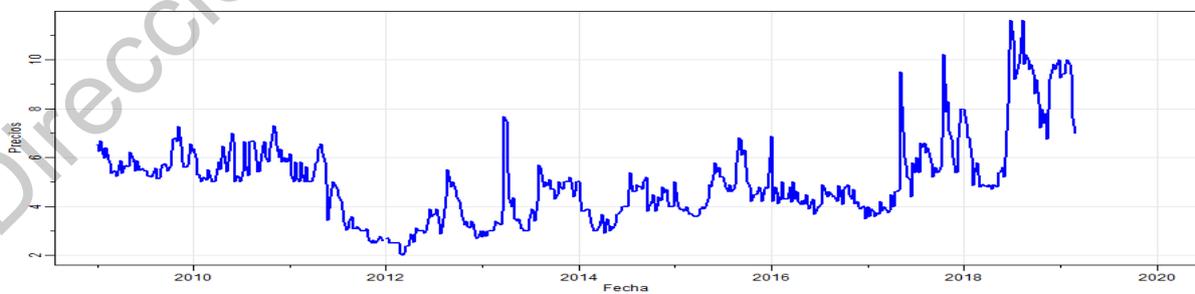
*Figura C.40: Precios de la guayaba*



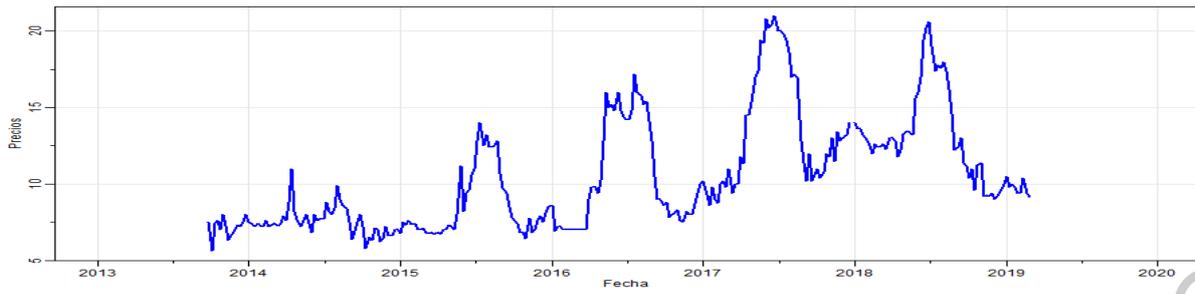
*Figura C.41: Precios de la jícama*



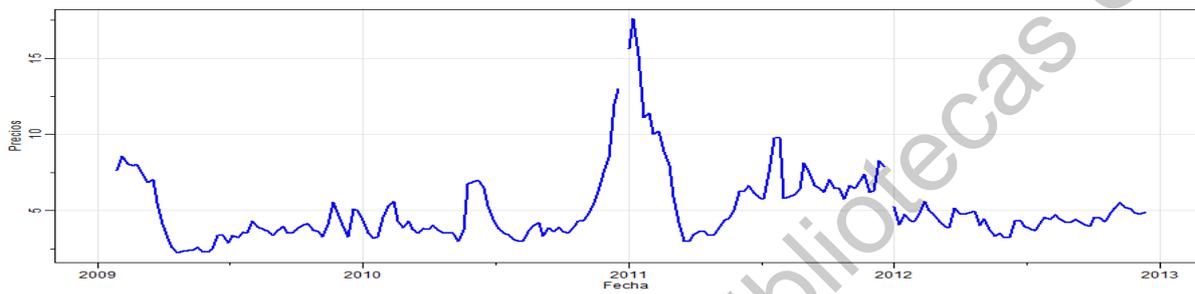
*Figura C.42: Precios del kiwi*



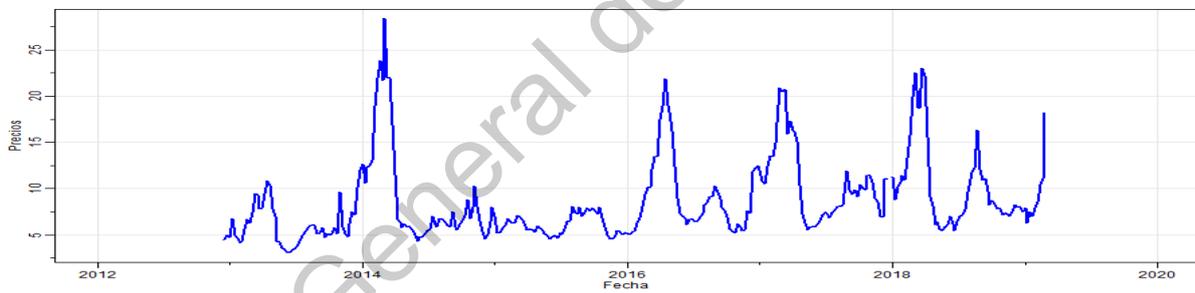
*Figura C.43: Precios de la lechuga romanita grande*



*Figura C.44: Precios de la lima*



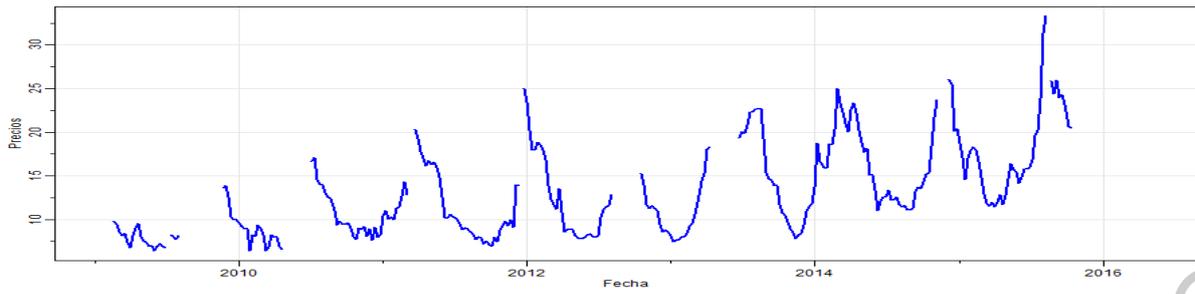
*Figura C.45: Precios del limón con semilla #3*



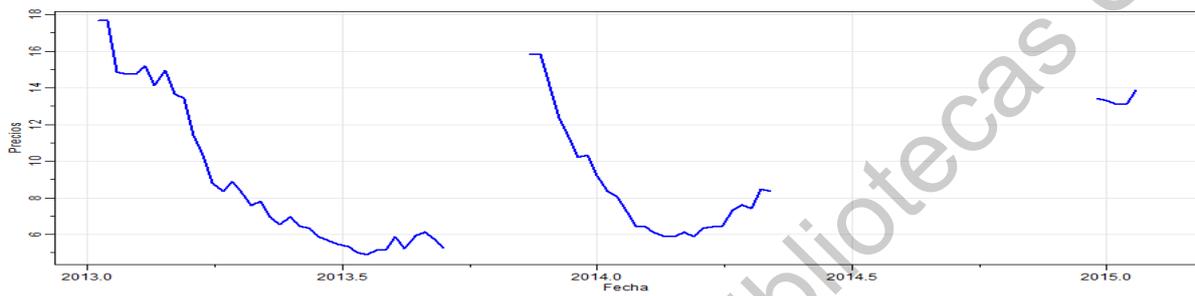
*Figura C.46: Precios del limón con semilla #5*



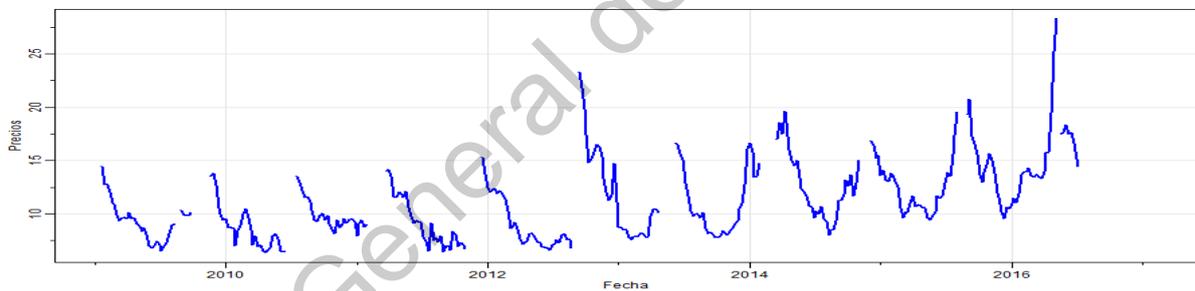
*Figura C.47: Precios de la mandarina*



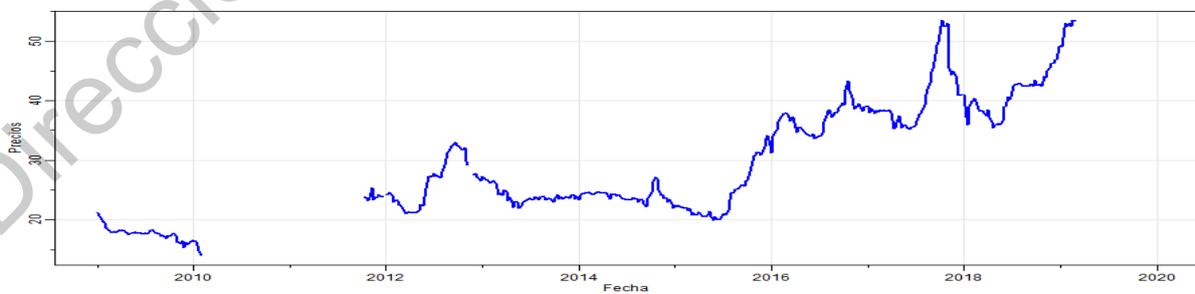
*Figura C.48: Precios del mango Ataulfo*



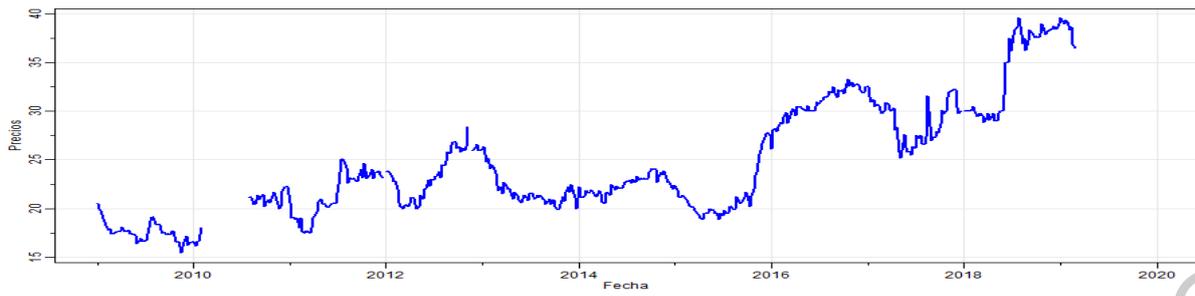
*Figura C.49: Precios del mango Haden*



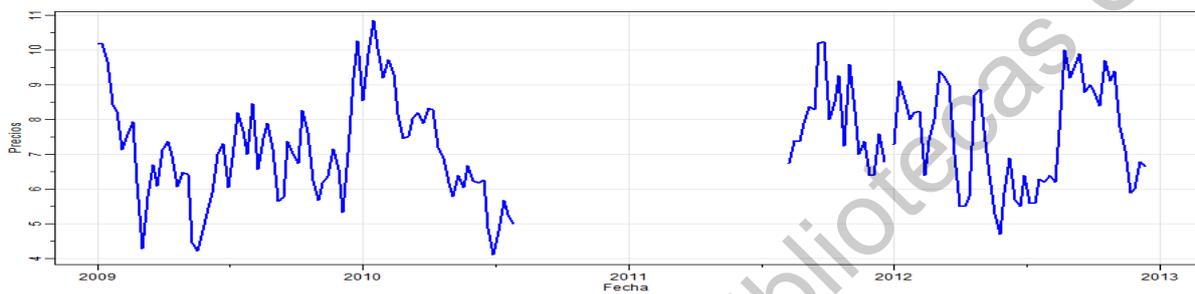
*Figura C.50: Precios del mango manila*



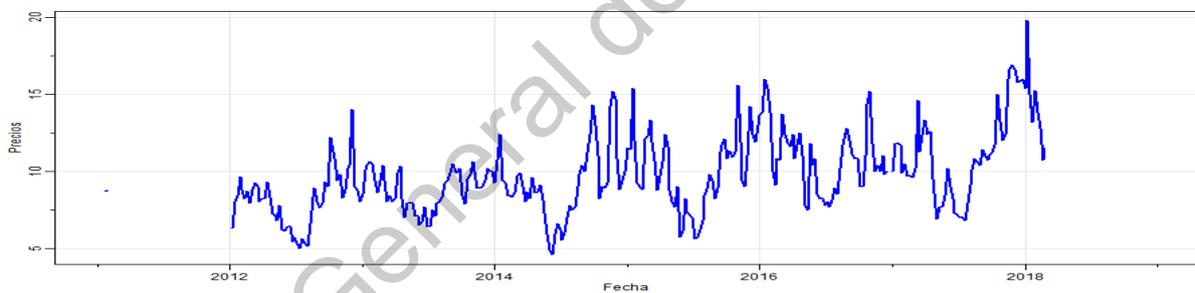
*Figura C.51: Precios de la manzana golden delicious*



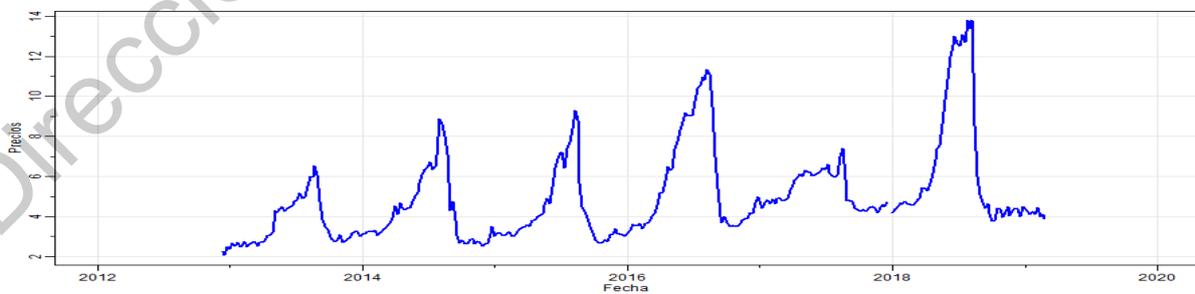
*Figura C.52: Precios de la manzana red delicious*



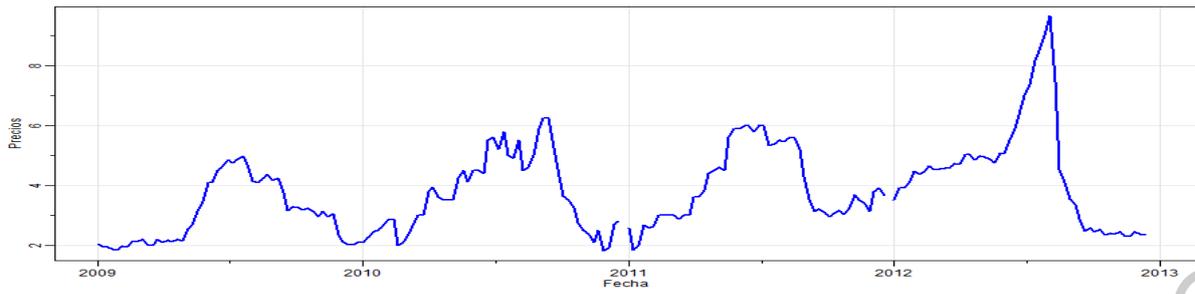
*Figura C.53: Precios del melón Cantaloupe #27*



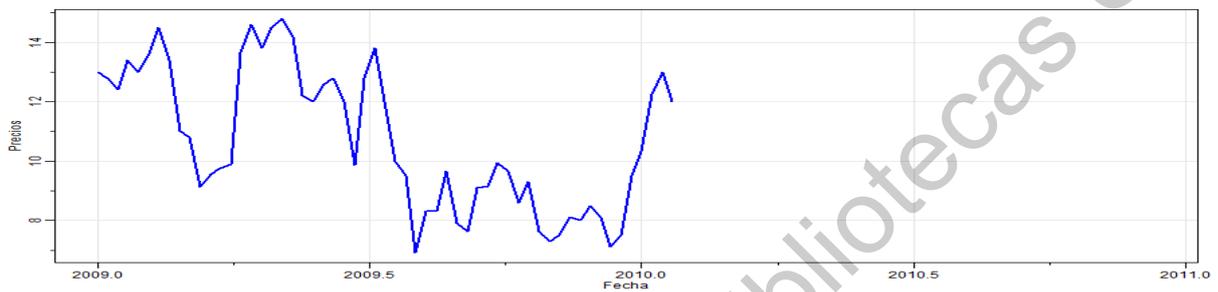
*Figura C.54: Precios del melón Cantaloupe sin clasificación*



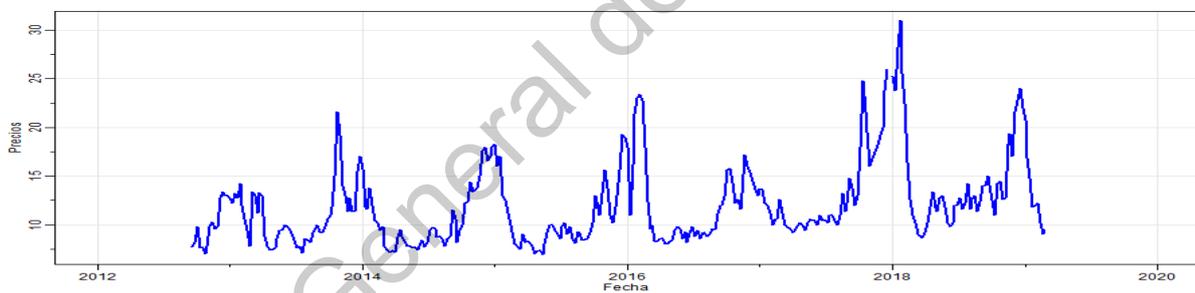
*Figura C.55: Precios de la naranja Valencia grande*



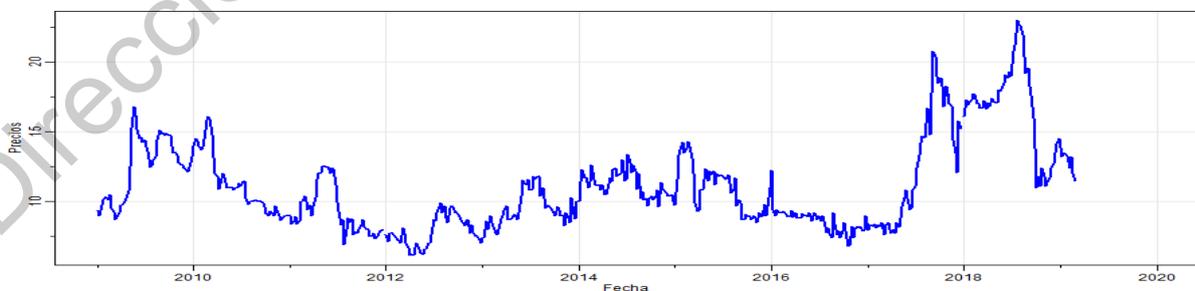
*Figura C.56: Precios de la naranja Valencia mediana*



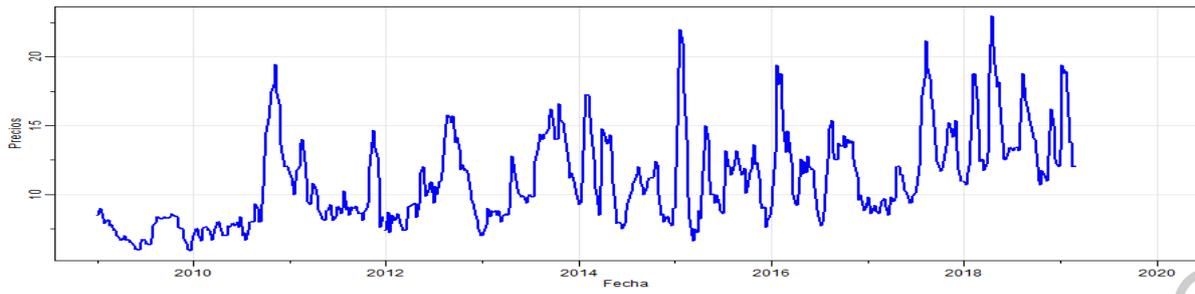
*Figura C.57: Precios del nopal*



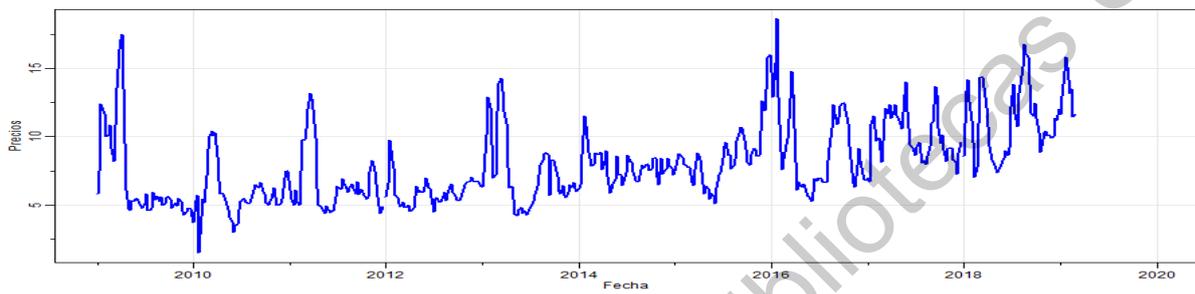
*Figura C.58: Precios del nopal grande*



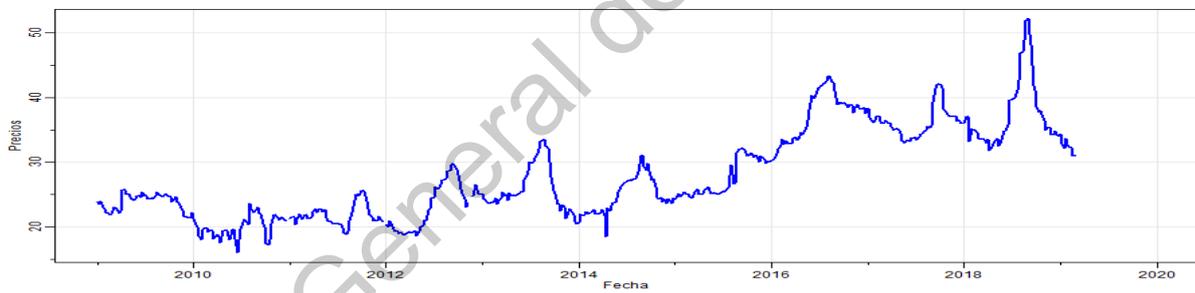
*Figura C.59: Precios de la papa alpha*



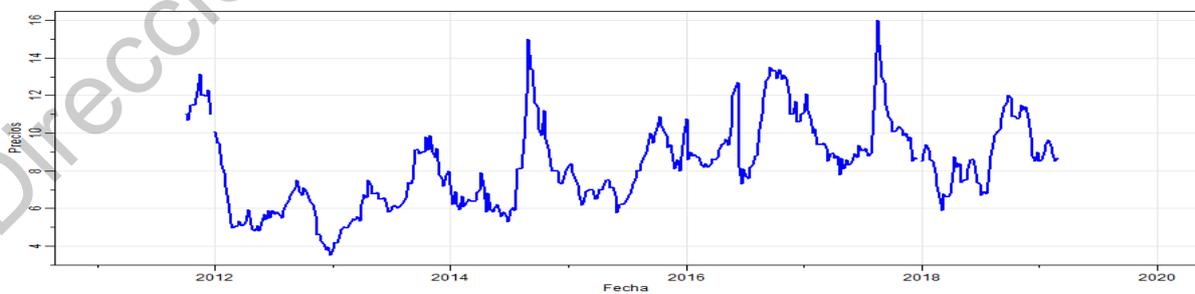
*Figura C.60: Precios de la papaya maradol*



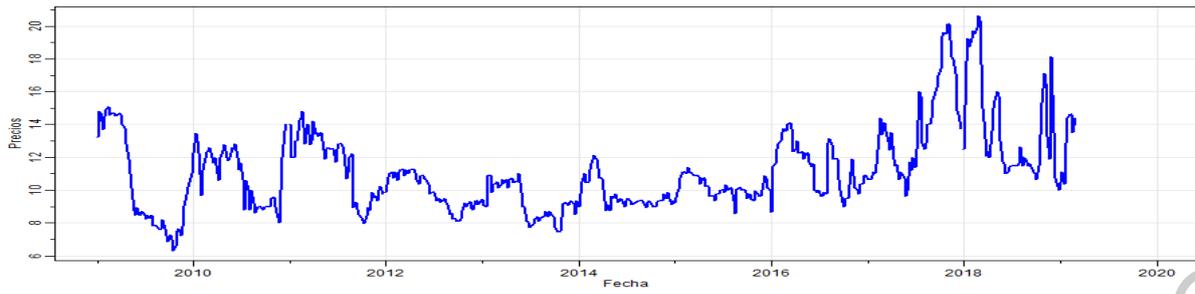
*Figura C.61: Precios del pepino*



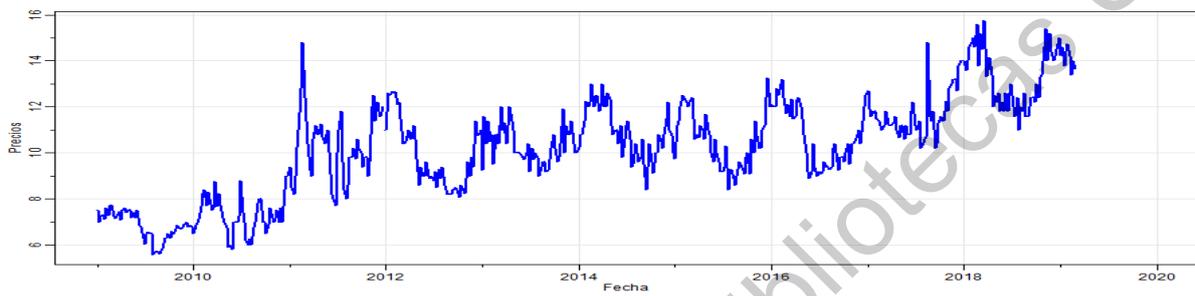
*Figura C.62: Precios de la Pera D´anjou #100*



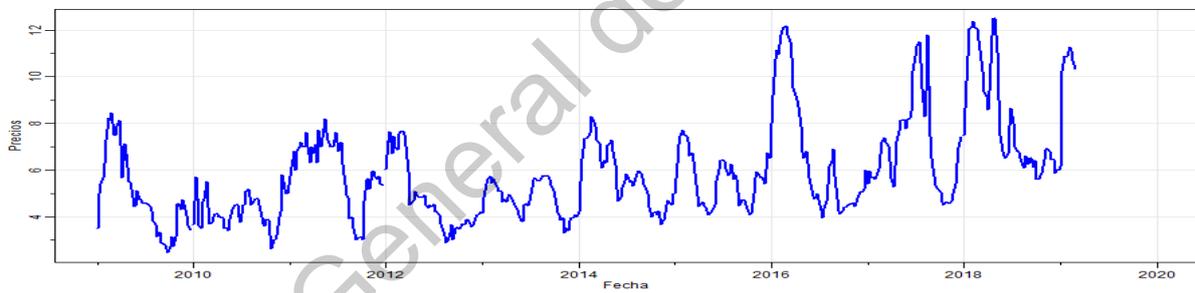
*Figura C.63: Precios de la piña mediana*



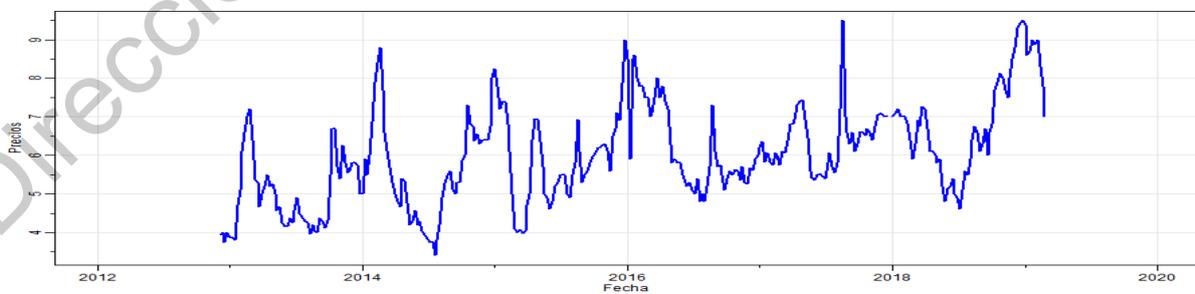
*Figura C.64: Precios del plátano dominico*



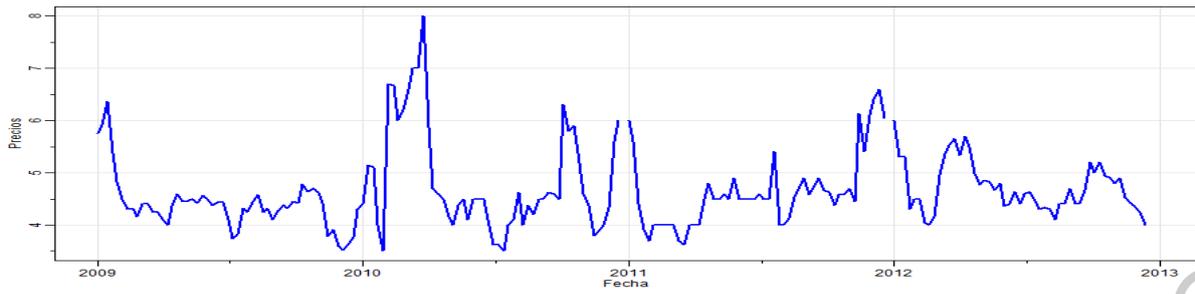
*Figura C.65: Precios del plátano macho*



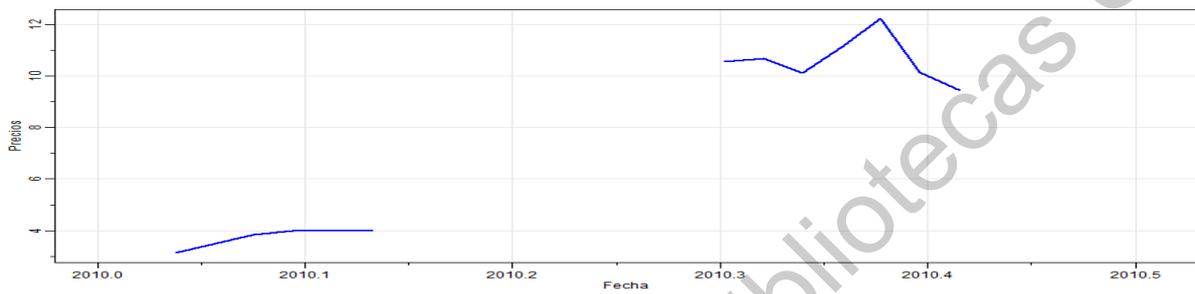
*Figura C.66: Precios del plátano Tabasco*



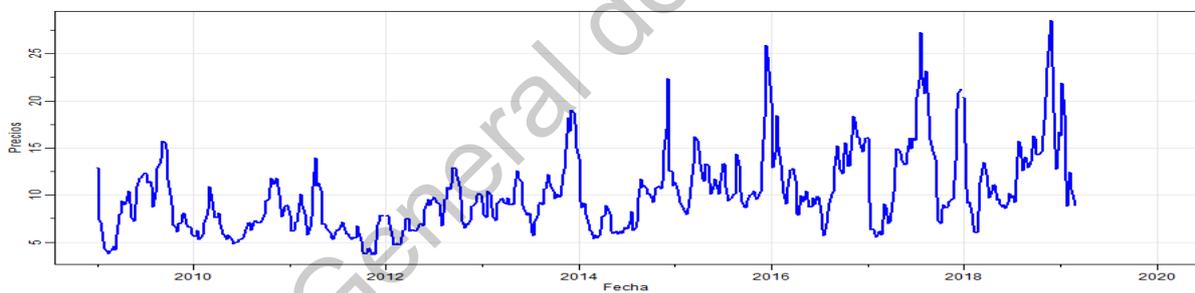
*Figura C.67: Precios de la sandía rayada*



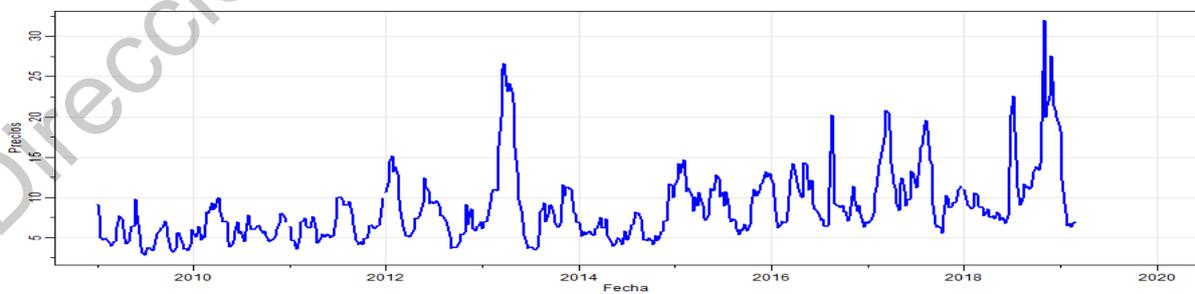
*Figura C.68: Precios de la sandía sangría*



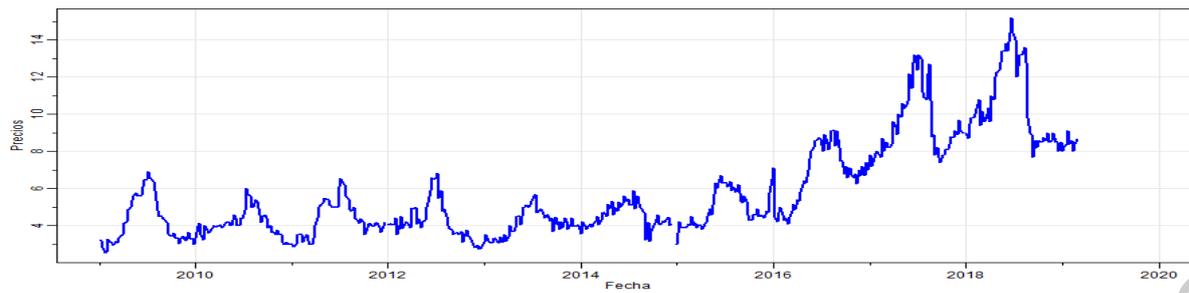
*Figura C.69: Precios del tejocote*



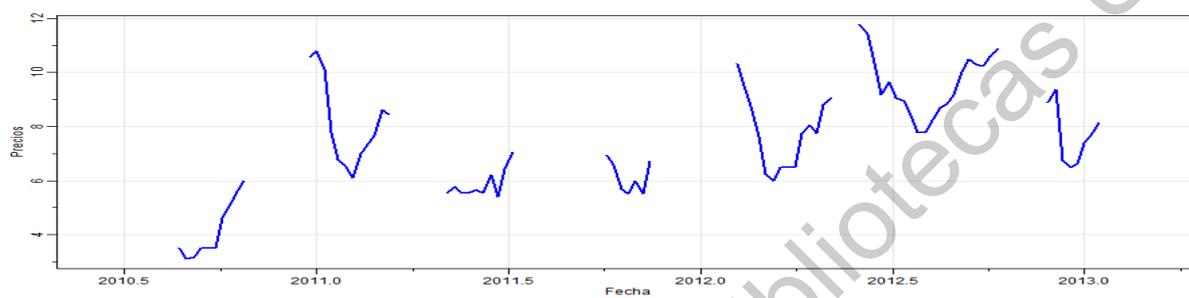
*Figura C.70: Precios del tomate saladette*



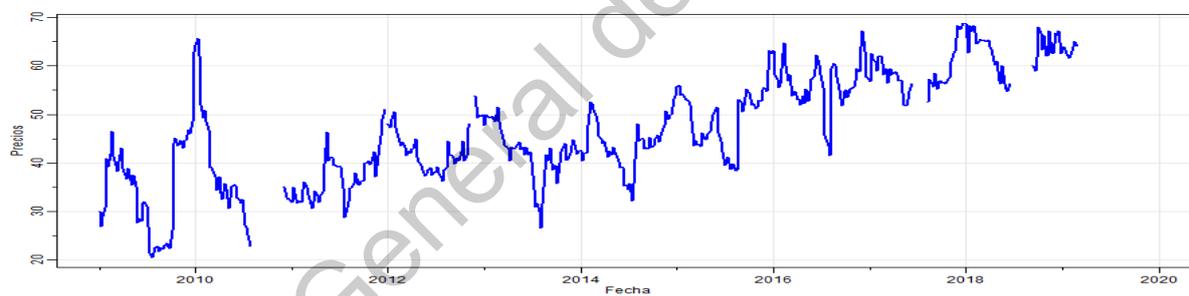
*Figura C.71: Precios del tomate verde*



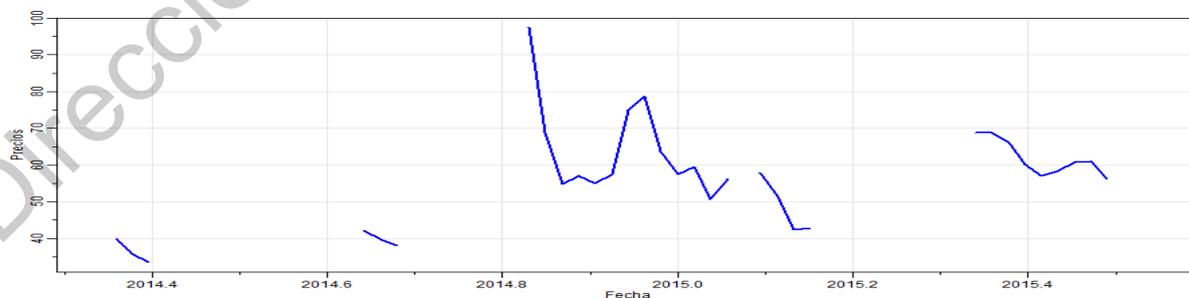
*Figura C.72: Precios de la toronja roja*



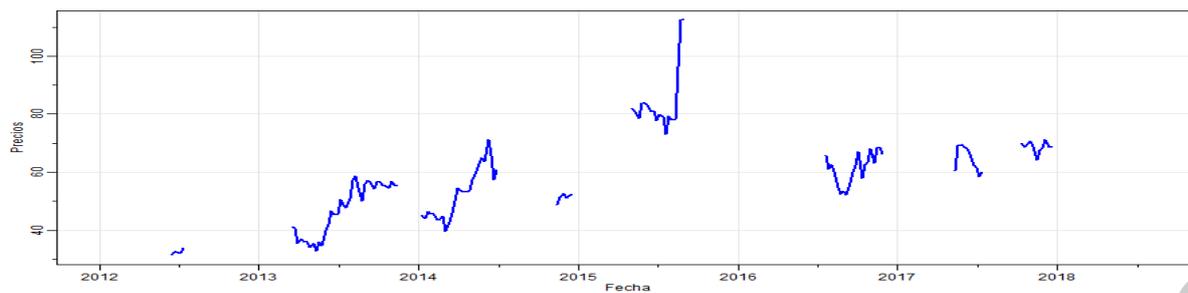
*Figura C.73: Precios de la tuna blanca*



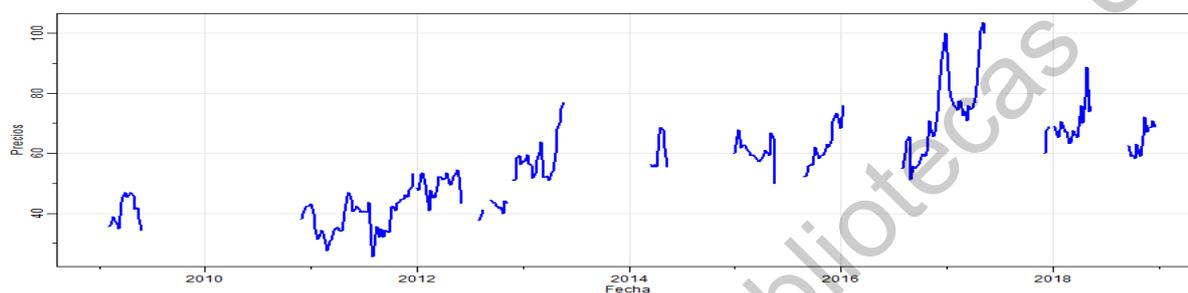
*Figura C.74: Precios de la uva globo*



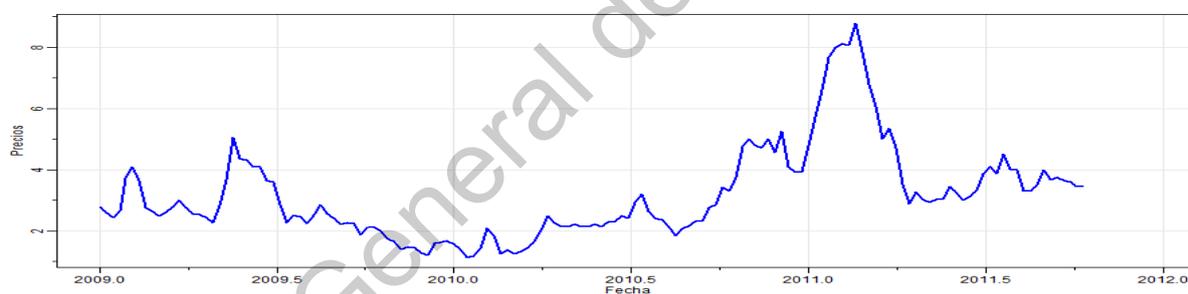
*Figura C.75: Precios de la uva Perlette*



*Figura C.76: Precios de la uva superior*



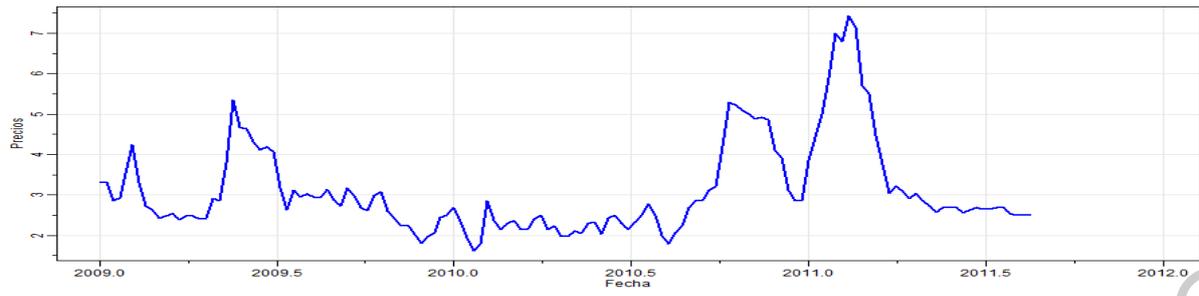
*Figura C.77: Precios de la uva Thompson*



*Figura C.78: Precios de la zanahoria leña*



*Figura C.79: Precios de la zanahoria mediana*



*Figura C.80: Precios de la zanahoria polvo*

Dirección General de Bibliotecas UAQ

## APÉNDICE D

### Tablas: Series de tiempo semanales

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	23	22.4	23.2	22	21.75	19.8	21.6	22	22	23	23.75	24.4	23.8	25	25.8	28.2	28	28.33	28.6	28.4	28	27.1	27	27.4	28	28.6	27.4
2016	23.75	21.4	19	20.2	20.75	20.6	19.8	19.2	19.2	18.6	19.2	18.83	19.75	19.2	20.2	20.6	21	22.75	25.2	28.2	35.6	38.8	38.4	38.8	41.6	54	56.2
2017	28.6	30	31	32.2	32	31.5	30.8	31	33.2	37.6	41	43.75	44.8	46.8	44.67	47.2	52.4	52.33	53.2	55.6	56.2	58	58.6	57.6	56.8	56.8	58
2018	30.5	32.6	34	34	34	34	33.3	32	33.5	32.8	34.8	34.5	35	35.4	37.4	38.2	39.2	38.5	39.6	38.6	39.4	39.4	39.6	38.8	39.2	39.2	39.6
2019	36.5	32.6	29.4	29.5	30.5	30.6	31	30.2																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	28.6	29.5	29.6	30.2	26.6	28.6	29.6	25.9	24.9	24.5	23.8	24.2	22.6	21.6	20.4	19	19.25	19.4	19.25	18.8	17.9	19	19.8	20.4	24	25	
2016	52	46	42.2	38	36.4	38	35	34.4	37.6	39.4	38.75	38	37.6	39.4	37.3	37	37	37.2	35.6	36.5	33.2	33.4	31.6	33.6	29.4	28	
2017	58.2	58.6	57.8	55.2	52.6	57.8	63.2	56.6	54.6	54.8	53.4	50	49.6	48.6	45	43	39	40	40.6	38.5	36.8	34.8	34	34	32.5	32.93	
2018	39.8	41	41.6	41.6	43	44.4	46	45.8	46.2	43.8	41.4	39.4	40.2	37.6	36.2	35.8	35.5	37	38	36.25	35.8	35	35.8	36.4	37.5	35	
2019																											

*Tabla D.1: Precios semanales del aguacate Hass*

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	28.5	30.4	32.4	33	32.25	31.8	31.6	31.4	30.4	30.8	30.5	29.6	30.6	30.67	30	30.8	30.8	30	29.8	31.6	31.8	33.2	33.8	33.6	34.4	33.6	32.8
2016	32.5	33.8	35	35	35.75	35.4	36.6	36.2	36.2	36.8	36.6	35.67	35.5	35.2	35.2	34.4	33.8	34	35.6	35.6	36.2	36.2	35.8	35.8	35.2	37.2	37.8
2017	38.8	37	38.8	38.8	38.4	38.75	37.8	39	38.6	38.1	38.8	38.25	42.8	42.6	43.33	43.4	43.4	43.83	43	43.8	44	45	45	47	46.8	47.8	48.6
2018	53	52.8	52	53.4	53.8	54	55.4	55	55.8	53	44.2	53.75	59	59	60	60.4	61.6	61	60	60	61.2	62.4	62.8	63.4	63.8	62.4	63.6
2019	66.75	65.2	65.2	66.4	65.5	65.8	66.8	67.6																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	34.6	36.8	36	37	36	35.4	36.8	37.4	38.2	43.75	45.6	45.4	45.8	43.2	42	41.8	40	40.6	39	39.2	38.4	35.8	35.8	34.4	34.25	34	
2016	38.7	38.6	39.6	39	39.4	39.2	38	36.8	38	39	38.75	39.4	39.2	39.4	39.6	39.6	39	39.2	39	39	38.2	38.4	38.2	37.2	38	39	
2017	48.2	49	48.6	47.8	47.2	47.6	53.6	48	46.2	46.6	48	48.8	47	51	50	57.6	56.25	55	55	55.75	54.6	52	51.6	53.8	53	52.7	
2018	63.6	63.2	62.8	62.8	63	62	62.2	63.8	64.2	64.8	65	66	65.8	67.2	71.6	73.8	71.75	71.8	70.4	72.75	69.6	68	68	67.8	66.75	66	
2019																											

*Tabla D.2: Precios semanales del cacahuete*

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	7.8	11.84	11.12	9.04	6.9	6.4	8	6.96	5.68	6.56	9.7	10.72	7.04	4.8	4.56	4.32	4	5.33	4.56	4.72	6.48	8.04	7.6	6.24	6.2	9.36	8.96
2016	8.3	14.16	23.92	22.56	19.6	13.68	13.52	10.08	9.52	11.92	14.96	8.27	7	9.12	7.8	6.08	7.92	7.3	5.04	4.92	6.64	5.28	9.6	12.96	9.44	9.04	9.28
2017	6.56	12	13.2	8.48	9.12	7.6	7.44	11.52	11.36	10.8	13.52	13	12.64	10.88	10.4	6.56	10.12	9.6	6.96	6.16	5.28	7.04	9.2	8.48	8.64	8	11.68
2018	22	15.68	17.36	15.68	10.68	8.3	6.8	7.68	6.24	7.04	5.88	5.6	5.53	5.36	5.12	5.28	5.12	5.2	6.16	7.44	5.28	7.84	8.72	7.96	11.44	12.56	9.76
2019	9.8	14.8	17.12	11.76	9.6	9.04	8.72	9.12																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	9.68	7.64	6.76	6.2	5.92	5.32	4.76	7.28	6.08	7.7	8.08	7.24	8.4	7.84	9.44	8.24	5.6	6.48	6.7	5.92	8.96	11.28	10.96	9.48	8.3	8	
2016	8.4	10.56	10.16	11.52	11.88	10.8	11.68	9.76	9.28	9.04	8.3	9.2	10.72	11.04	10.64	9.36	8.7	8.16	9.44	9.4	8.64	5.64	6.72	7.2	6.56	6.8	
2017	11.2	10.4	10.8	10.88	11.12	13.68	11.52	7.84	10.32	11.2	10.64	6.24	12.08	13.12	10.4	11.6	11.7	8.64	12.72	13.1	17.08	17.44	19.12	24	23.7	20.31	
2018	8.08	7.36	6.56	6.56	7.2	7.6	7.2	6.32	6.64	7.12	6.64	6.08	6.24	6.88	7.12	12.88	11.4	11.44	12.48	18.6	15.68	14.8	11.68	17.04	17.9	12	
2019																											

Tabla D.3: Precios semanales de la calabacita italiana

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	5.13	6.11	7.47	8.42	9.47	9.95	8.74	7.79	5.26	4.68	6.58	9.79	6.53	4.91	4.32	4.42	4.74	6.4	5.32	5.11	7.47	11.37	7.68	6.95	5.05	4.58	4.32
2016	6.58	5.42	9.05	12.11	12.76	11.68	11.79	10.21	6.84	9.58	11.68	10.7	10.53	9.58	7.21	9.79	8.84	8.82	7.47	9.16	8.21	7.47	11.16	12.74	6.74	7.68	7.89
2017	5.37	6	10	8.95	7.68	6.84	5.68	6.53	11.05	9.47	7.37	8.03	9.89	13.47	12.11	8.42	5.79	9.91	10.95	9.58	5.68	5.58	11.79	11.58	11.16	9.63	10.95
2018	11.05	10.53	15.89	20.42	17.16	11.71	9.05	8.53	7.26	7.05	7.26	5.92	5.61	4.84	5.26	6.21	6.42	6.05	8	8.84	7.37	6.63	10.21	7.26	5.05	5.05	7.89
2019	4.74	9.58	11.05	11.05	15.13	16.42	12.74	10.42																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	4.32	3.84	4.58	5.63	6.16	5.68	6	4.74	4.05	4.8	5.47	5.89	5	6.21	5.05	4.63	4.5	4.05	5.86	6.89	6.11	5.58	5	5.58	6.51	6.32	
2016	4.74	7.26	5.58	5.47	7.26	6.95	6.32	8.79	8.21	5.05	5.53	5.68	6.42	6.53	5.37	7.58	6.05	6.21	6.95	7.76	5.79	5.26	5.47	6.11	5.37	5.79	
2017	10.95	8.11	6.21	5.79	5.26	5.89	6.58	6.95	6	6.95	7.32	6.74	6.63	6.32	6.42	10.84	11.84	8.74	6.42	9.47	11.37	10.11	9.68	10.74	11.32	11.54	
2018	11.89	11.05	9.79	8.74	7.05	7.58	8.84	12.32	10.63	6.53	6	4.95	5.79	6.32	5.05	5.58	6.32	6.42	6.84	11.32	8.53	6.11	5.89	6.11	6.18	4.74	
2019																											

Tabla D.4: Precios semanales del chayote sin espinas

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	8.75	7.2	6.6	6.7	7.25	8.8	7.5	7.3	7.2	9.8	12.5	16.6	15.8	15	14.4	15.9	16.9	17.17	18.8	19.4	19	16.5	14	14.4	14.6	16.2	14
2016	12.62	12	14.1	20.5	22.25	26.3	29.2	28.6	21.8	19.8	20.4	18	16	13.3	11.1	9.8	10.8	11.25	12	14	12.9	10.8	10.6	11.3	11.1	10.2	10.1
2017	7.9	8.6	8.3	8.7	9.8	12.12	14.8	13.7	14.8	17.4	17.3	15.88	17.1	16.2	17.33	23	22.2	20.67	21	21.4	21	21.4	19.6	17.5	18.2	24.6	24.7
2018	11.62	8.6	7.7	7.4	7.3	7.75	10.5	8.8	8.5	8.3	9.6	10.25	11.5	11.5	13.2	14.7	16.4	16	12	10.4	8.5	8	7.1	7.5	10.1	12	12.9
2019	55.5	81.4	83.4	54.8	32.5	37.2	39.6	31.6																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	13.3	14.7	16.8	16.4	15.5	14.5	13.7	16.8	17.4	16.5	15.9	16.1	13.6	16.6	18.1	21.2	20	16	14	15.1	13.7	12.3	11.9	11.7	11.75	12	
2016	10.2	11	10.6	12	13.6	16.4	17.2	16.6	13.3	12.3	11.88	12.1	11.8	12.8	15.9	19.4	18.5	19.5	19	15.25	10.9	10.1	8.8	8.5	8.2	10	
2017	26.5	26.1	26.3	21.6	16.6	18.8	20.4	18.1	19.8	18.5	16.8	18.4	17.3	18.6	23.6	24.1	25.25	25.9	24.6	21.75	14	12.9	13.2	12.8	13.5	11.24	
2018	13.9	13.4	14.5	14	15.9	15.1	16.2	16.6	16.6	13.8	14.4	13.7	15.4	16.4	18.6	22.8	23	20.8	20.6	22.5	25	25.8	46	42.4	41.25	65	
2019																											

Tabla D.5: Precios semanales del chile serrano

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	9.75	9.6	9.6	11.2	10.75	11.2	11.2	11.2	11	11	9.75	9.6	9.8	9.33	9.4	9.2	8.6	9.33	8.4	8.6	9.6	10.4	10	9.8	10.2	9.6	9.4
2016	11.75	11.1	9.6	10.2	10	9.8	10.6	11.6	10.4	10	10	10.33	9.75	9	9.8	9.8	11.2	11.25	11.2	10	10.6	11	10.6	11.4	10.4	10	10.6
2017	10.9	10.5	10.8	11.8	11.2	11.5	12	11.8	11.2	11.2	10.4	11	11.4	11	11	10.8	10	9.33	10.8	10.4	11.2	11.8	11.8	11	11	11.6	11.4
2018	15	12.6	13.6	14.4	13.4	11.75	11.4	11.6	10.6	11	11.4	11.25	11.67	12.2	12	12	12	12.25	12.4	12	12.8	12.8	11.2	12.4	12.6	11.2	11.6
2019	13.5	12.6	12.8	12.8	12.5	13	12.8	12.8																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	10.2	10.2	9.9	10.2	10.2	10.9	10.8	11.2	11.2	11.5	10.8	12.4	12.6	12.4	11.6	11.2	11	10.9	10.5	9.8	9.2	9.4	9	8.4	10.75	12	
2016	10.8	10.8	10.8	9.8	10.2	10.8	10.6	11.4	11.2	10.1	10.75	11.3	12	12.4	11.6	11.4	10	11.8	11.2	12.25	11.4	11.4	11.6	11.9	10.7	11	
2017	12	12.6	12.6	12.1	11.8	13.3	14	12.2	11.6	13.8	14	15.2	14.2	13.8	13.6	15	14	13.6	12	13.75	15.4	16	16.8	14.4	14.25	14.44	
2018	11.6	12	12	12	12.2	12.4	12.4	12.4	12.8	13	13	12.6	13.8	14.4	14.2	13.2	13.5	13.6	12.8	13.25	14	12.4	13	13	13.5	15	
2019																											

Tabla D.6: Precios semanales de la coliflor grande

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	67	57.6	50	53.6	52.5	49	50	48	47	43	41.25	43	41	40	43	43	42	45	42.4	44	48	50.6	54	51	52.2	49	53
2016	71.25	56	61	56	60	62	61	58	52	47	50	46.67	52.5	48	50	52	51	53.75	55	51	54	53	52	52	68	72	56
2017	59	53	52	58	55	60	63	61	58	47	52	48.75	44	45	43.33	43.6	47	45	44	49	41	42	46	44	55	74	85
2018	52.5	48	51	51	49	42.5	50	52	50	45	50	47.5	51.67	46	47	41	42	45	45	44	43	44	43	49	63	60	63
2019	72.5	76	68	67	67.5	66	64	62																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	58	50.2	45.8	55	46.8	41.6	41.3	46.4	49	48.75	54	55.8	54	51	47	48	46.25	50	50	50	50	55	72	82	73.75	75	
2016	63	49	44	47	47.6	43.4	58	56.2	56	53	53.75	52	45	46	45	44	42.5	43	50	45	46	44	49	56	58	60	
2017	83	78	67	62	67	62.8	59.4	53	50	56	65	69	71	69	72	63	57.5	48	49	50	47	46	51	45	50	49.58	
2018	64	61	53	61	63	61	58	58	59	57	62	54	52	53	58	52	57.5	61	60	57.5	63	61	66	63	65	90	
2019																											

Tabla D.7: Precios semanales del epazote

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	5	5.9	6.2	6	7	6.8	6.8	7.2	6.9	6.8	6.75	6.8	7	7	6	6.6	6.2	7	6	5.8	6.2	6	5.8	6.1	6.7	7.1	6.9
2016	6	5.72	6.2	5.9	6	6.4	5.6	5.7	5.8	5.7	5.3	5.5	5.25	4.7	4.8	4.5	4.6	4.25	4.3	4.3	4.3	4.7	4.5	4.4	4.6	5.2	5
2017	4.4	4	4	3.9	4.1	4.12	4.4	4.1	4.2	4.6	4.6	4.25	4.5	4.3	4.33	4	3.9	4.33	4	4.3	4.3	4.3	4.3	3.8	4.3	4.5	4.6
2018	6	6	5.7	6.6	6.6	6.62	6.8	6.7	6.9	6.8	7	6.75	7	7	7	6.8	7	7	6.9	7	7.1	6.8	7	7	7.1	7.8	10.8
2019	7.5	7.8	7.8	8.6	10.25	8.9	8.8	8.6																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	7.2	7.4	7.2	7.4	7.1	6.3	6.4	6.8	6	5.75	5.8	5.8	5.1	5.2	5.2	5.2	5	5.2	5.12	5.2	5.3	5.6	5.2	5.1	4.75	5.5	
2016	6.1	6	6	9.2	8.6	9.1	6.8	6.6	5.5	6	5.25	5.2	4.5	4.7	4.6	4.5	4.38	4.4	4.5	4.75	4	4	4.2	3.7	4.2	4.5	
2017	4.5	5.1	8.4	12.6	10.6	11.2	12.2	10	7.7	6.8	6.8	6.5	6.1	6.2	6.3	6.4	6.12	5.7	5.4	5.38	5.7	5.5	5.5	5.5	6	5.78	
2018	10.4	11.8	13.4	12.8	12	10.4	10.2	8	8.6	8.6	7.4	7	7	7.2	7.2	7	7.62	7.4	7.3	7.5	7.7	7.1	7.2	7.5	7.88	8	
2019																											

Tabla D.8: Precios semanales de la jícama

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	48.75	48.4	48.8	47.8	47.75	48	48	48	48	48	48.25	48.6	50.8	50	50	49.6	48.6	45.33	45.6	43.4	43	40.4	42	41.4	38.2	36.6	37.4
2016	48	46.8	50.4	53.8	54.25	57	56.8	59.4	64.4	65.6	65.2	66	69	68.4	65	53.6	44.2	39	37.6	37	34	32.4	32	33.2	33.2	40.6	43.8
2017	48	49.6	50	49.6	49.2	50	50	51.6	51.4	52.6	52.4	52.75	51.8	51.2	51.33	50.2	50.6	49.33	49.2	49	48.8	46.4	50	52.8	47.8	50.8	50.4
2018	57	57.4	58	57.8	57.4	57.5	57.6	57.6	57.4	56.4	56.4	56.5	56	56.6	56	55.6	56	54.25	54.8	51.2	51.2	50.4	51.2	51.6	50.4	50.8	50.8
2019	50	49.6	52.6	58.8	59	56.6	57.4	58.4																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	37	38	38.4	39.2	39.2	39.5	39.8	40.6	40.4	39.5	39.2	36.8	36.2	35	34.6	33.8	33.75	37.1	41	43.4	48	49.4	48.2	48.4	49.5	50	
2016	44.8	46.4	47.6	47.6	46.4	44.2	43	37	31.2	30.2	32	32.6	30.8	33.8	42	46.8	47.5	47.8	46.6	48.75	50	50.4	51.4	50.6	48	48	
2017	47.8	48	47.4	48.6	48.6	51.4	52.2	49	53.6	55	55	57	58.2	63.6	64.8	65.4	65	64	62.4	59.5	59.6	56.6	57.8	57.6	57	57.47	
2018	50.4	49.6	48.8	49.6	50.2	49.2	48.4	48	48	48.4	47	46.2	45.2	45.4	45.4	45.6	47.5	52	50.8	50	50	50.4	49.2	50.4	52.5	50	
2019																											

Tabla D.9: Precios semanales del kiwi

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	6.75	7.5	7.3	7.6	7.38	7.4	7	7	7.1	6.8	6.75	6.8	6.7	6.83	6.7	7	7	7.33	7.2	7	8	11.2	8.2	9.4	9.6	10.6	11
2016	8.62	6.9	7.2	7.2	7	7	7	7	7	7	7	7	9	9.8	9.8	9.4	10.25	12.8	16	15	15.2	14.8	16	14.8	14.4	14.2	
2017	10.2	9.6	8.6	9.8	9	8.75	10	10.2	9.8	11	9.4	10	10	11.8	11.33	14.5	14.5	15.33	17	17.4	19.4	19.2	20.8	20.2	20.4	21	20
2018	14	13.6	13.6	13.2	13	12.5	12	12.6	12.4	12.4	12.6	12.25	13	13	12.8	11.8	12.2	13.25	13.4	13.4	13.2	15.6	16	17	19.4	20.2	20.6
2019	10.5	9.8	10	9.8	9.38	9.4	10.4	9.4																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	12.8	14	12.5	13.2	12.4	12.4	12.5	12.8	10.7	9.75	9.4	8.5	7.8	7.6	7.4	6.8	6.88	6.4	7.75	6.8	7	7.6	7.9	7.5	8.12	8.5	
2016	14.2	14.8	17.2	16	15.8	15.2	15.4	14.1	12.8	10.5	9	9	8.6	8.8	7.8	8	8.12	8.3	7.6	7.5	8.2	8	8	8.7	9.4	10	
2017	20	19.8	19.4	18.6	17	17.2	17	12.8	11.4	10.2	12	10.2	10.6	11	10.4	10.8	12	11.8	13	11.5	13.4	12.8	13	13.2	14	13.57	
2018	19	17.4	17.8	17.6	18	17.4	16.2	14.6	12.2	12.4	13	11.4	11.2	10.4	11	9.6	11.25	11.4	9.2	9.25	9.2	9.4	9	9.2	9.5	10	
2019																											

Tabla D.10: Precios semanales de la lima

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	7	5.25	5.2	5.7	6	6.8	6.5	6.25	6.3	7.1	6.94	6.6	5.95	5.5	5.65	5.5	5.3	5.92	5.7	5.45	5.1	4.95	4.6	4.8	4.95	4.65	5.2
2016	5.12	5.05	5.4	5.4	6.31	6.85	8.1	9.4	10.1	10.2	12.1	13.5	13.5	17.45	18.9	21.9	19	18.31	16.25	12.4	9	7.3	7	6.75	6.15	6.65	6.7
2017	12	10.85	10.5	12.4	13.55	13.5	14.9	17.1	20.9	20.5	20.7	15.88	17.3	16.3	16.17	15.1	10.35	7.25	6.25	5.5	5.9	5.9	6.15	6.55	7.2	7.4	
2018	11.25	8.8	10.2	10.7	11.4	10.94	13.3	15.75	19.9	22.55	18.75	18.69	23	22.2	15.6	9.1	8.15	6.12	6.45	5.65	5.45	5.9	6.15	7	6.2	5.45	6.4
2019	7.88	6.25	7.5	7.05	7.88	8.6	10.45	11.1																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	5.05	5.6	6.4	6.55	8.05	7.4	7.25	8	7.05	7.44	7.9	7.55	7.9	7.6	7.25	7.95	6.88	5.9	5	4.6	4.6	5	5.45	5.4	5.06	5.25	
2016	6.47	6.55	6.95	7.35	8.2	8.45	9.1	9.1	10.25	9.6	8.88	7.95	7.7	6.9	5.65	5.4	5.25	6.2	5.85	5.44	5.5	7.6	7.3	11.8	12.2	12.5	
2017	7.15	6.85	7.45	7.65	8.1	8.05	8.2	10.65	11.95	9.9	9.35	9.85	9.1	10.45	10	9.9	11.44	11.4	10.65	9	8.55	7.05	6.9	11	11.12	9.88	
2018	7.05	7.2	7.75	9.65	11.05	11.75	12.25	16.35	12.05	10.9	11.1	9.7	8.2	8.75	8.3	7.85	7.81	7.2	7.3	7.19	7	7.5	8.25	8	8	7.5	
2019																											

Tabla D.11: Precios semanales del limon #5

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	3	3.16	3.2	3.04	3.05	3.22	3.24	3	3.06	3.26	3.38	3.46	3.5	3.6	3.52	3.8	3.8	4	4.06	4.16	4.56	4.92	4.66	5.46	6.44	6.88	7.18
2016	3.15	3.32	3.62	3.54	3.55	3.66	3.38	3.64	3.64	3.84	4	4.13	4.58	5.16	5.2	5.76	6.5	6.3	6.38	7.36	7.74	8.3	8.74	9.2	9.04	9.04	9.02
2017	4.8	4.42	4.68	4.82	4.64	4.88	4.64	4.9	4.9	4.8	4.84	4.9	5.34	5.58	5.8	5.94	6.12	6	6.32	6.24	6.2	6.04	6.08	6.16	6.24	6.44	6.34
2018	4.2	4.38	4.5	4.68	4.62	4.78	4.64	4.62	4.58	4.68	4.86	4.9	5.43	5.4	5.3	5.66	6.14	6.85	7.38	7.56	8.62	9.84	10.84	11.98	12.56	13	12.68
2019	4.38	4.32	4.12	4.24	4.12	4.46	4	4.1																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	7.2	6.44	7.4	7.7	8.4	9.3	8.8	5.66	4.5	4.25	3.92	3.5	3.2	2.84	2.72	2.68	2.73	2.82	2.75	3.1	3.12	3.38	3.18	3.14	3.1	3	
2016	9.74	10.4	10.56	10.98	10.8	11.34	11.1	9.5	7	5.4	4	3.7	4	3.7	3.5	3.56	3.5	3.56	3.64	3.8	3.9	3.9	4.2	4.14	4.68	5	
2017	6.62	6.2	6.04	5.96	6.08	7	7.4	5.86	4.78	4.8	4.76	4.4	4.34	4.26	4.34	4.26	4.3	4.46	4.48	4.4	4.2	4.44	4.62	4.5	4.7	4.48	
2018	12.52	13.1	12.7	13.8	13.4	13.8	8.54	6	5.1	4.7	4.4	4.64	4.14	3.78	3.8	4.4	4.38	3.98	4.24	4.5	4.02	4.26	4.4	4.36	4	4.5	
2019																											

Tabla D.12: Precios semanales de la naranja Valencia grande

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	18.25	16	17	13	12.5	11.2	10.6	9.3	8	7.8	7.5	9	8.2	8.33	8.2	7.8	7	7.33	7.3	6.9	8	9.2	10	10	9.4	9	8.6
2016	18	11	16.2	21.4	23	23.4	22.8	18	12.4	9.6	10	8.33	8.25	8.6	8.6	8	8	8.25	8.4	9.4	9.8	9.6	8.6	9.3	8.2	9.1	9.8
2017	13.7	13.6	12.2	12	11.2	10	10.6	10.6	12.6	11.2	10	9.75	9.6	9.2	9.33	9.8	10.2	10	9.4	10.2	10.6	10.4	10.4	10	11	10.4	10.4
2018	25.25	23.8	28.2	31	25.6	22.5	16.4	12.6	11	10.2	9	8.88	8.67	9.2	10.2	11.8	13.4	11.75	11.4	12.8	13	11.8	10.2	9.8	10.2	12	12
2019	20.75	16.9	14.8	11.8	12	12.2	10.2	9																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	9.8	10.2	9	9.8	8.6	8.1	9.3	9.2	8.4	8.5	8.6	9.2	10.2	13	11.7	11	13.25	15.6	13.75	11	10.2	11.8	13	15.6	19.25	19	
2016	8.8	9.4	8.6	9.2	9.2	8.8	9	9.6	9.5	11.1	12	12	12.9	15.6	15.8	14.3	12.25	12.6	11.6	13.25	17.2	16	15.4	14.5	13.6	13	
2017	10.2	11	11	10.4	10	11	13.2	11.4	12.2	14.8	13.8	12	13	18	24.8	21.6	19.5	16	16.8	17.5	18.2	19.2	20	23.8	26	24.51	
2018	12.8	11.6	12.2	14.2	11.6	12.6	13	11.4	12.4	14	14	15	14.1	12.5	11	14.2	14.5	12.6	12.8	18.25	19.4	17	21.6	22.8	24	22	
2019																											

Tabla D.13: Precios semanales del nopal grande

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	8.25	8.4	7.9	7.6	7.25	6.7	6.16	6.3	6.9	7	7	7	6.5	6.5	7.04	7	7.2	7.5	7.5	7.1	7.16	6.7	5.78	6.2	6.2	6.2	6.4
2016	10.75	8.6	9	8.8	8.88	8.7	8.7	8.3	8.2	8.4	8.2	8.33	8.62	8.6	8.7	9.1	9.4	9.5	9.7	9.36	10.3	12	12.4	12.7	8.2	7.3	8.1
2017	11.1	12.1	11.2	10.8	10	10.25	9.4	9.4	9.4	9.5	9.2	8.5	8.9	8.9	8.67	8.8	7.8	8.67	8.1	8.6	8.3	8.4	8.5	8.9	8.7	9.36	9.06
2018	8.5	8.9	9.4	9.1	8.6	8.5	7.9	7.2	6.6	5.9	6.8	6.62	6.67	6.8	7.9	8.76	8.2	8.38	7.4	7.5	7.5	8.1	8.6	8.6	8.3	7.4	7.3
2019	8.5	8.6	9	9.4	9.62	9.3	8.9	8.5																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	6.7	6.8	7.3	7.6	8.02	8	8.4	8.7	9	9.38	10	9.5	9.9	10.3	10.9	10.4	10.12	9.9	9.25	9.4	8.5	8.1	8.6	8	8.75	10	
2016	7.7	7.6	8.2	8.3	8.8	8.8	10.2	11	11.9	12.8	13	13.5	13.3	13.3	12.9	13.4	12.88	13.1	12.9	12.12	11	11	11.7	10.6	10.6	11	
2017	9.1	9.2	8.8	8.9	9.9	12.56	16	14.7	13	12.7	11.5	10.9	10.7	10.1	10.1	10.3	10.38	10.2	9.9	10	9.6	9.8	8.5	8.7	8.62	8.77	
2018	6.7	6.9	6.8	7.4	8.8	9.5	9.9	10.1	10.2	10.8	11.4	11.6	12	11.9	10.9	10.9	10.75	10.8	11.5	11.25	11.4	10.8	9.8	8.8	8.5	9	
2019																											

Tabla D.14: Precios semanales de la piña mediana

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	9.75	11.2	11.4	12	12.5	12.4	12.2	12	12.2	12.4	11.5	10.6	10.8	10.67	11.2	11	10.6	11.67	11	10.8	10	9.6	10.6	10	9.6	9.2	9.2
2016	12.12	12	12	12.8	12.5	13	13.2	12.4	11.7	12.1	11.6	12.33	11.5	11.6	12.3	12.4	12	11.5	10.8	10	9.2	8.9	9.2	10.4	9.4	9	9.2
2017	12.7	11.9	11.6	11.8	11.8	11.5	11.4	11.2	11	11.2	11.8	11.25	11.2	11.2	11.33	11.6	11	10.67	11.2	11.2	10.6	11.2	10.8	10.8	11.8	12.2	11.2
2018	14	13.6	14	14.6	14.8	15	14.6	15.6	13.8	15.2	14.5	15.75	13.33	13.6	14.16	13.4	12	12.5	12.2	12.6	11.8	12.2	11.8	12.6	11.8	12.6	13
2019	14.25	14.6	13.8	14.4	14.75	14.2	13.4	14																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	9.4	10.3	8.4	9.2	9.3	8.9	8.6	9	9.6	9.5	9.4	9.1	9.8	10	9.1	10.6	10	11.4	10.75	10.2	10.2	11	11.2	12.4	13.25	12	
2016	9.1	9.2	9.6	9.4	9.4	9.3	9.4	10.2	10.4	9.6	9.25	9.8	9.8	9.6	10	10.4	9.5	10.4	10.4	10.5	10.8	11	10.4	11.4	12	12.5	
2017	11	11.4	10.2	10.3	10.6	10.7	14.8	11.6	11.4	11.8	10.2	10.4	11.3	11.6	11.4	11.4	12.25	11.8	12.8	13	13.2	13.2	12.7	13.8	14	13.68	
2018	12	11.6	12.4	11	12	11.8	12.6	11.6	11.6	12	12.4	12.4	12.2	13	12.4	13.3	13.38	14.6	15.4	14	15.2	14.4	14	14.2	14.5	15	
2019																											

Tabla D.15: Precios semanales del plátano macho

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	5	6.28	7.1	7.44	7.7	7.52	7.44	7.22	6.6	6.74	6.38	5.86	4.9	4.43	4.52	4.62	4.46	4.4	4.08	4.12	4.18	4.36	4.52	5.54	6.04	6.4	6.42
2016	8	9.2	10.76	11.14	10.95	11.8	12	12.1	12.18	11.8	11.5	10.83	9.53	9.2	9.02	8.7	7.88	6.78	6.5	6.8	6.28	5.68	5.16	4.9	4.76	5	4.72
2017	5.4	5.96	5.68	5.68	5.6	5.72	6.06	6.68	7.18	7.4	7.2	7.03	6.38	5.64	5.27	6.2	7.22	7.67	8.12	8.14	8.14	7.78	8.16	8.24	8.52	10.24	10.9
2018	7.5	9.8	10.9	12.04	12.14	12.38	12.12	12.06	11.58	11.14	10.24	9.38	9.17	8.6	9.4	11.76	12.5	12.5	11.06	9	7.5	6.8	6.62	6.5	6.72	7.9	8.66
2019	6.15	10.24	10.9	10.86	10.9	11.26	11.1	10.7																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	6.38	6.06	5.76	5.82	6.26	5.6	5.8	5.6	4.48	4.58	4.74	4.7	4.5	4.31	4.1	4.16	5	5.66	5.92	5.78	5.66	5.7	5.4	6.04	6.75	6.5	
2016	4.28	3.96	4.34	4.66	5.38	6.16	6.38	6.9	5.74	4.78	4.25	4.12	4.24	4.4	4.42	4.5	4.55	4.54	4.6	4.45	4.8	4.92	5.04	5.08	5.2	6	
2017	11.3	11.5	10.6	9.5	8.3	9.4	11.8	7.44	6.52	6.1	5.34	5.26	5.18	5.04	4.64	4.5	4.62	4.6	4.58	4.67	4.98	5.4	6.3	6.98	7.42	8.15	
2018	8	7	6.82	6.76	6.5	6.18	6.1	6.62	6.24	6.5	6.14	6.4	5.62	5.6	5.76	5.88	6.46	6.86	6.95	6.8	6.46	6.46	5.86	6.03	6		
2019																											

Tabla D.16: Precios semanales del plátano Tabasco

Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27
2015	11.42	10.67	9.93	9.2	8.75	8.4	7.93	8.6	9.53	12.13	13.33	16.2	15.8	14.78	13.33	11.73	11.47	13.33	13.13	11.33	10.07	10.87	11.67	11.13	10.13	11.47	13.33
2016	19.83	12.93	14.53	18.47	15.17	13.27	11.53	10.2	9.07	10.8	12.67	12.67	12.83	11.53	8	7.87	9.93	9.33	9.4	10.4	9.53	8.73	8.73	9.67	9	9.87	9.8
2017	16	9.2	6.33	6.33	5.87	5.5	6.07	6.13	5.8	9.07	8.73	7	7.13	8.07	10.33	12.87	14.93	14.78	14.47	13.73	13.2	13.27	14.53	16	14.93	16.07	15.8
2018	20.33	10.67	9.13	9.33	8.13	6.83	6.07	6	8.27	11.33	12.8	13.5	12.22	10.87	9.67	10.4	11.07	11.08	9.47	9.6	9.07	8.73	9	8.6	9.2	10.2	9.73
2019	16.25	21.87	18.53	11.53	8.83	12.4	10.6	10.13																			
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53	
2015	13.33	10.83	9.4	9.6	9.6	10.1	10.13	14.33	13.13	10.33	9.27	8.73	8.67	9.4	10	10	10.42	9.87	9.5	10.13	10.33	13.53	18.27	25.93	24.67	21.67	
2016	8.13	6.33	5.67	7	8.27	9.13	10.13	10.33	13.07	15.2	13.75	12.73	12.27	14.4	15.53	13.07	13.08	16.33	18.33	17.92	16.13	16.13	15.4	14.6	15.47	16	
2017	20.33	22.2	27.27	22.53	20.73	23.13	20.93	15.87	15.33	14.67	13.8	9.8	7.27	7	8.53	9	8.67	8.73	9.33	9.5	9.53	13.6	19.2	20.87	21.25	16.91	
2018	9.93	9.2	14.07	15.73	15.27	12.6	13.67	14	12.87	13.13	13.53	16.27	16.07	14.33	14.27	14.47	14.58	17.93	20.4	23.5	27	28.53	18.73	14.93	12.75	16.67	
2019																											

Tabla D.17: Precios semanales del tomate saladette

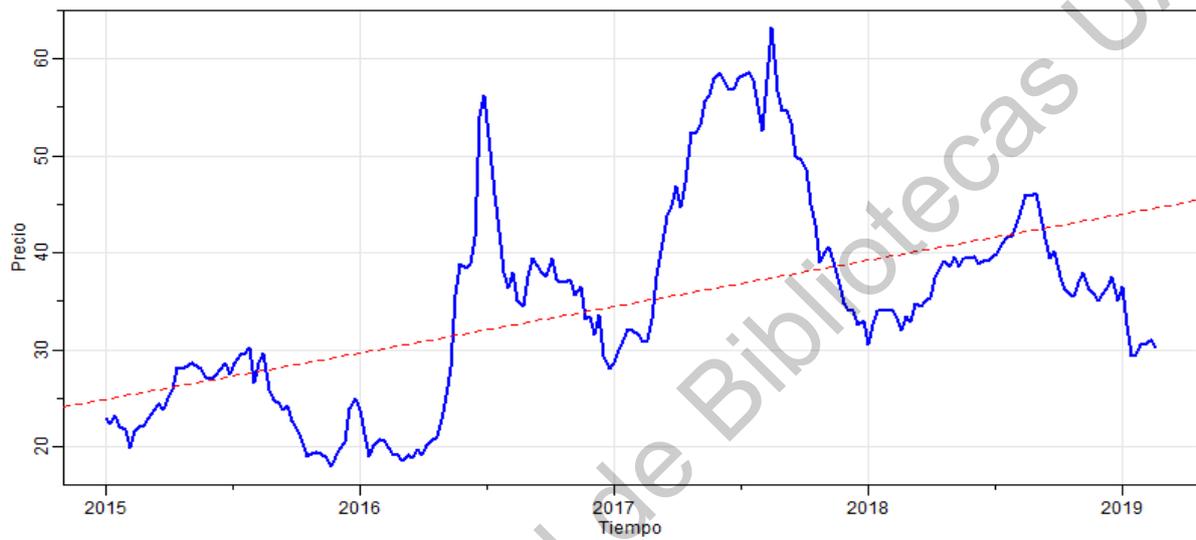
Año	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	
2015	3	4.2	3.9	3.9	3.88	3.9	4.1	4	4.5	4	4.12	3.9	4.04	4	3.8	4.1	4.18	4.5	4.9	4.6	5.1	6.3	6	6.3	6.7	6.3	6.3	
2016	7.12	4.44	4.2	4.7	5	4.6	4.3	4.4	4.1	4.46	4.8	5.17	4.88	5.38	5.3	5.8	6.2	6.42	6.28	7	7.6	8	8.2	8.6	8.6	8.76	8.66	
2017	7.8	7.2	7.4	7.9	8	7.88	7.7	8.7	8	8.5	8.2	8.25	8.4	9.6	9.33	8.9	10	9.83	10.6	10.3	10.5	10.7	12.2	11.4	12.8	13.2	12.4	
2018	9	8.7	9.5	9.8	9.8	10	10.3	10.8	9.4	10.2	9.6	9.75	10.33	9.6	11	10.8	12	12.25	12.4	12.8	13.4	13.4	13.8	13.4	14.4	15.2	14.4	
2019	8	8.2	8.4	9.1	8.38	8.6	8	8.4																				
Año	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44	S45	S46	S47	S48	S49	S50	S51	S52	S53		
2015	6.3	6.1	6.4	5.9	6.1	5.8	6.02	6.2	5.3	5.25	5.6	5.3	4.6	4.3	4.3	4.6	4.62	4.9	4.5	4.6	4.5	4.4	4.8	4.7	5.75	6.5		
2016	8	8.9	8.7	8.1	8.5	9.1	9.1	8.3	9.1	8.2	7.5	7.5	6.8	7.2	6.6	7.1	6.75	6.6	6.8	6.25	7	6.7	7.2	6.7	7.4	7		
2017	13.2	13	11.2	10.9	10.8	12.2	12.7	8.8	8.9	7.8	8.2	7.7	7.4	7.8	7.8	8.1	8.12	8.4	8.8	8.75	9.1	8.9	9.7	9.1	9	9.17		
2018	14	12	13.2	13.2	13.2	13.6	12.8	9.8	9	8.9	7.7	8.6	8.2	8.6	8.5	8.5	8.75	8.6	9	8.5	8.6	9	8.5	8.5	8	8.5		
2019																												

Tabla D.18: Precios semanales de la toronja roja

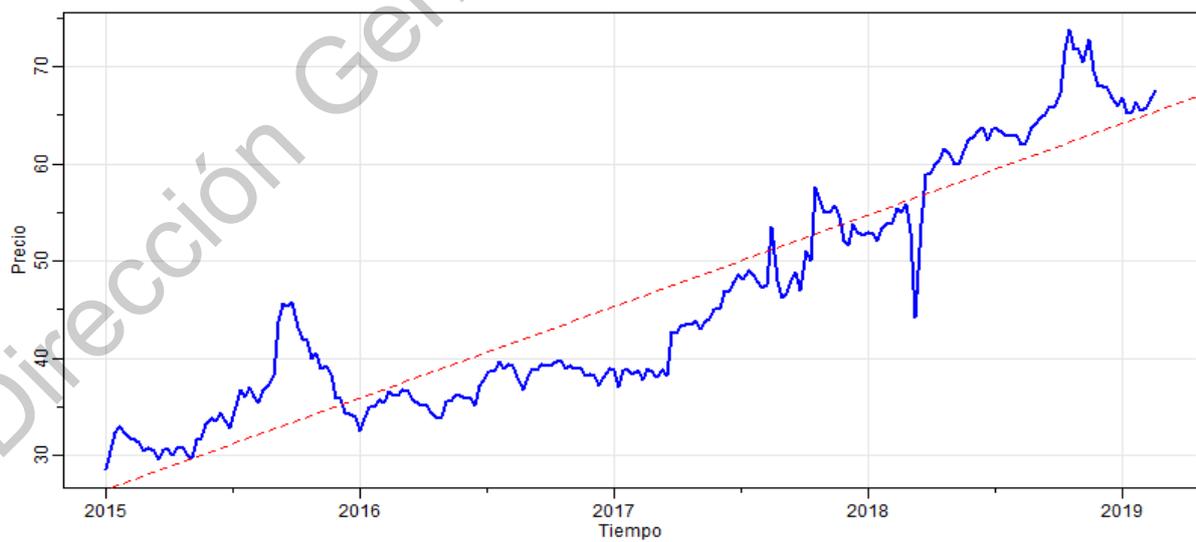
Dirección General de Bibliotecas UAQ

## APÉNDICE E

## Tendencia lineal de las series de tiempo



*Figura E.1: Serie semanal del aguacate Hass*



*Figura E.2: Serie semanal del cacahuete*

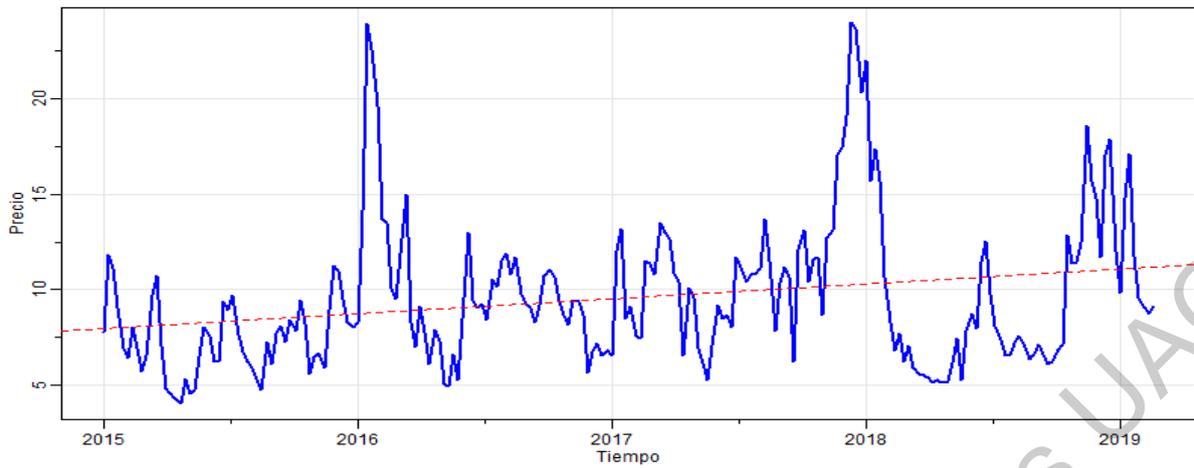


Figura E.3: Serie semanal de la calabacita italiana

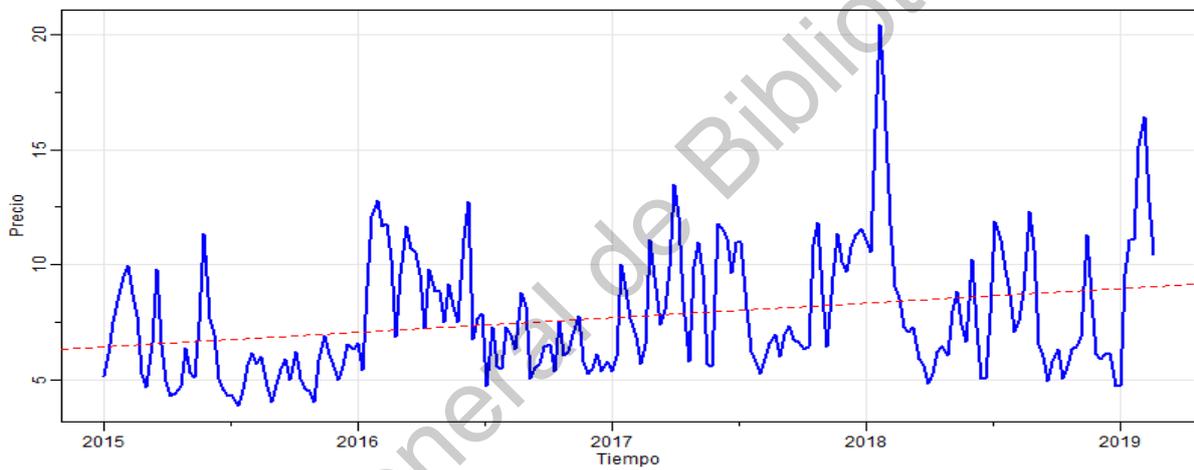


Figura E.4: Serie semanal del chayote sin espinas

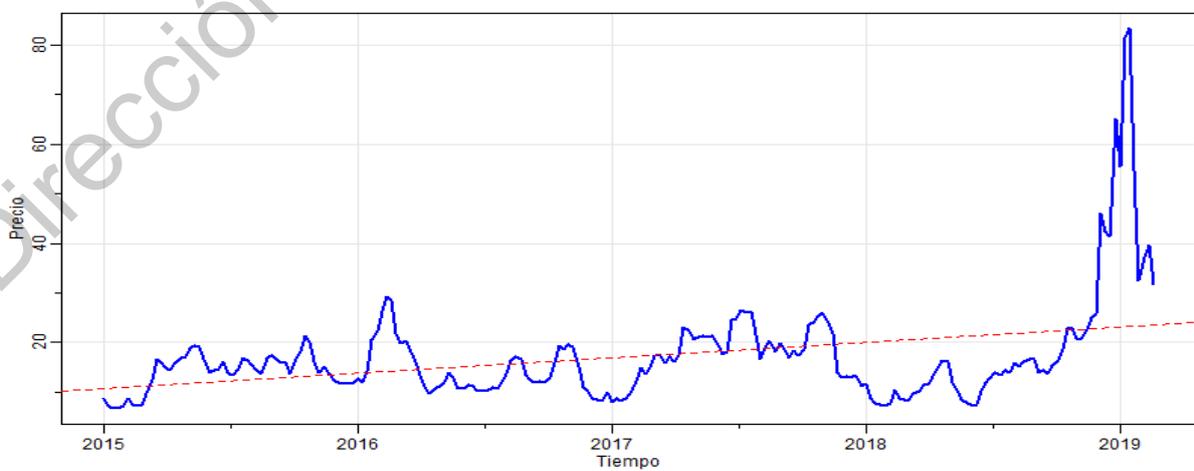


Figura E.5: Serie semanal del chile serrano

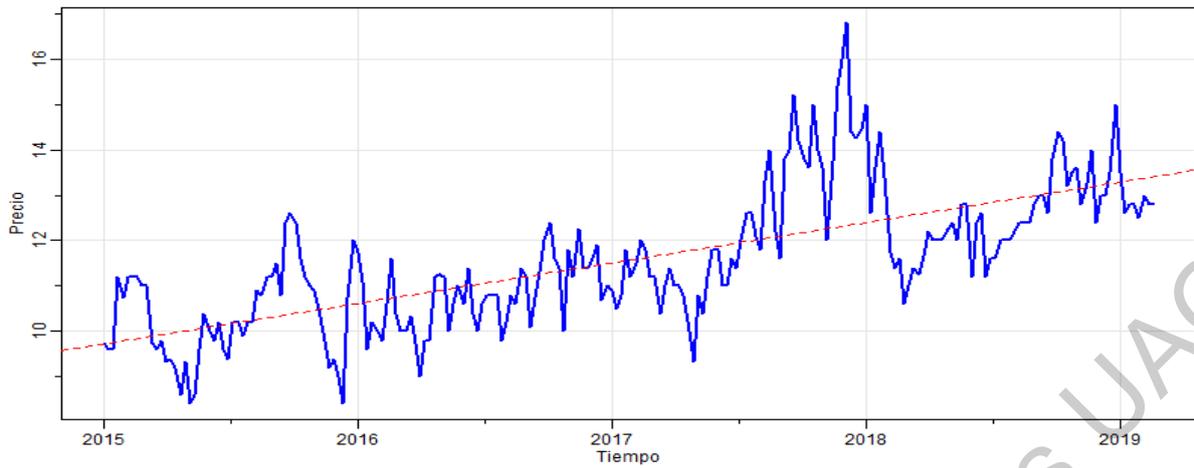


Figura E.6: Serie semanal de la coliflor grande

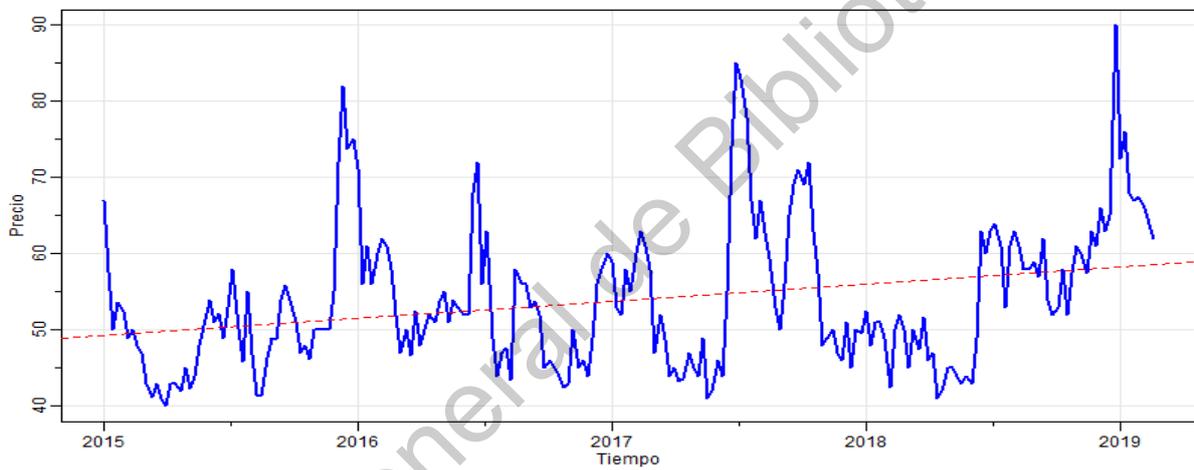


Figura E.7: Serie semanal del epazote

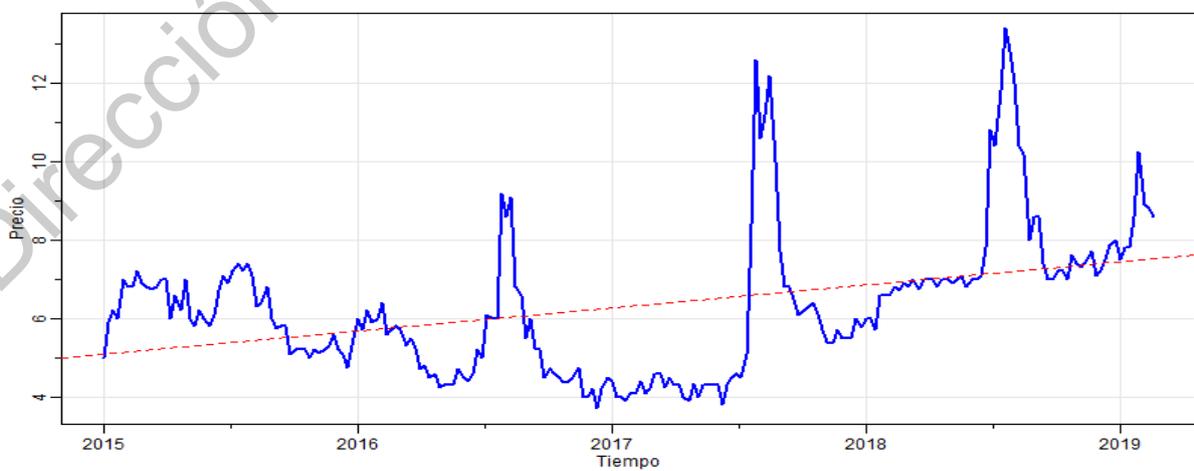


Figura E.8: Serie semanal de la jícama

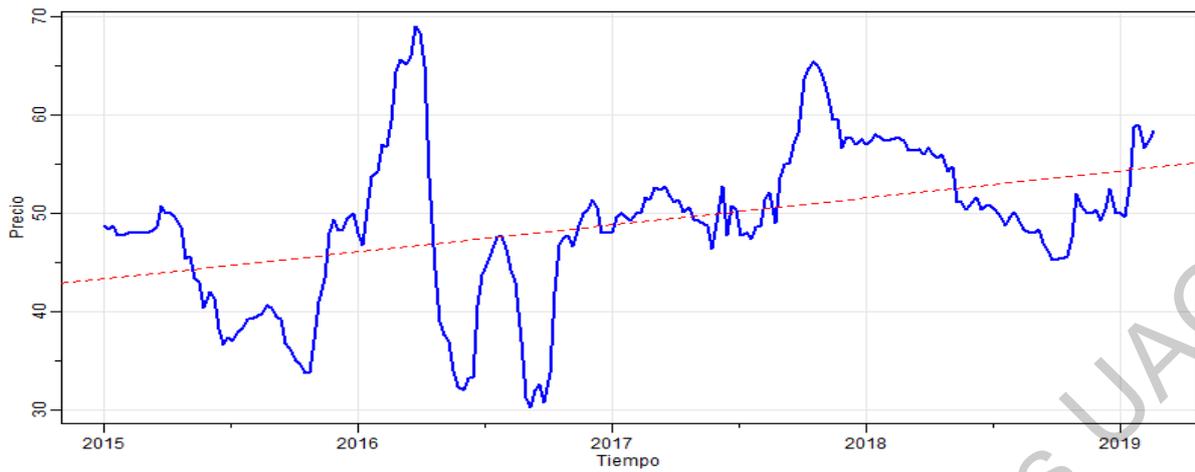


Figura E.9: Serie semanal del kiwi

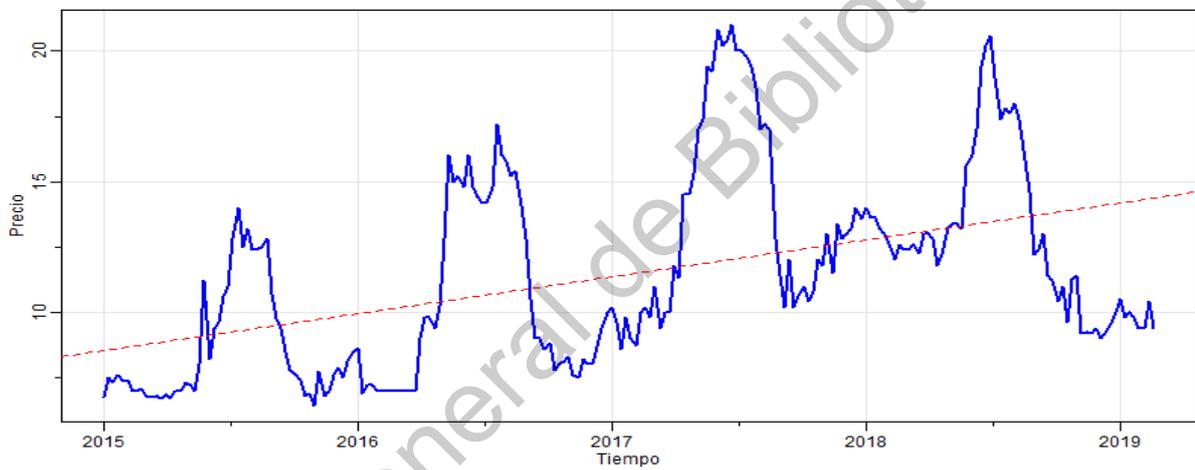


Figura E.10: Serie semanal de la lima

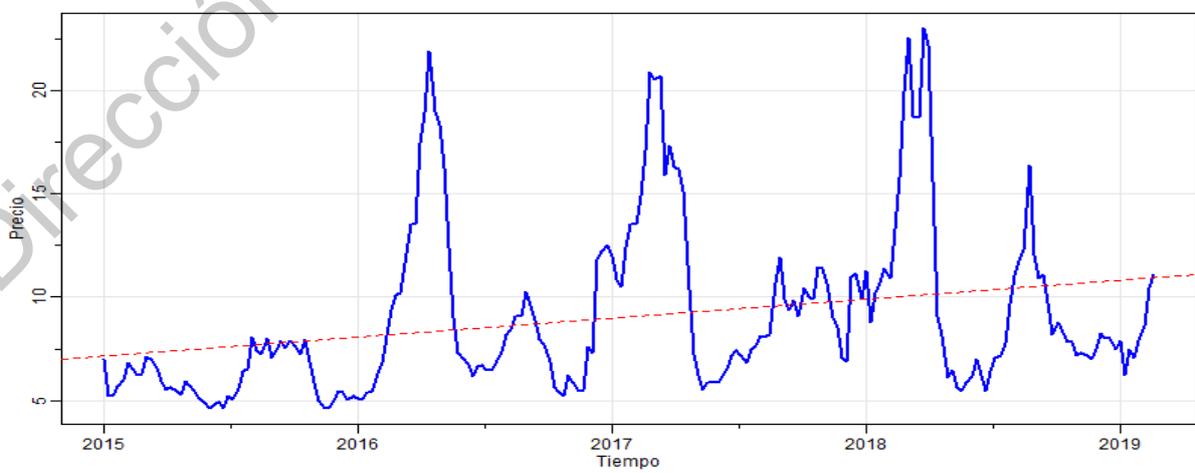


Figura E.11: Serie semanal del limón #5

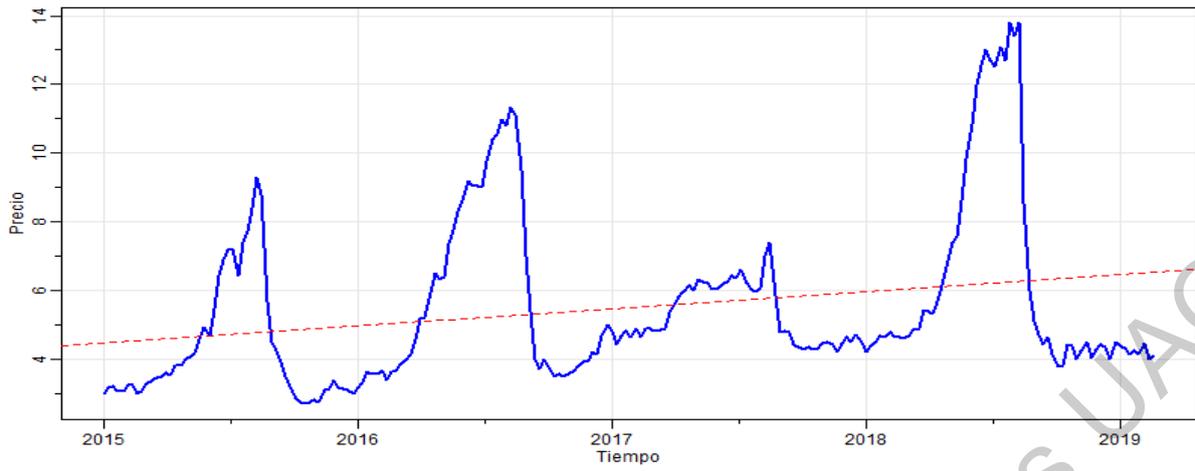


Figura E.12: Serie semanal de la naranja Valencia grande

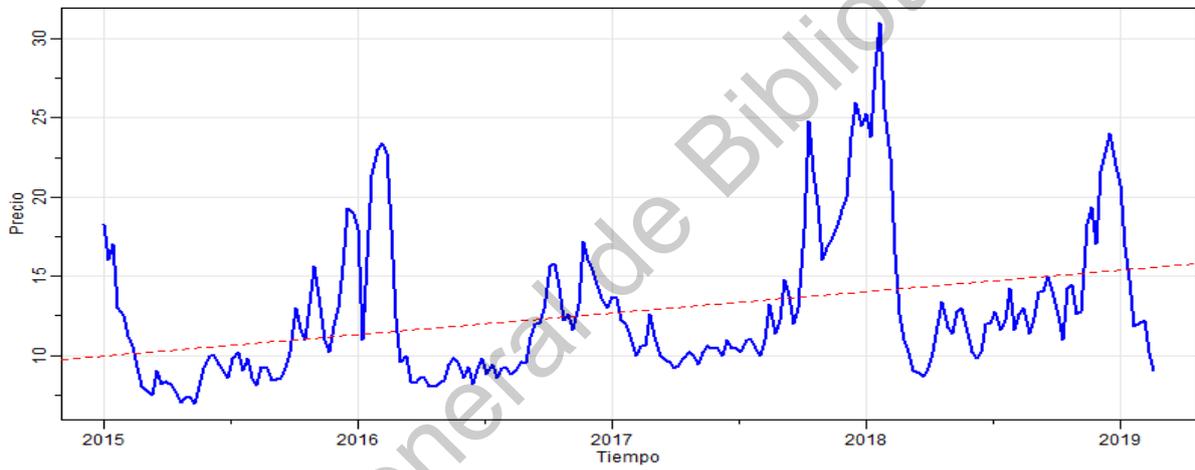


Figura E.13: Serie semanal del nopal grande

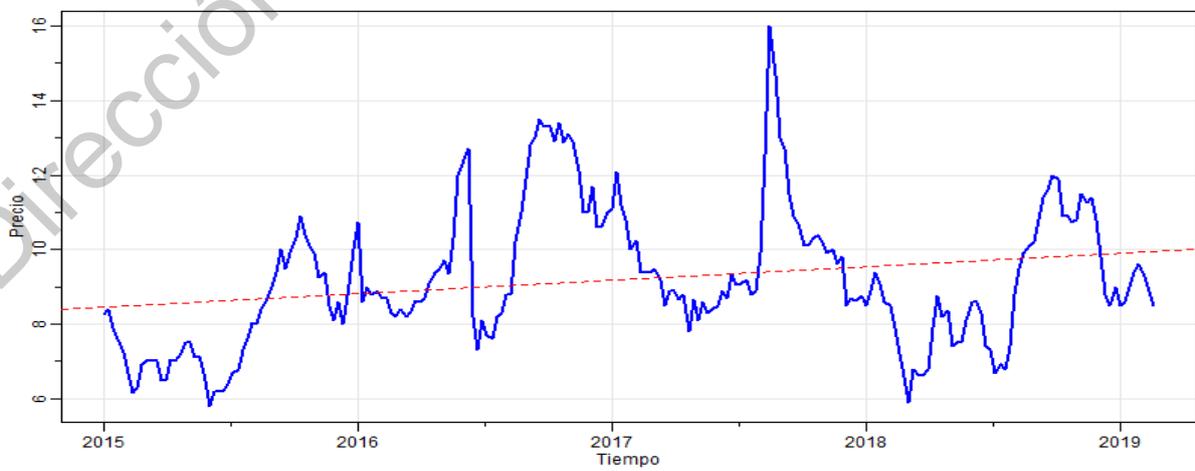


Figura E.14: Serie semanal de la piña mediana

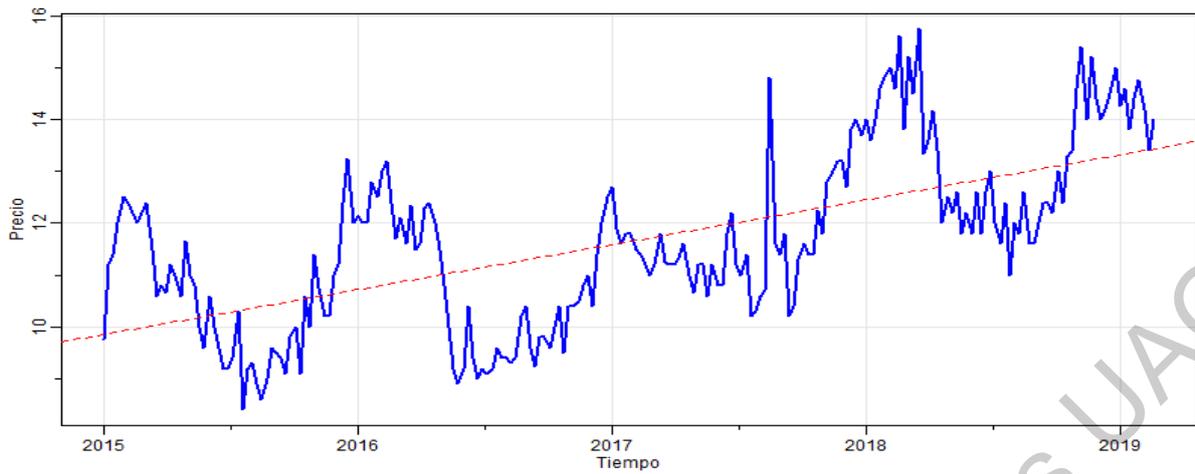


Figura E.15: Serie semanal del plátano macho

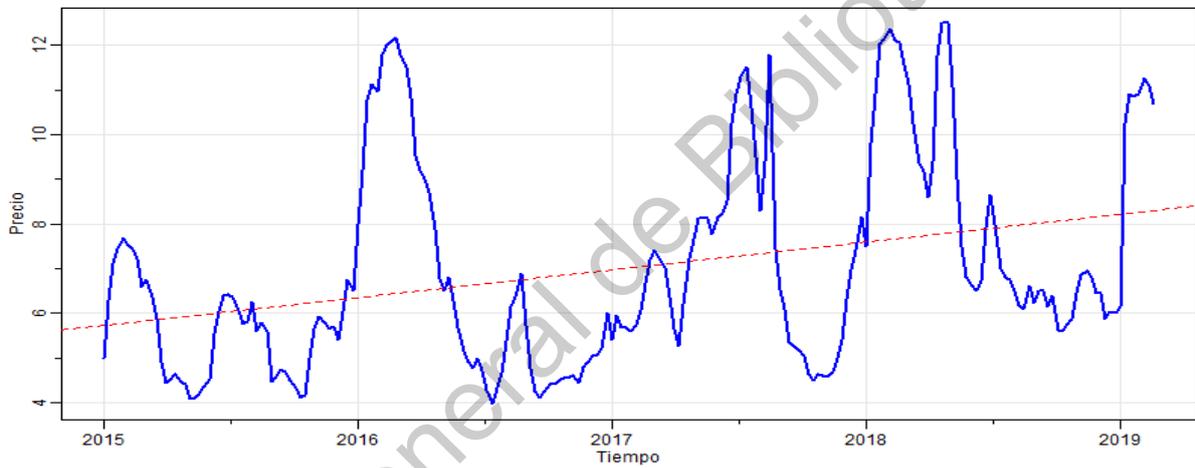


Figura E.16: Serie semanal del plátano Tabasco

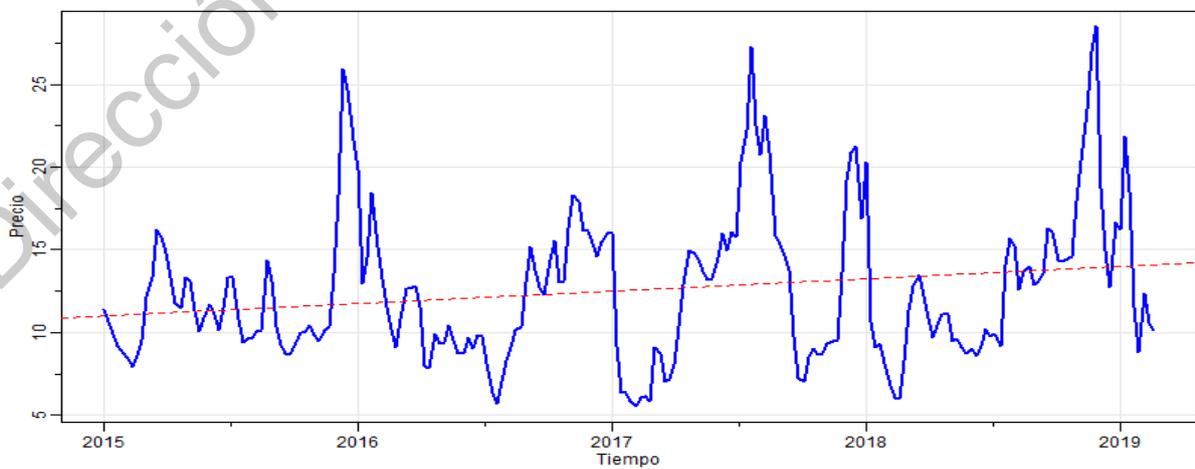


Figura E.17: Serie semanal del tomate saladette

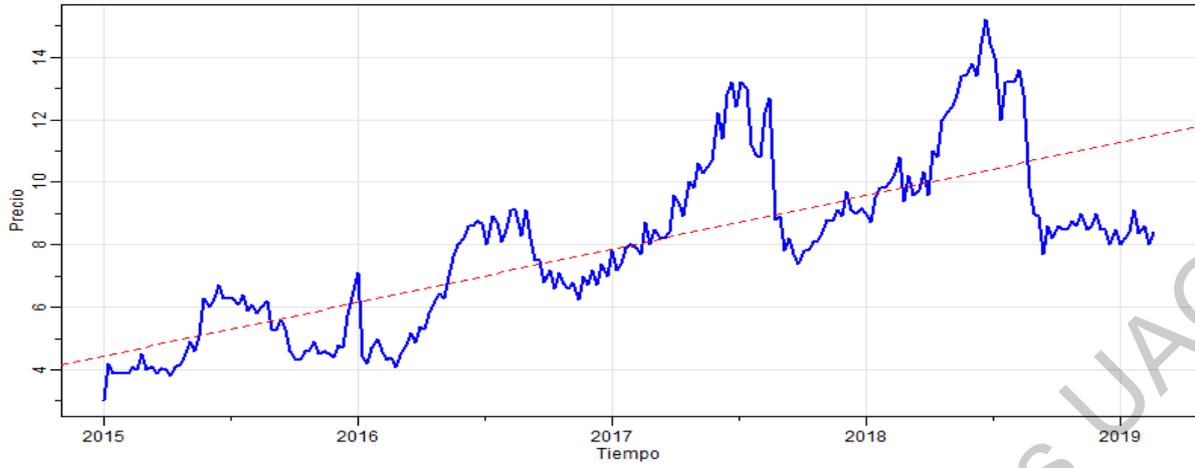


Figura E.18: Serie semanal de la toronja roja

Dirección General de Bibliotecas UAQ

## APÉNDICE F

## Series de tiempo estacionarias

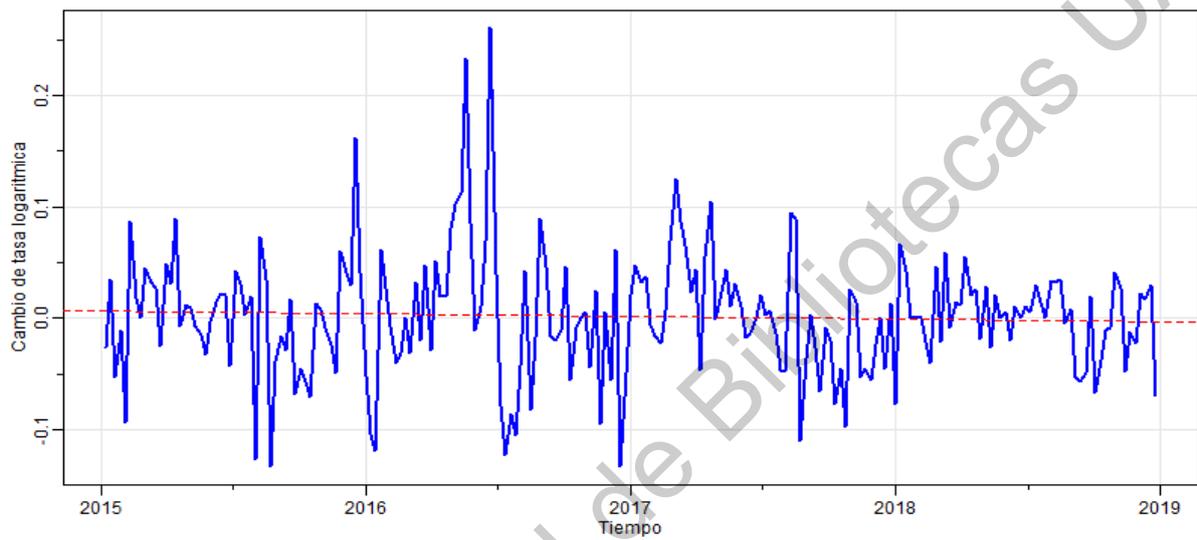


Figura F.1: Serie estacionaria del aguacate Hass

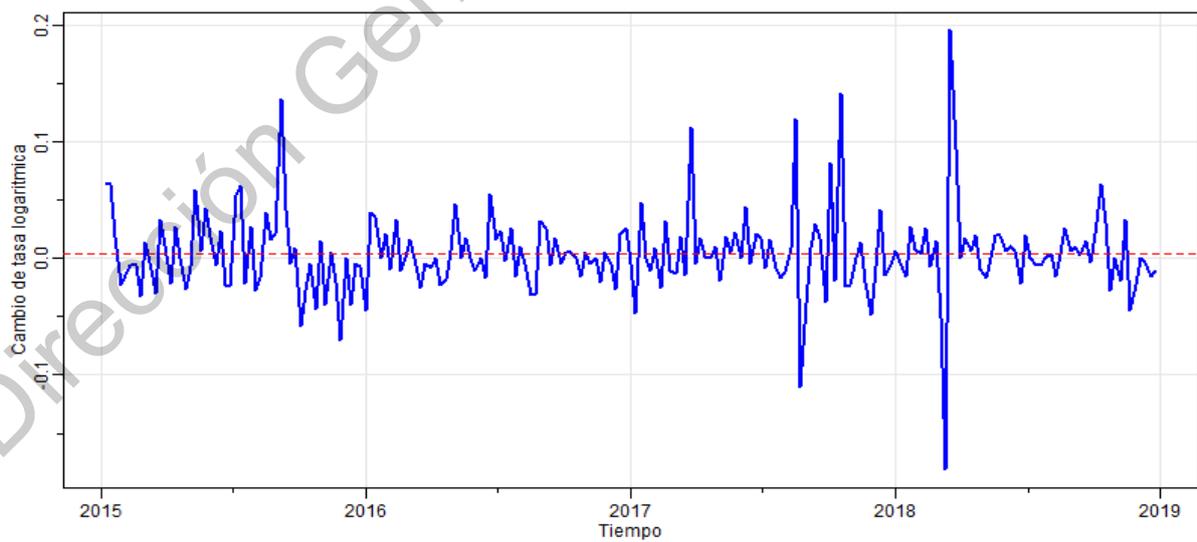
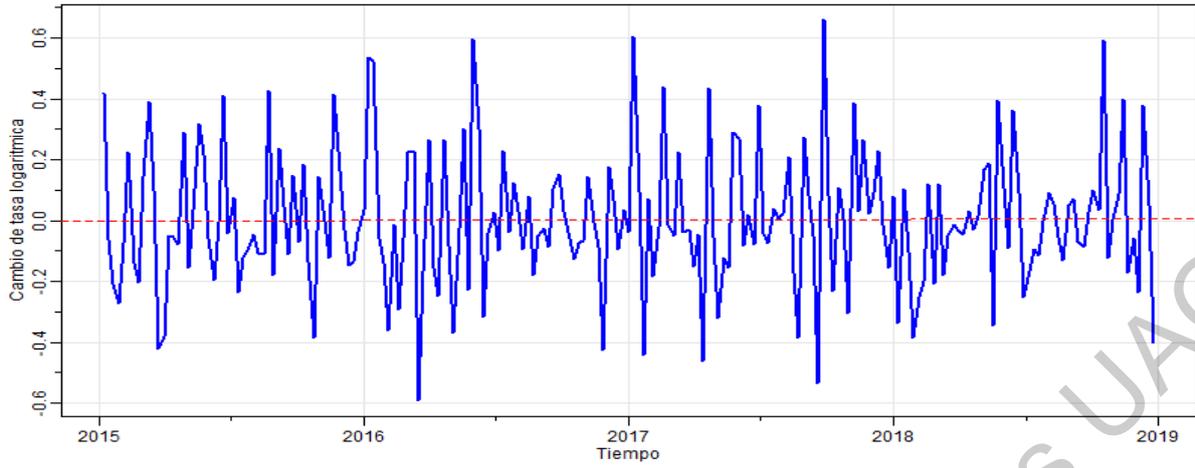
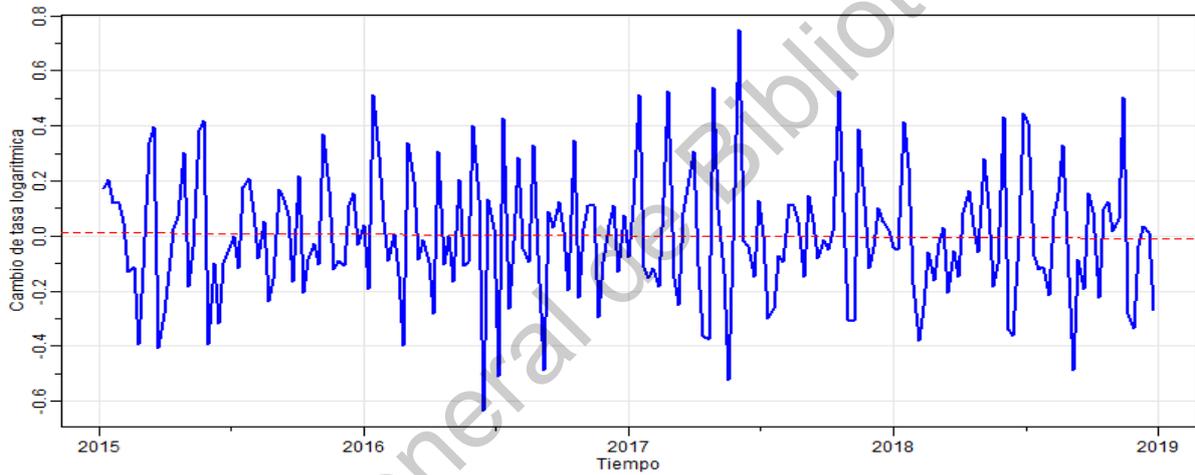


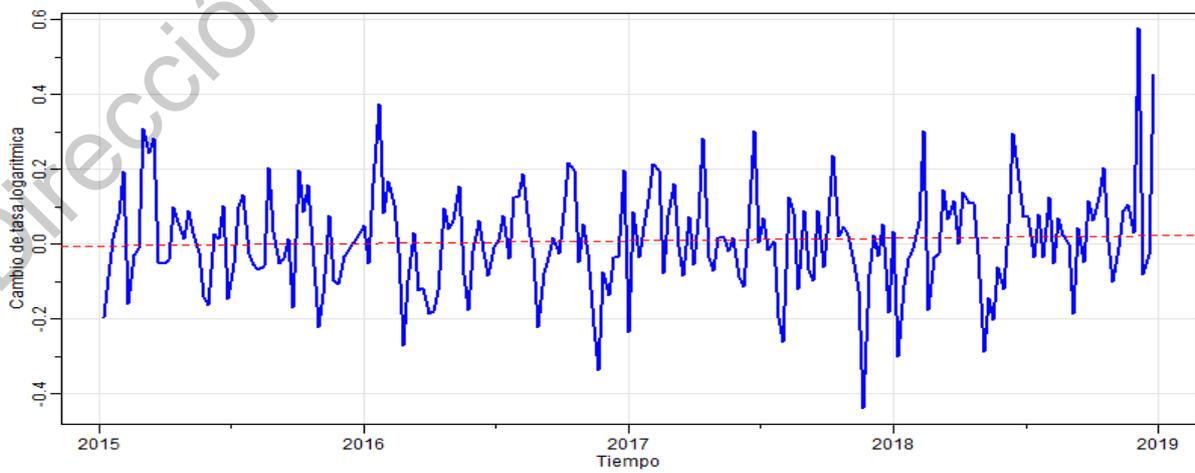
Figura F.2: Serie estacionaria del cacahuate



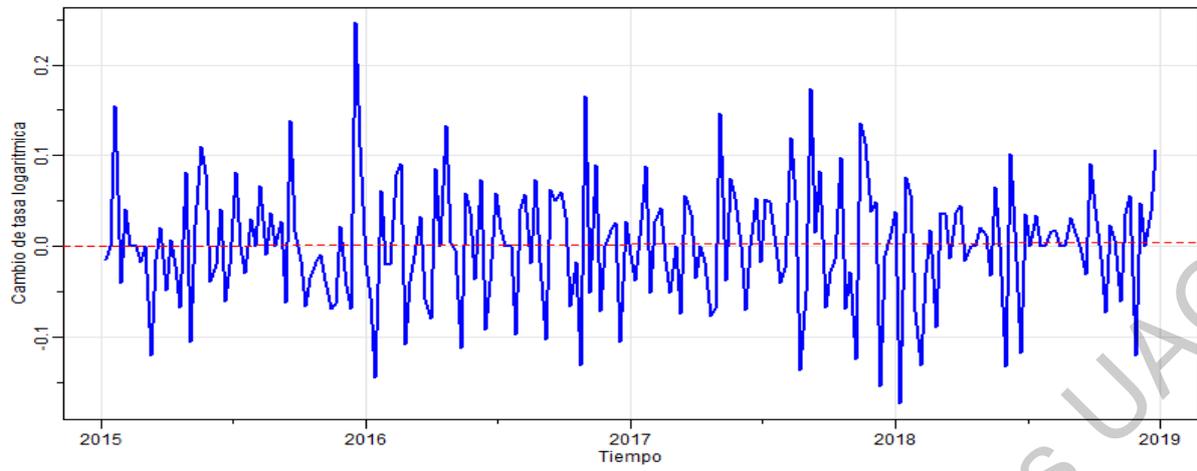
*Figura F.3: Serie estacionaria de la calabacita italiana*



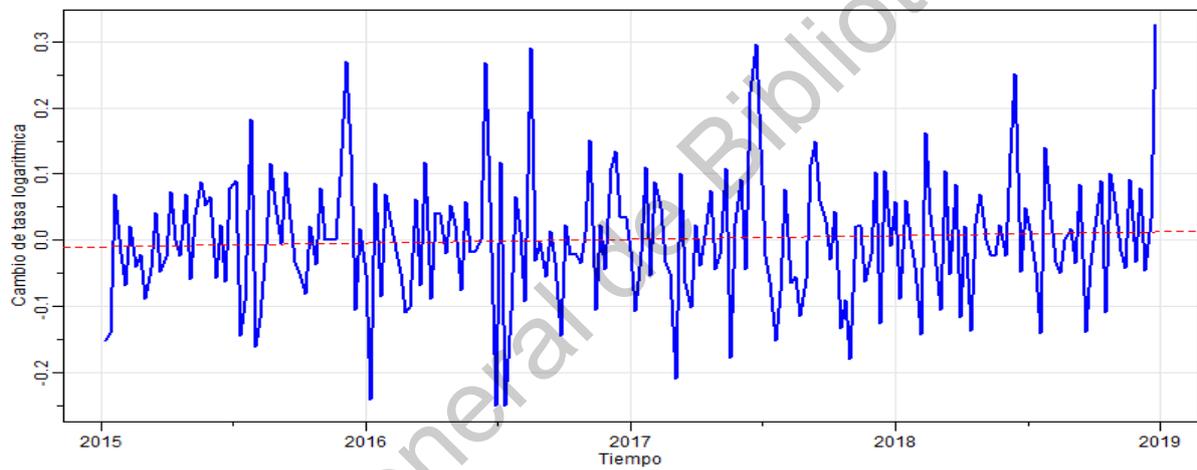
*Figura F.4: Serie estacionaria del chayote sin espinas*



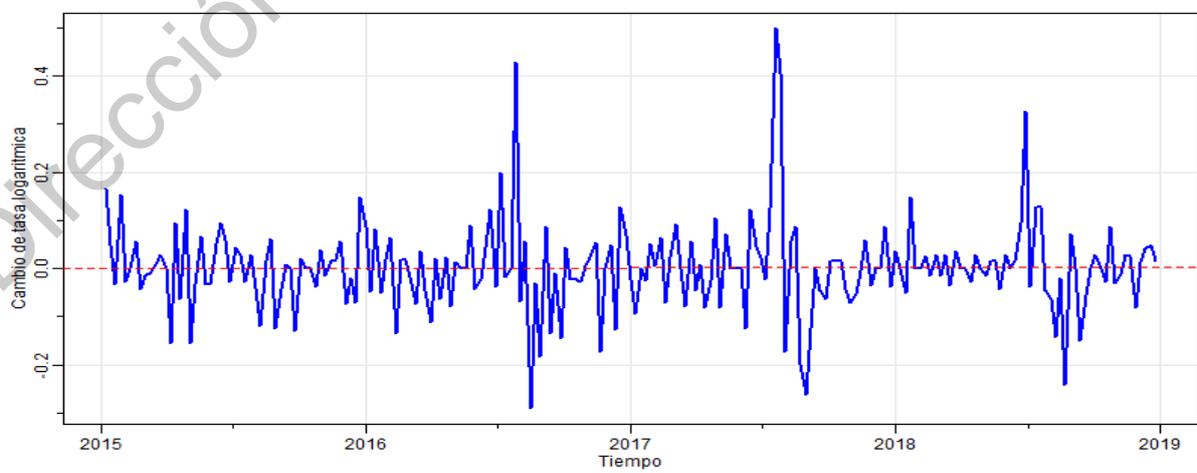
*Figura F.5: Serie estacionaria del chile serrano*



*Figura F.6: Serie estacionaria de la coliflor grande*



*Figura F.7: Serie estacionaria del epazote*



*Figura F.8: Serie estacionaria de la jícama*

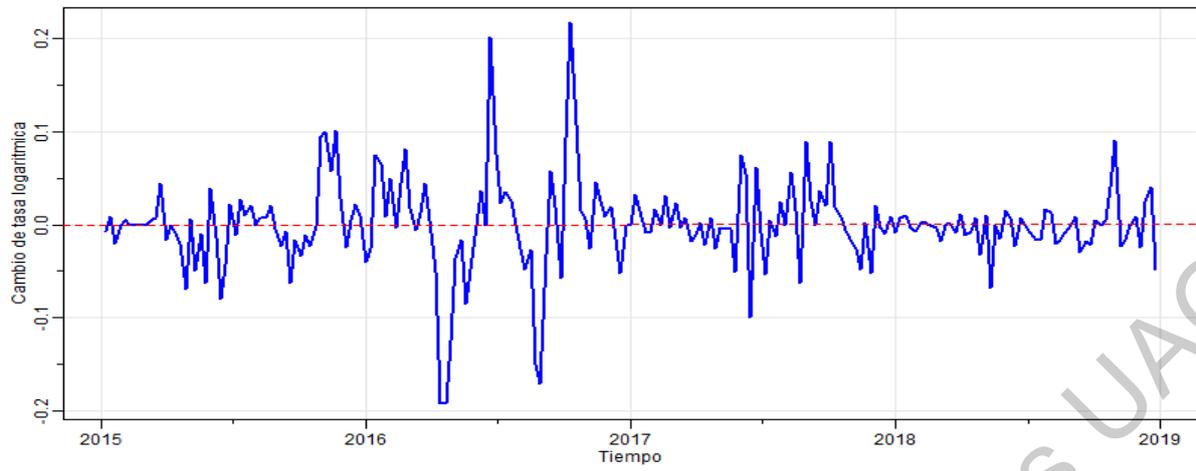


Figura F.9: Serie estacionaria del kiwi

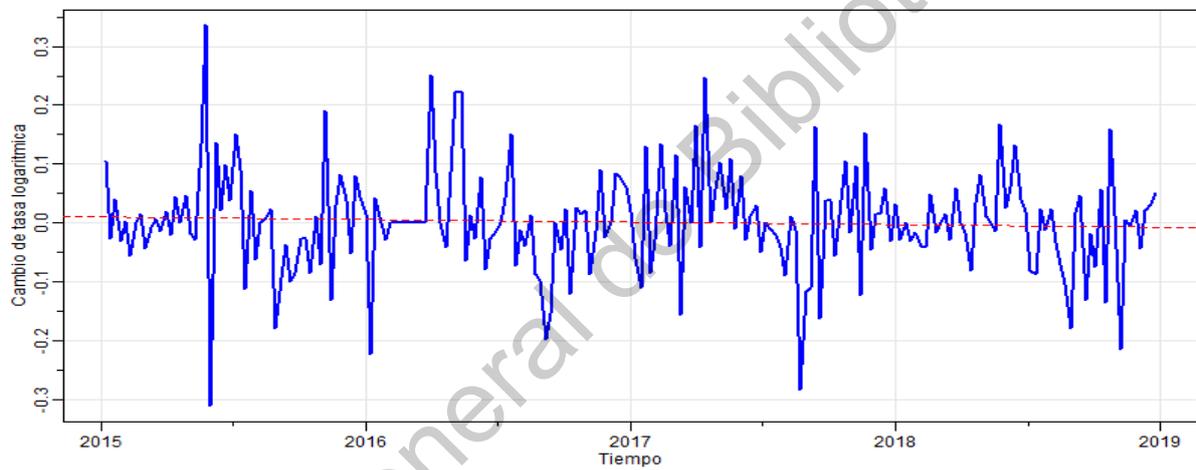


Figura F.10: Serie estacionaria de la lima

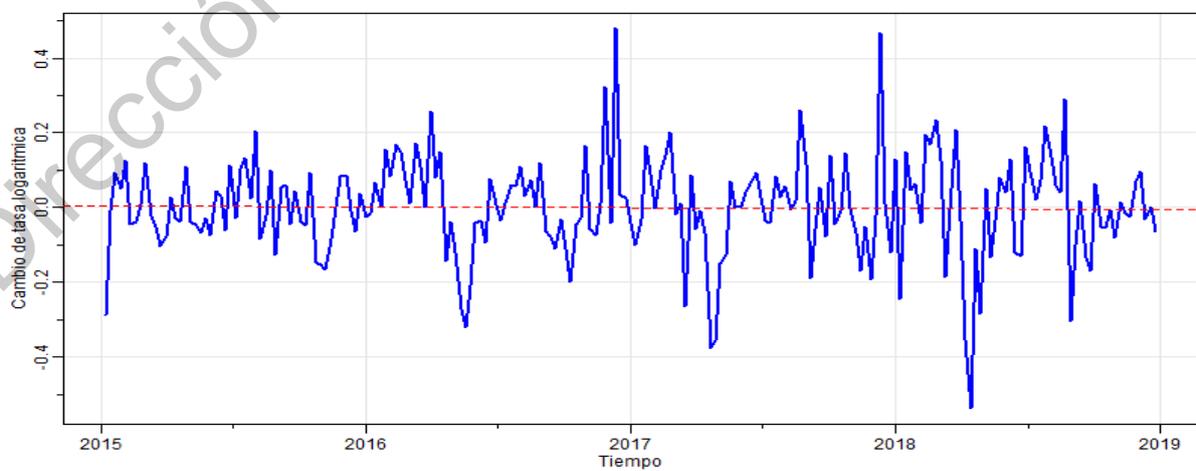


Figura F.11: Serie estacionaria del limón #5

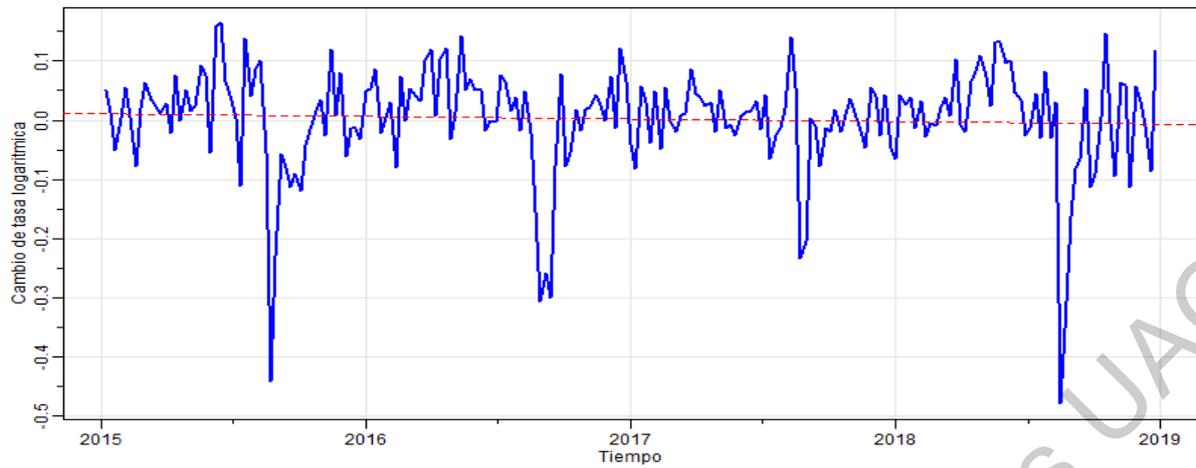


Figura F.12: Serie estacionaria de la naranja Valencia grande

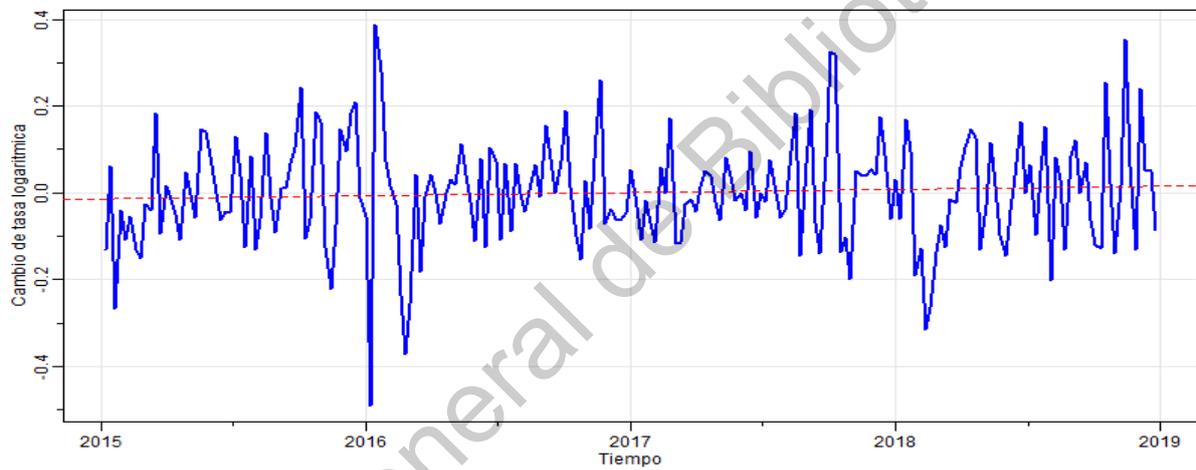


Figura F.13: Serie estacionaria del nopal grande

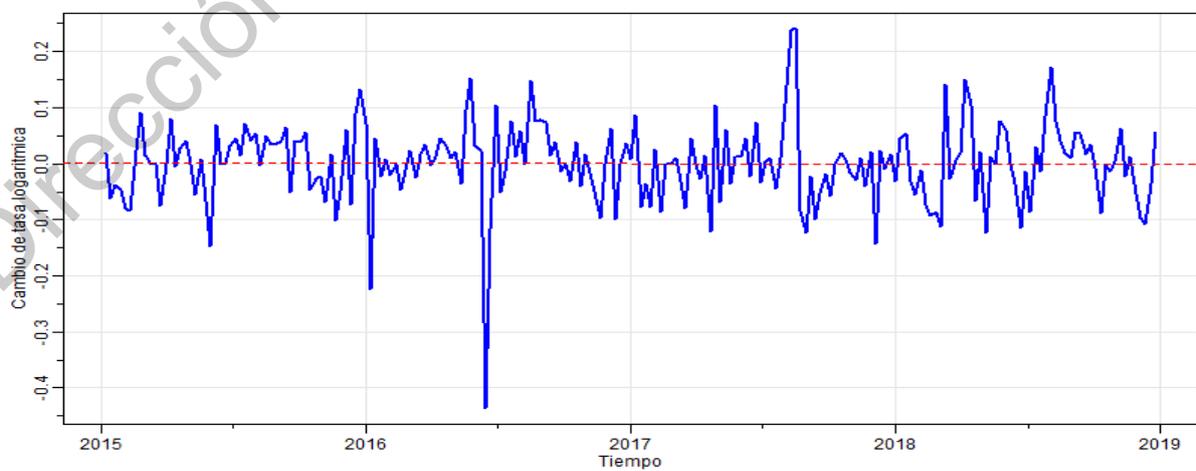


Figura F.14: Serie estacionaria de la piña mediana

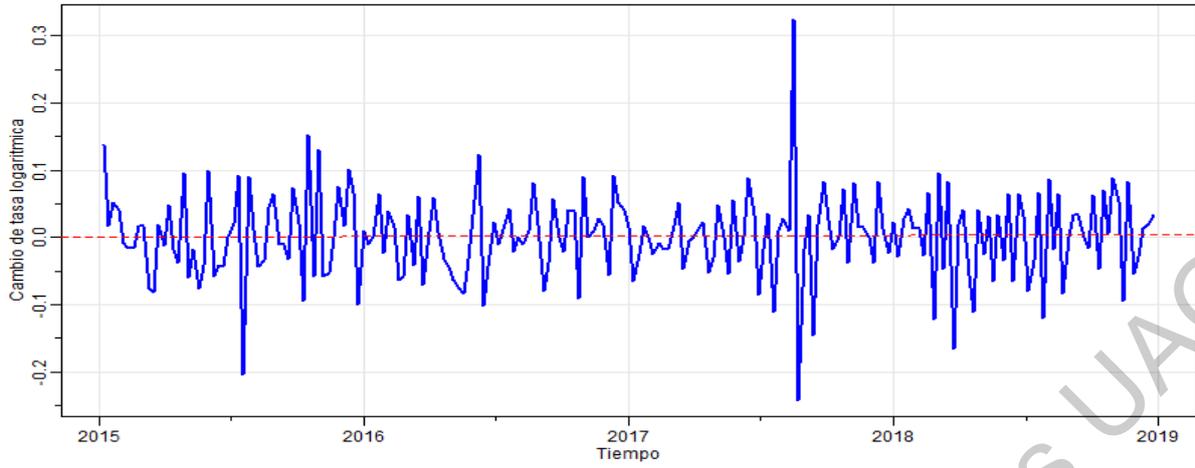


Figura F.15: Serie estacionaria del plátano macho

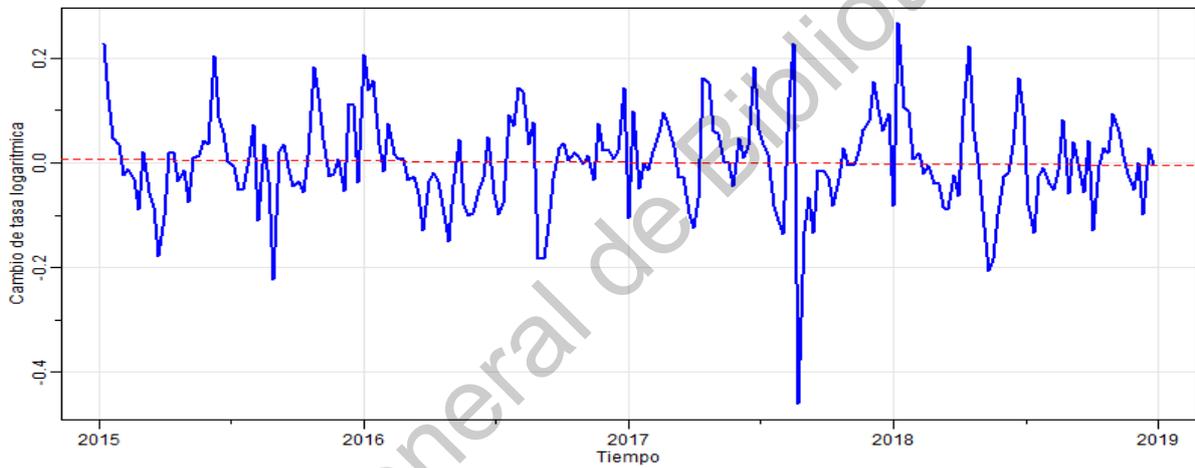


Figura F.16: Serie estacionaria del plátano Tabasco

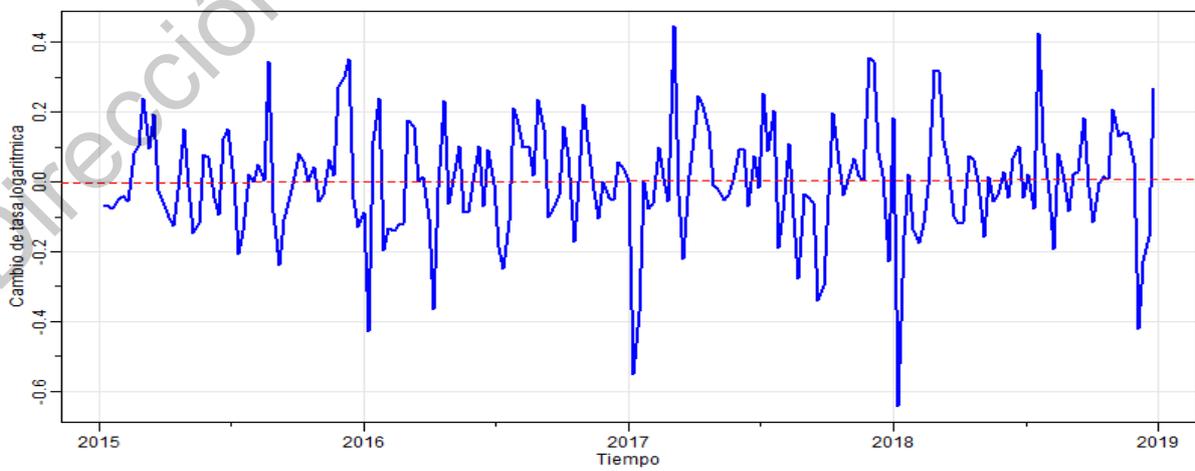
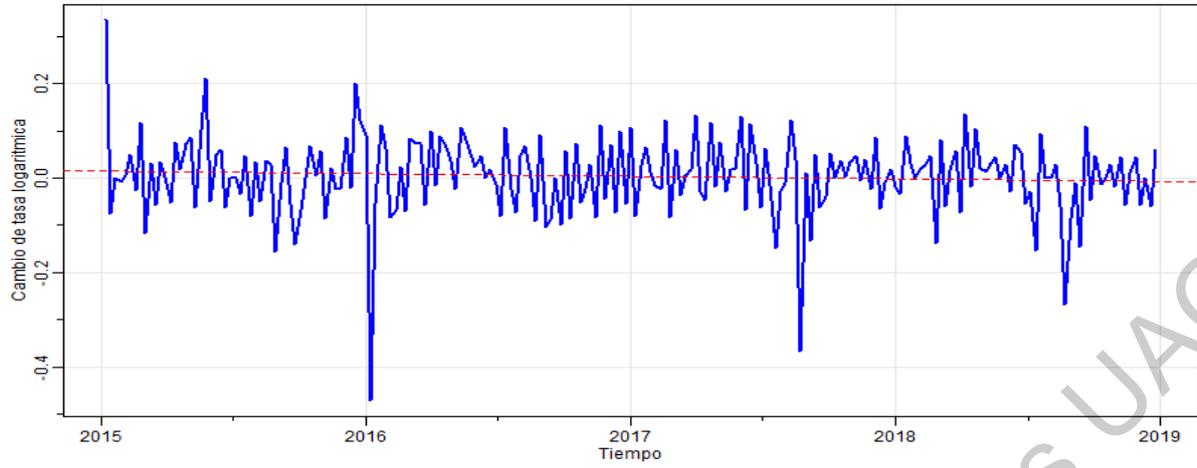


Figura F.17: Serie estacionaria del tomate saladette



*Figura F.18: Serie estacionaria de la toronja roja*

Dirección General de Bibliotecas UAQ

## APÉNDICE G

## Funciones de autocorrelación muestral

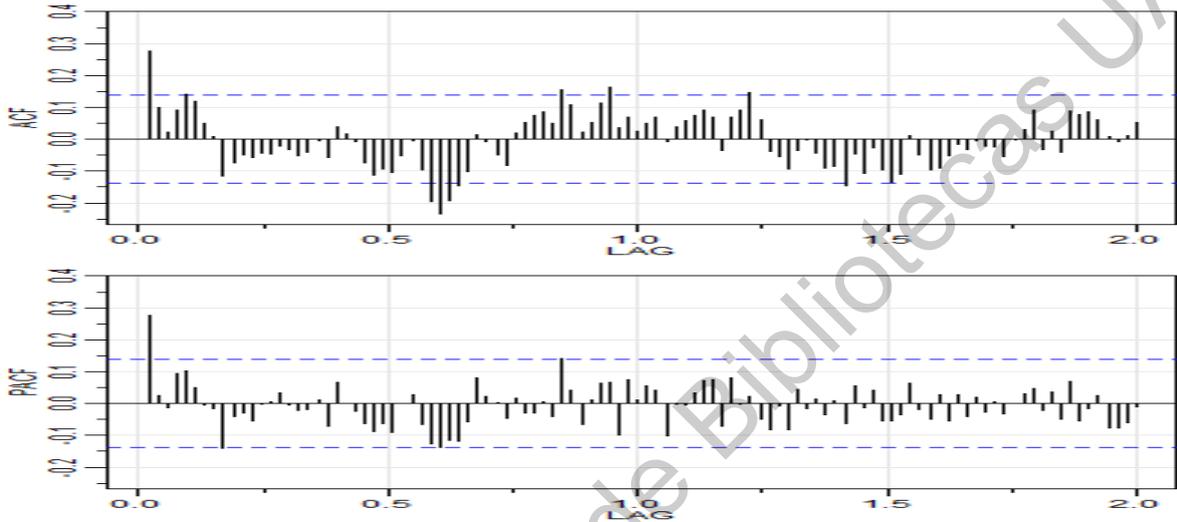


Figura G.1: Análisis de correlación del aguacate Hass

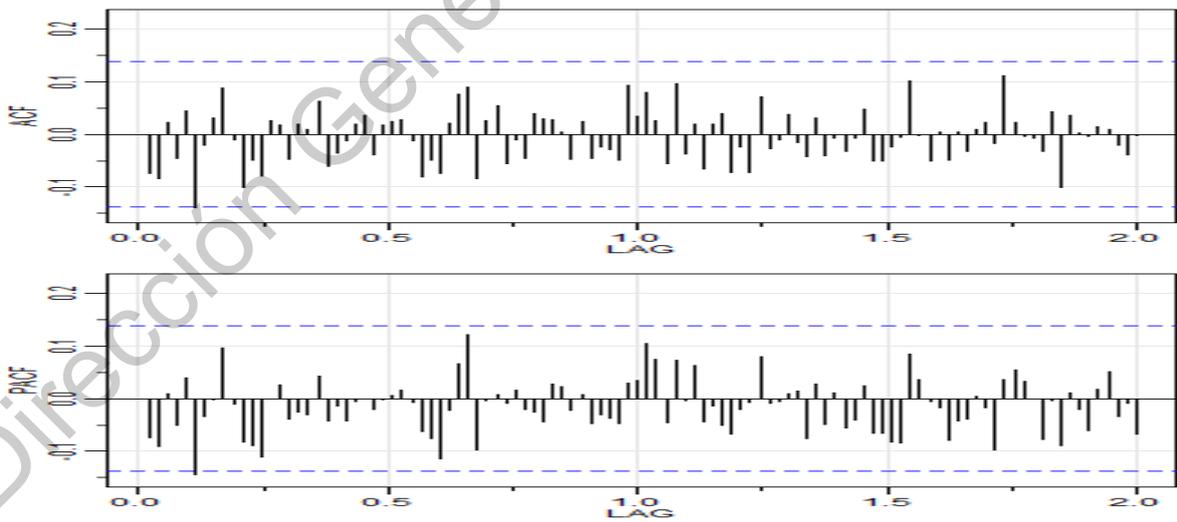


Figura G.2: Análisis de correlación del cacahuete

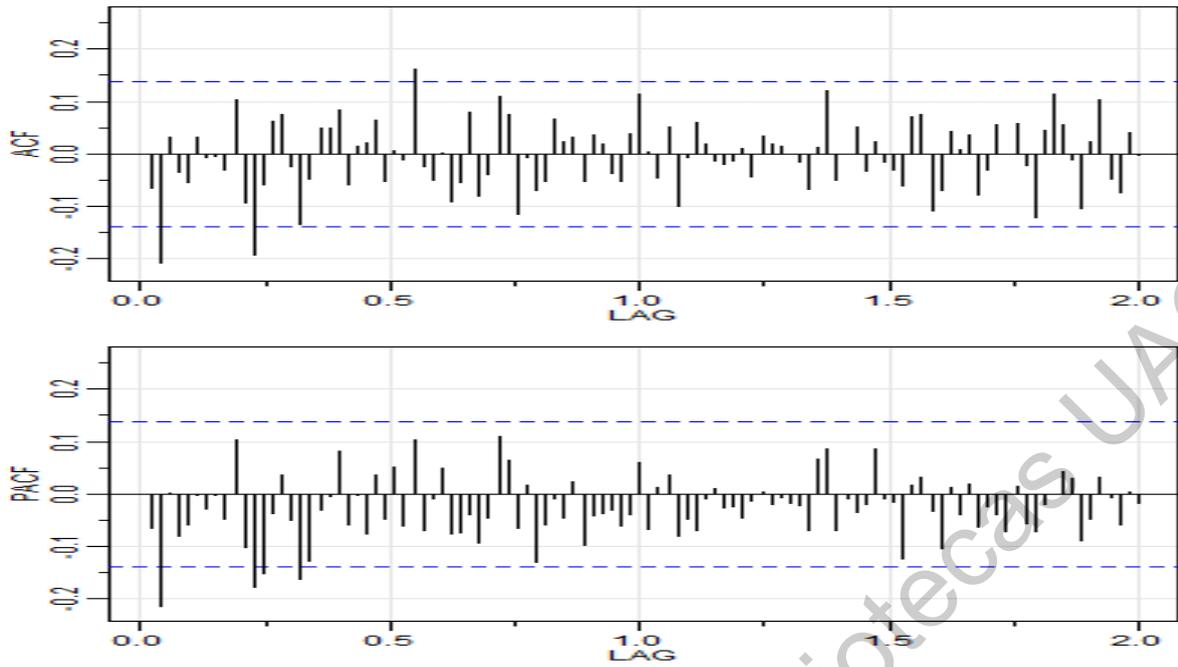


Figura G.3: Análisis de correlación de la calabacita italiana

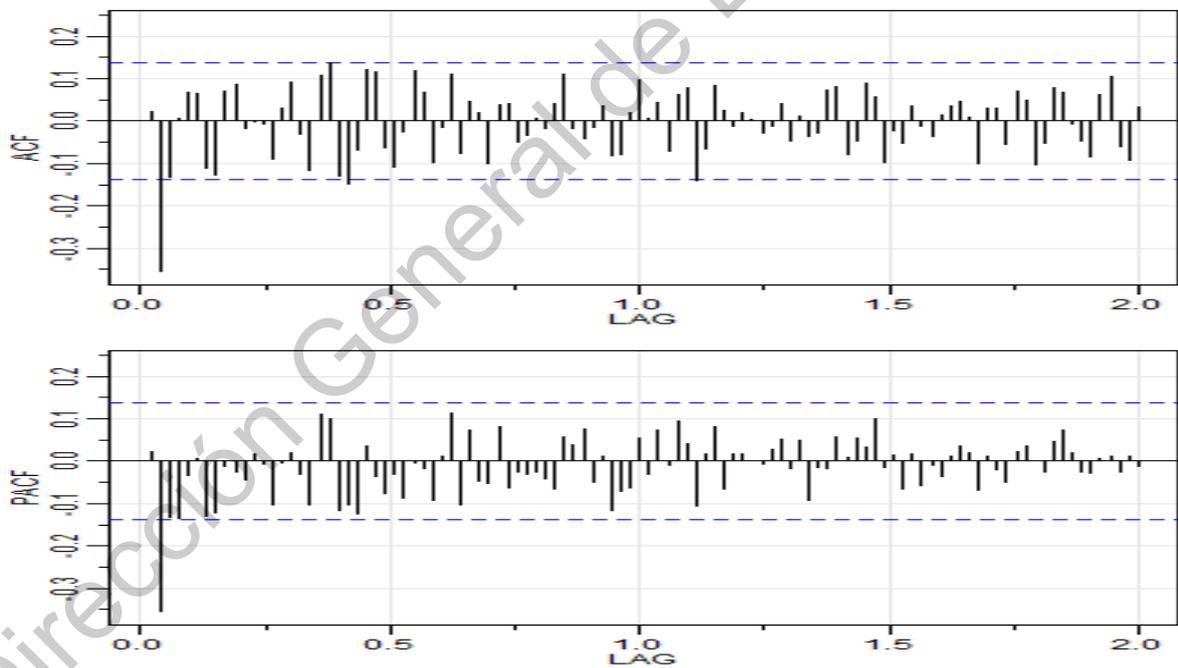


Figura G.4: Análisis de correlación del chayote sin espinas

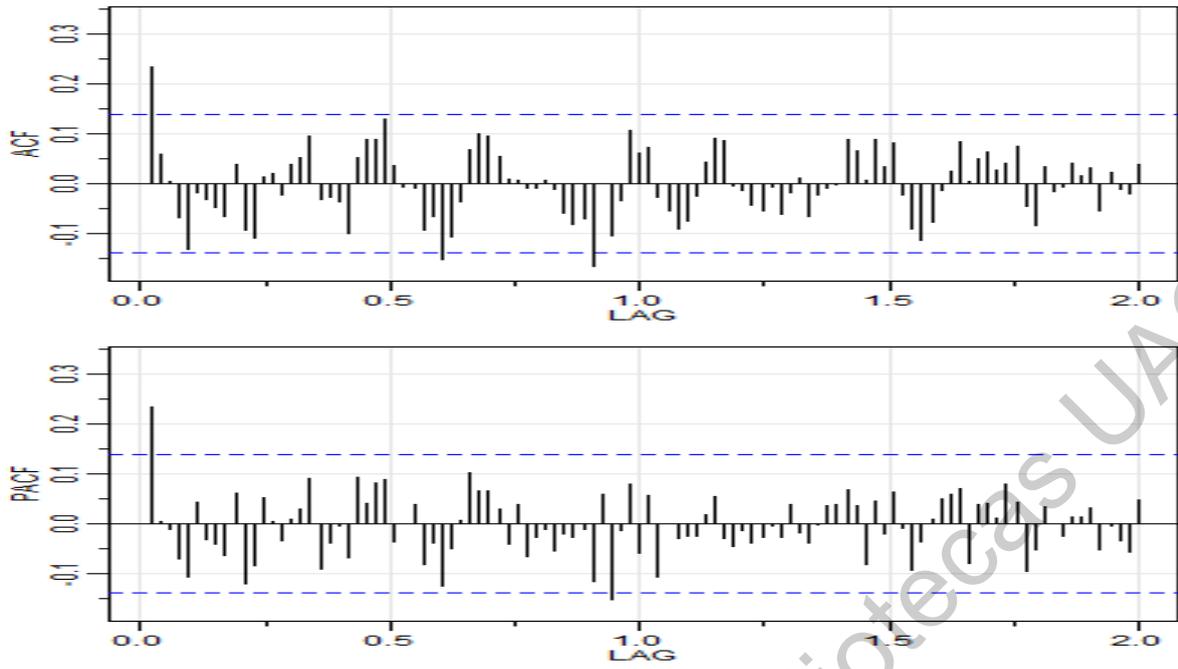


Figura G.5: Análisis de correlación del chile serrano

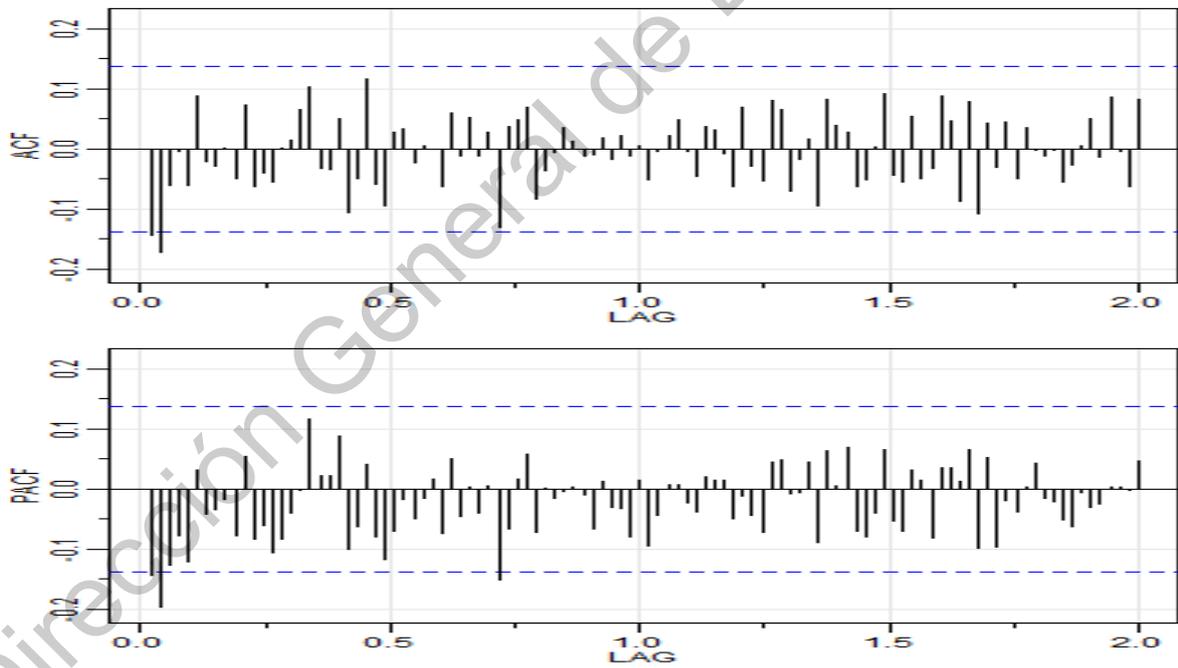


Figura G.6: Análisis de correlación de la coliflor grande

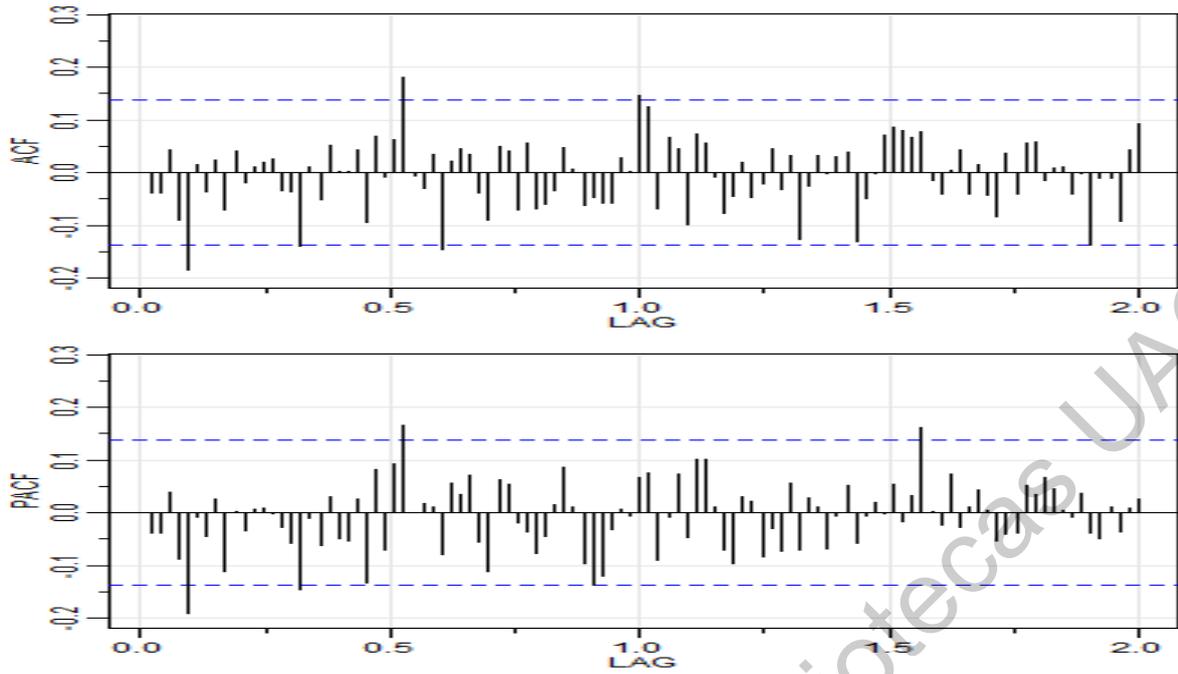


Figura G.7: Análisis de correlación del epazote

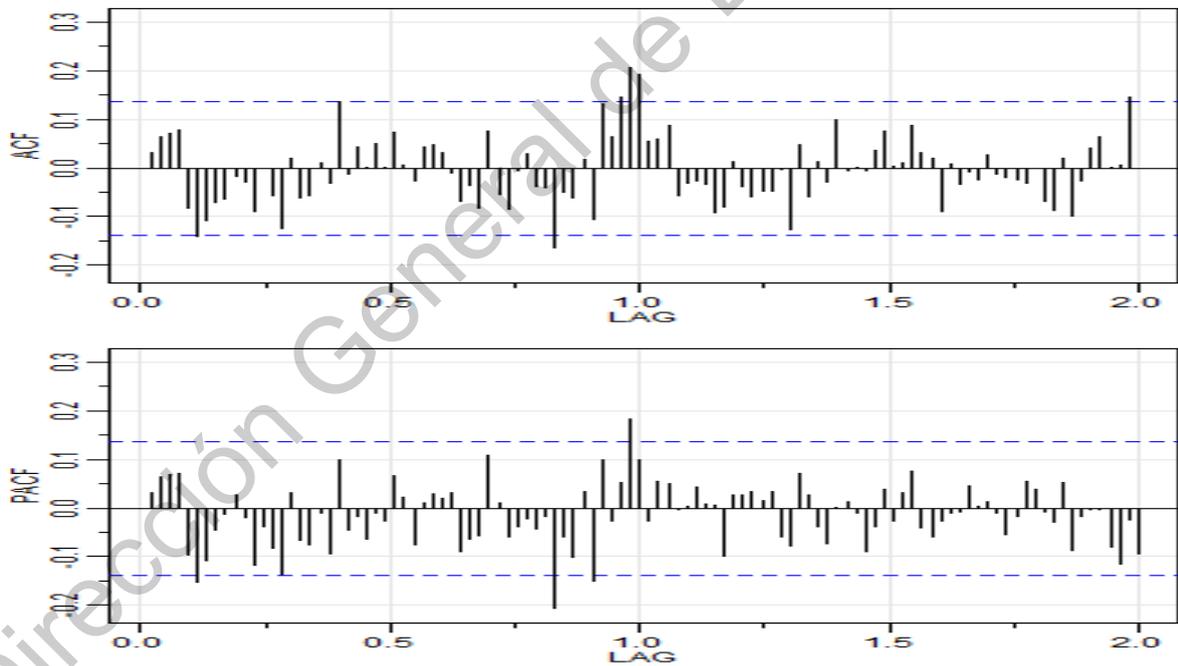


Figura G.8: Análisis de correlación de la jícama

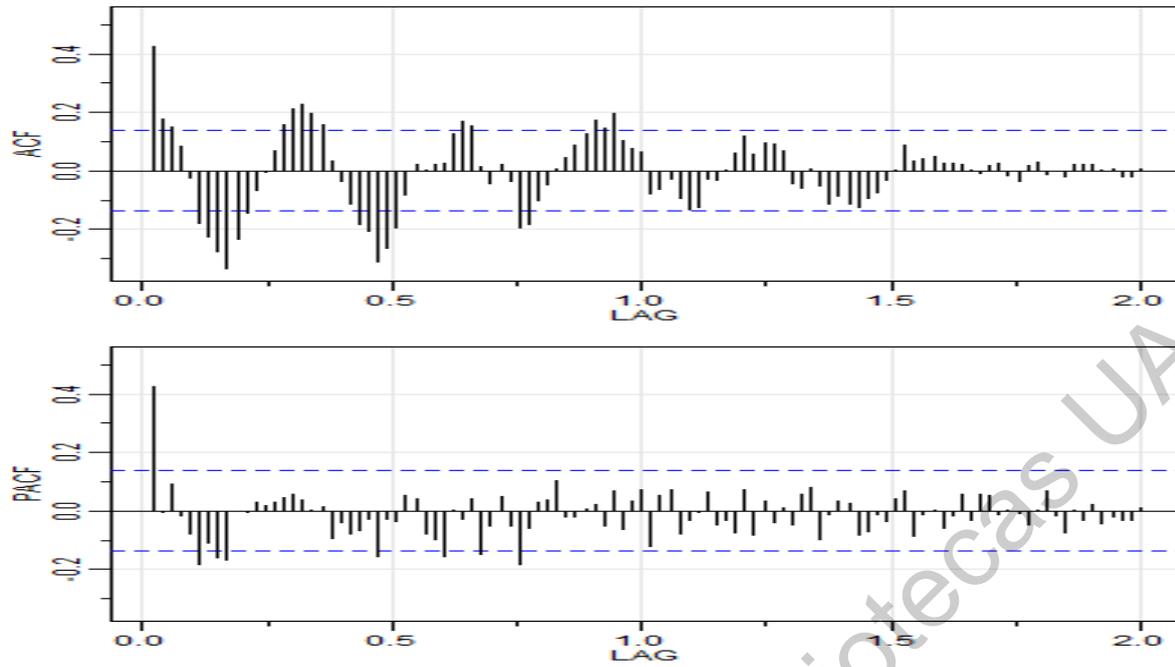


Figura G.9: Análisis de correlación del kiwi

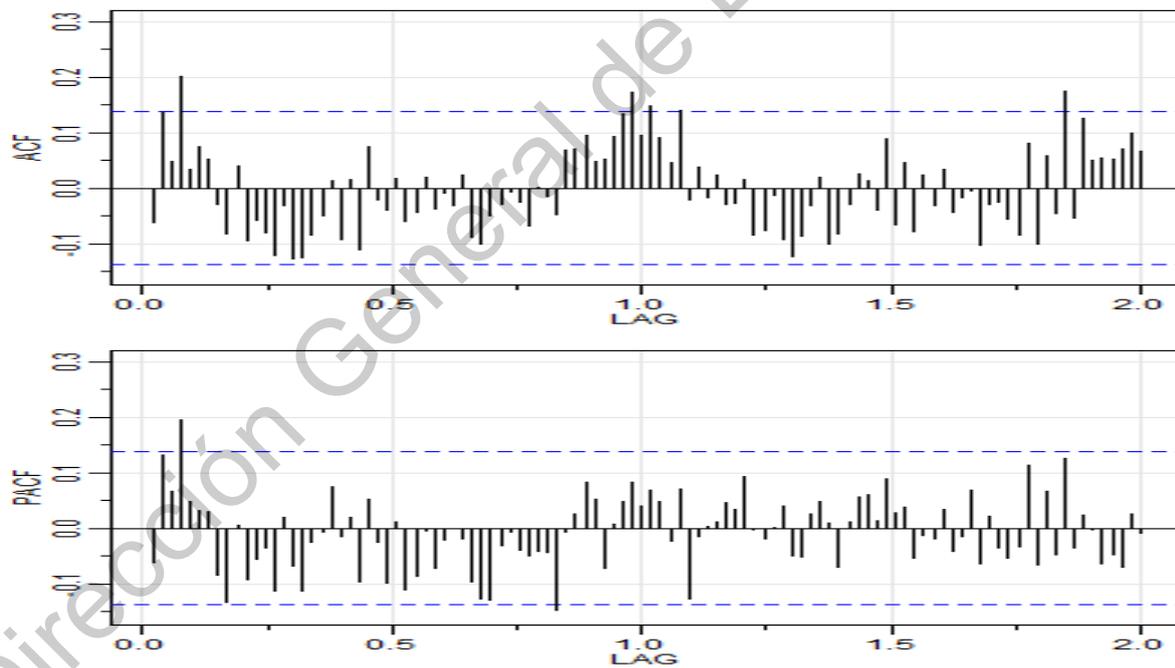


Figura G.10: Análisis de correlación de la lima

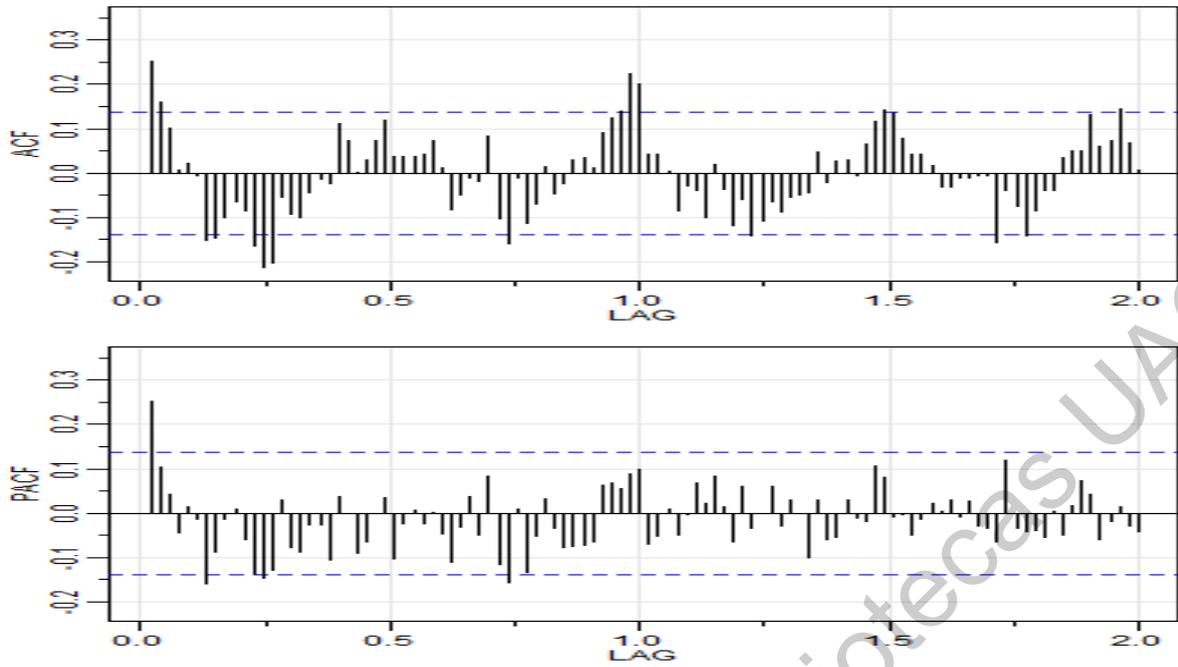


Figura G.11: Análisis de correlación del limón #5

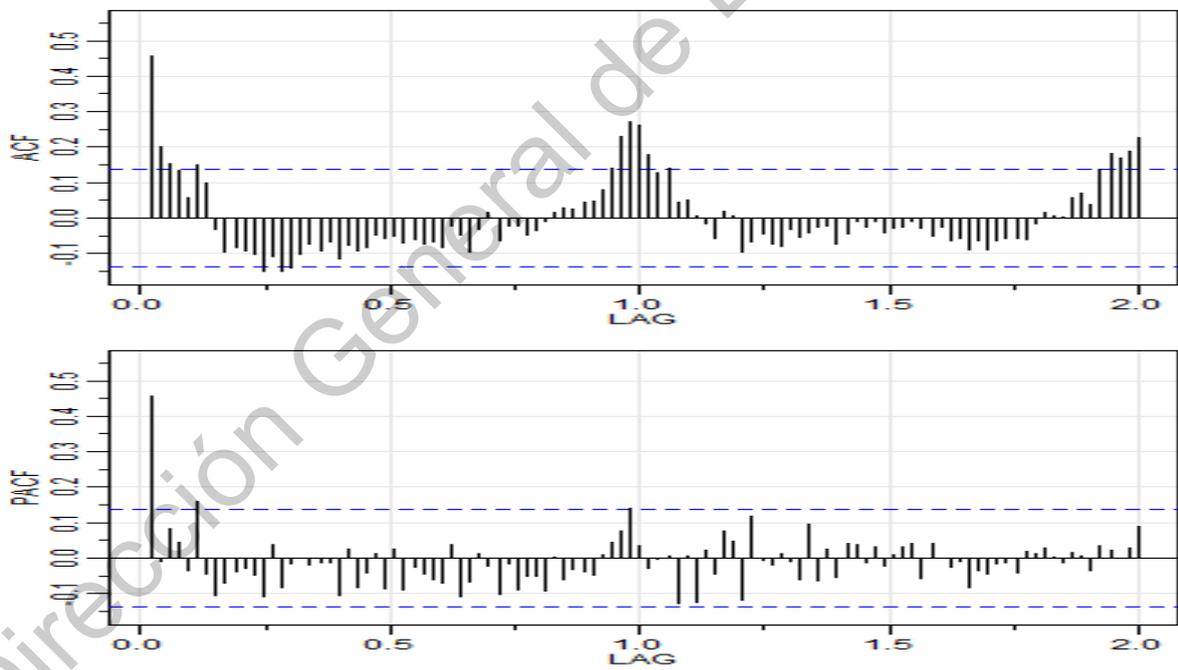


Figura G.12: Análisis de correlación de la naranja Valencia grande

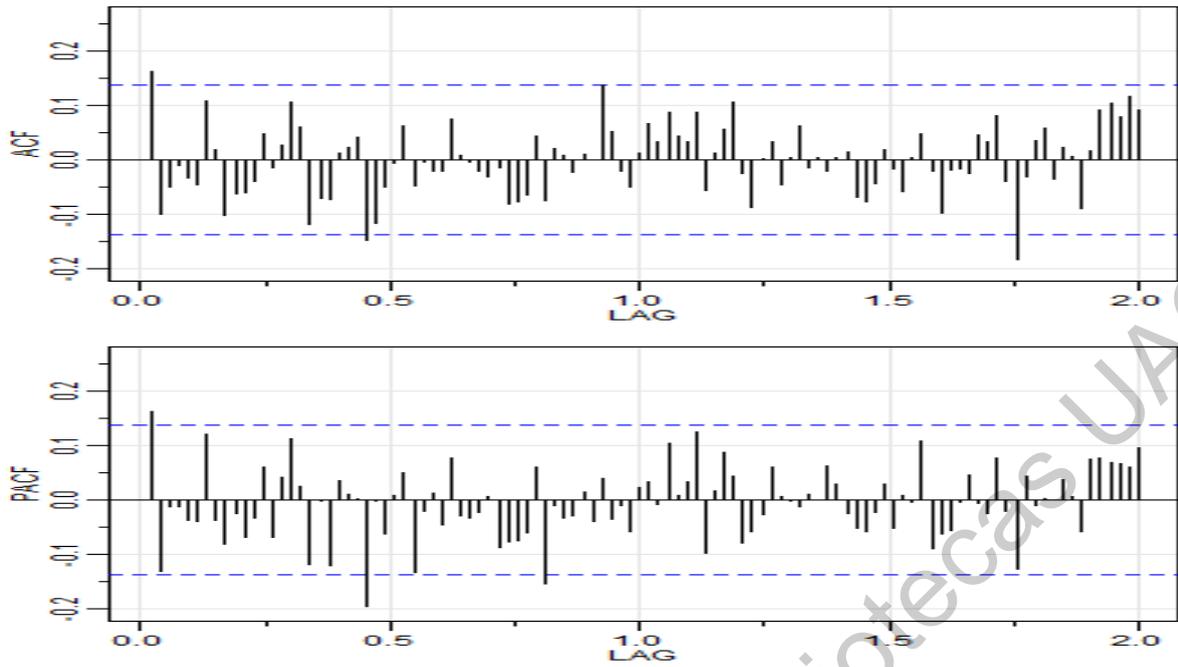


Figura G.13: Análisis de correlación del nopal grande

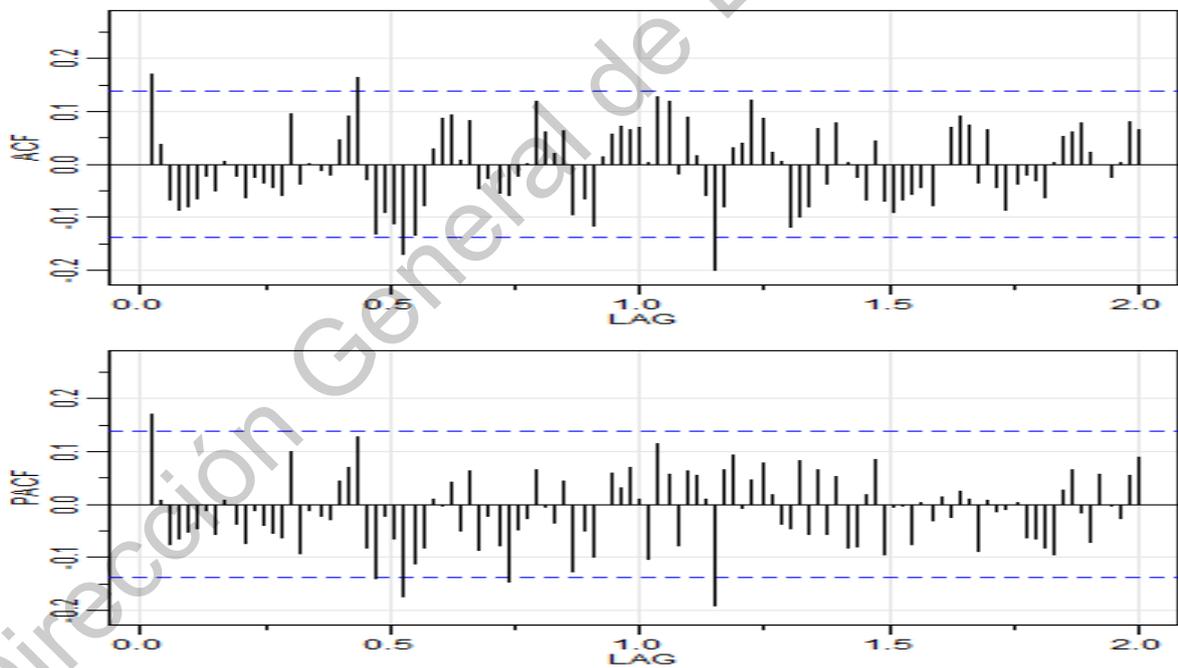


Figura G.14: Análisis de correlación de la piña mediana

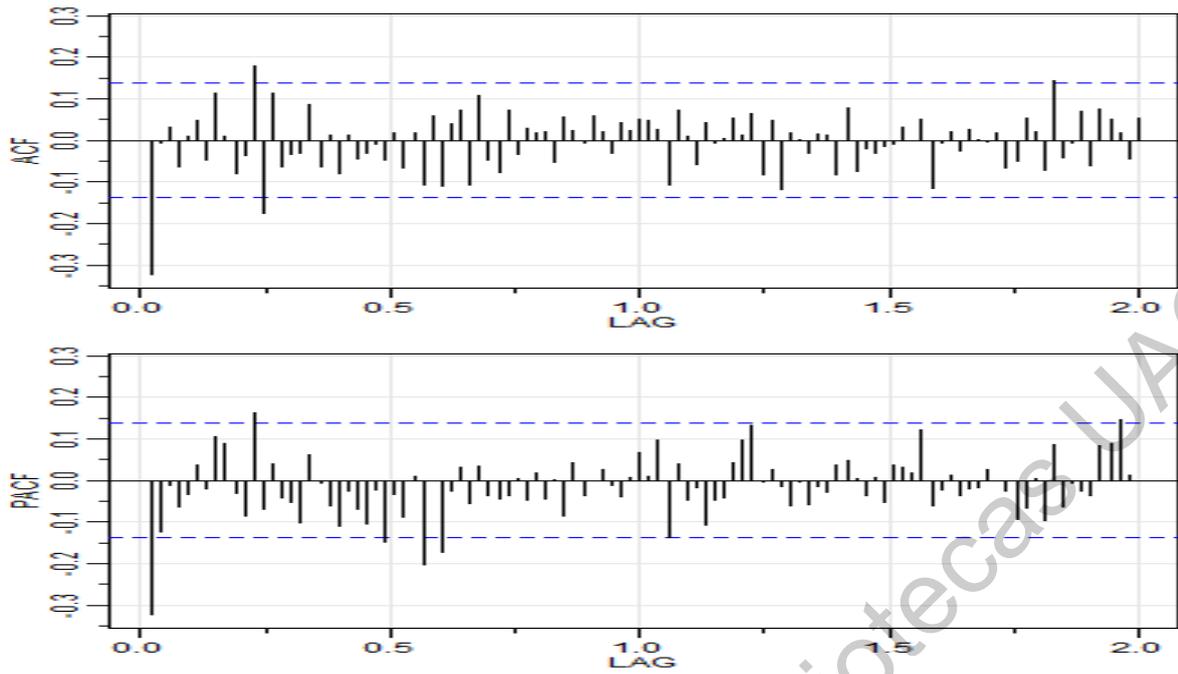


Figura G.15: Análisis de correlación del plátano macho

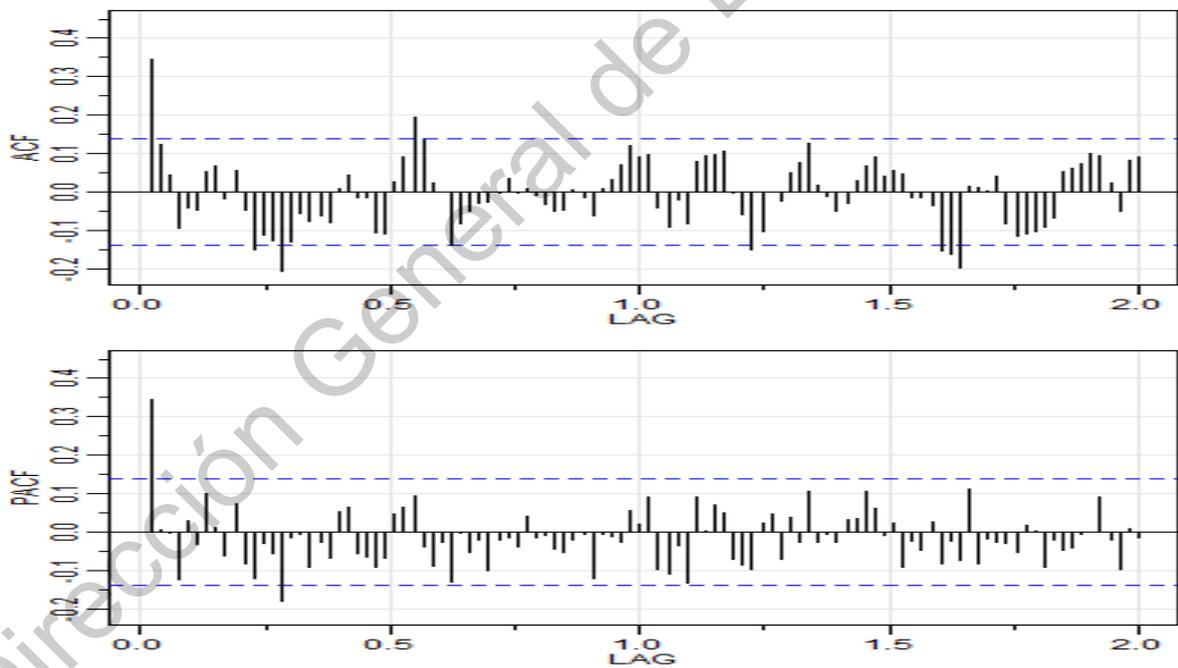


Figura G.16: Análisis de correlación del plátano Tabasco

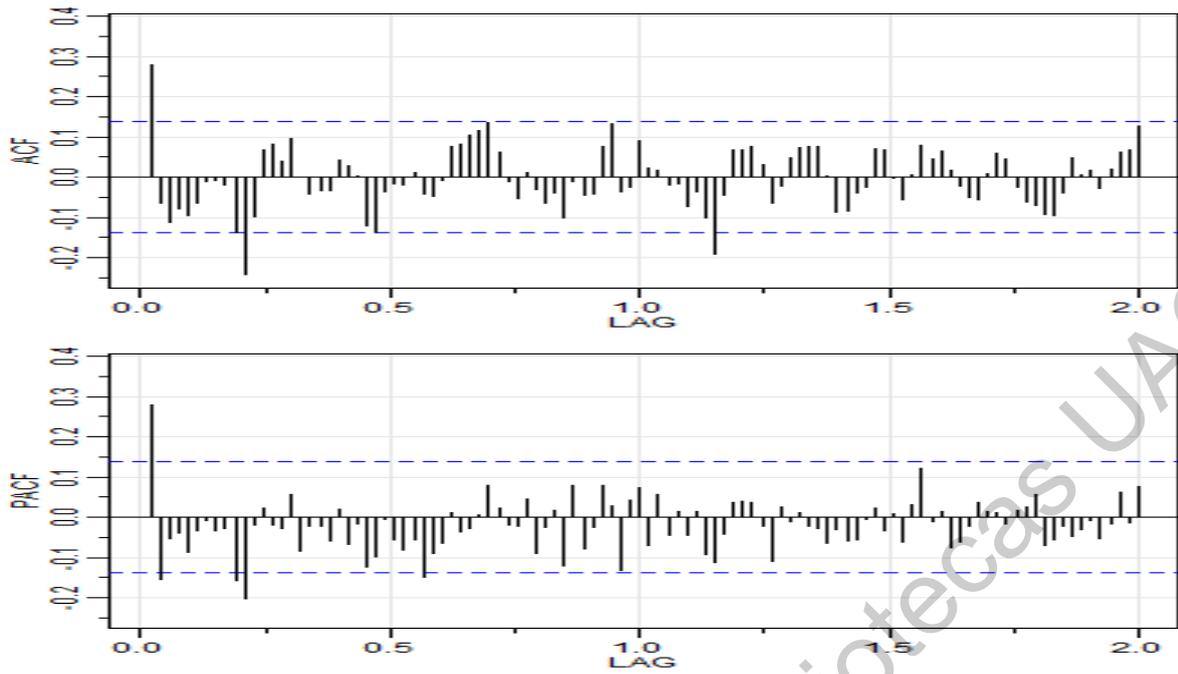


Figura G.17: Análisis de correlación del tomate saladette

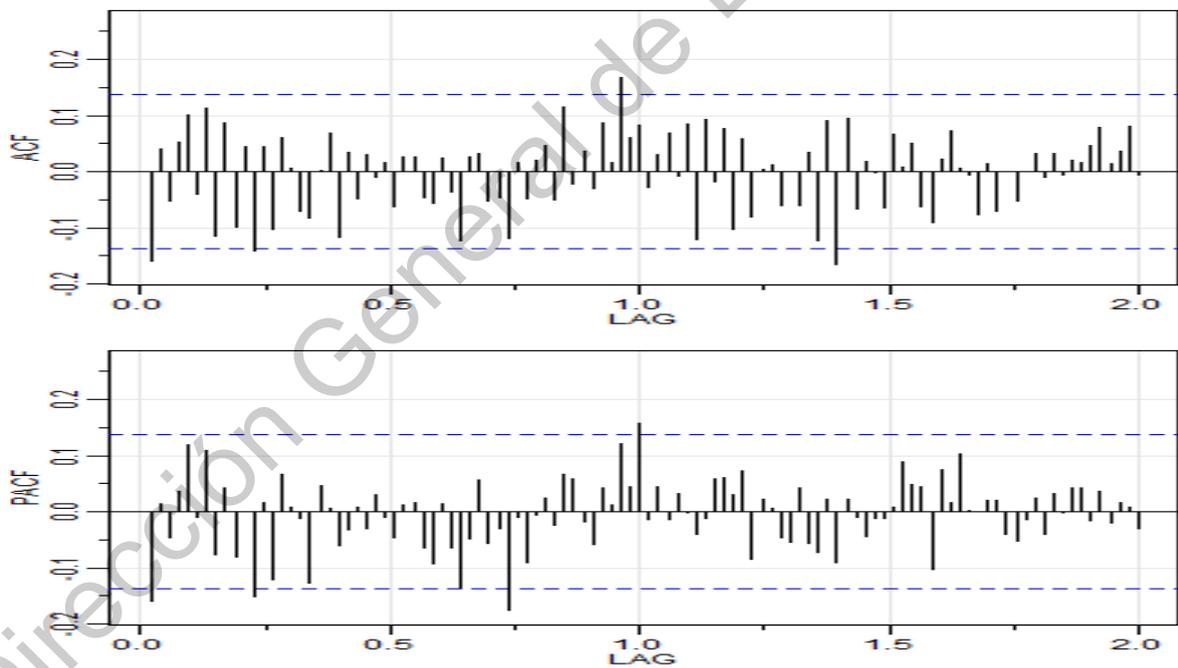


Figura G.18: Análisis de correlación de la toronja roja

## APÉNDICE H

### Análisis de los residuales

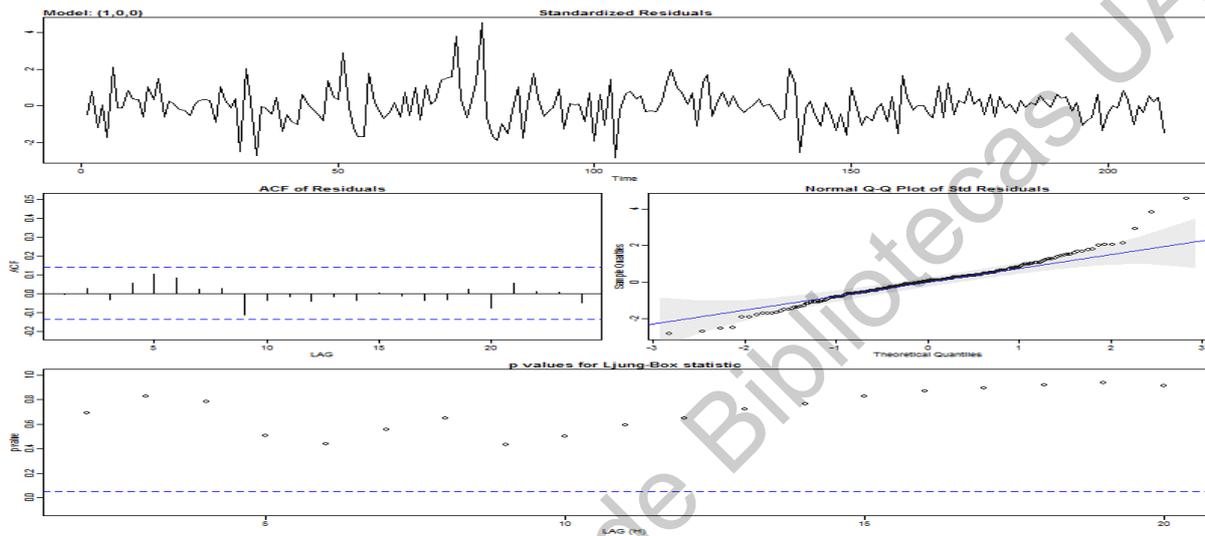


Figura H.1: Aguacate Hass

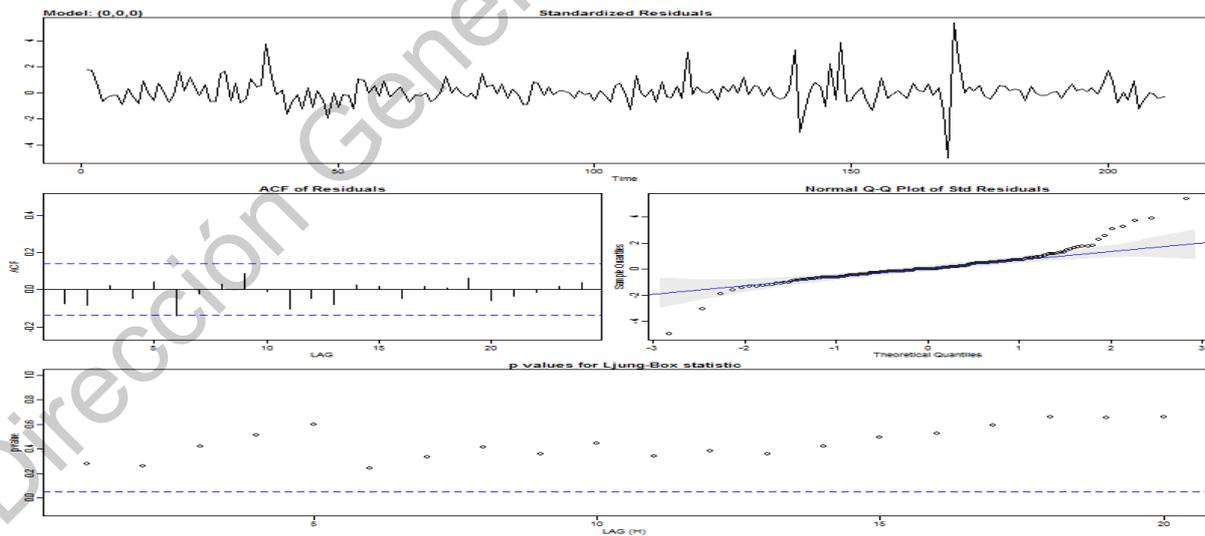


Figura H.2: Cacahuete

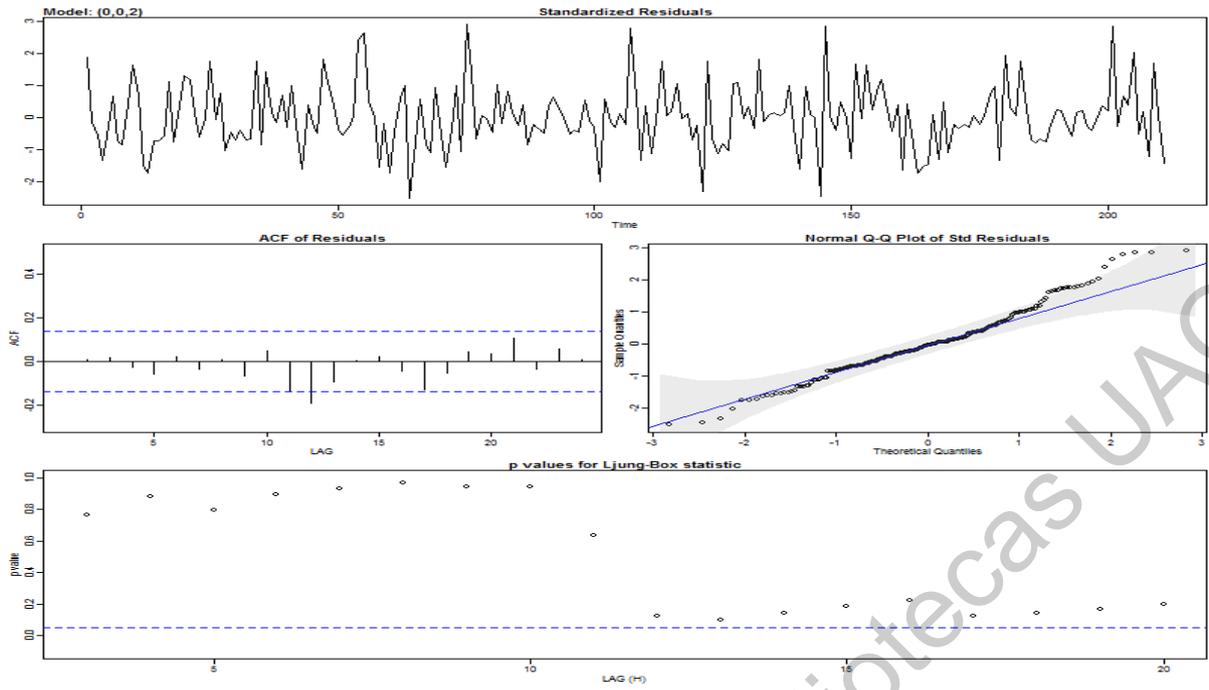


Figura H.3: Calabacita italiana

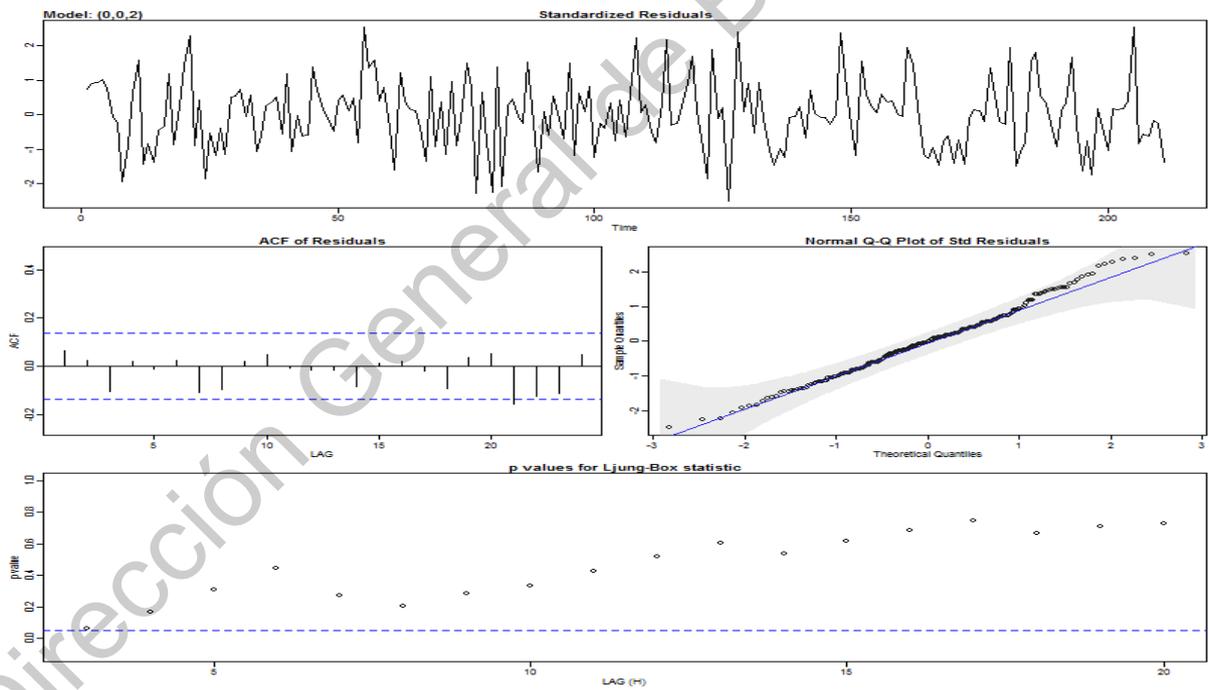


Figura H.4: Chayote sin espinas

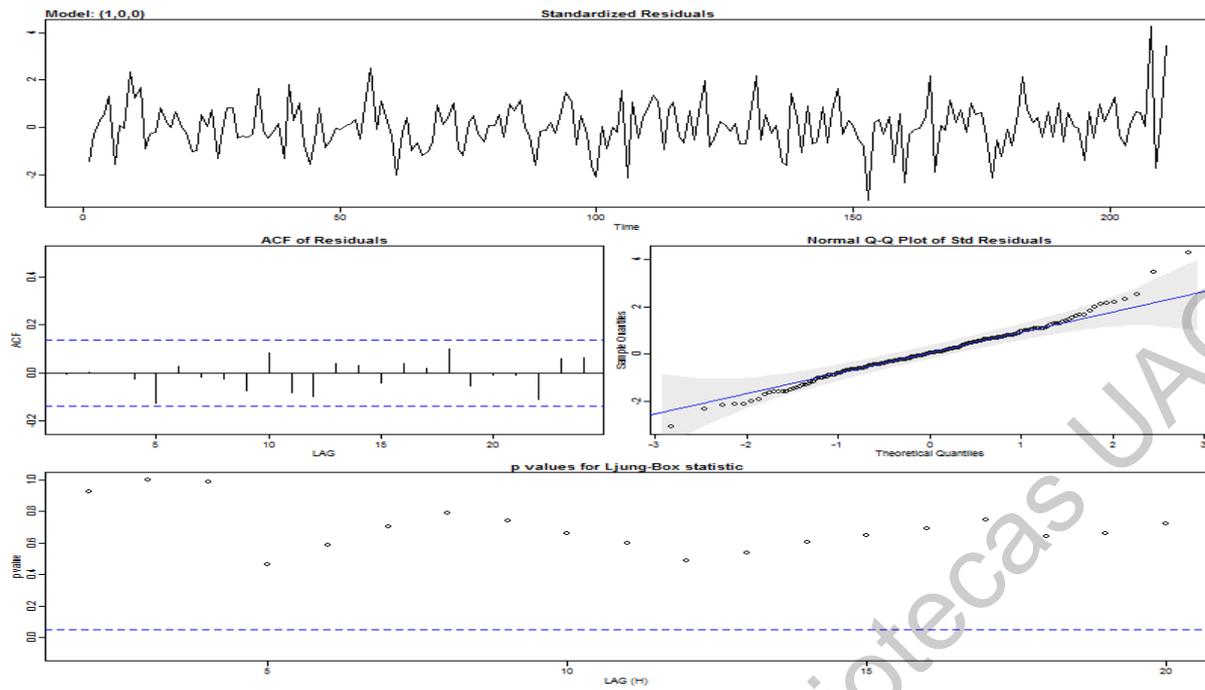


Figura H.5: Chile serrano

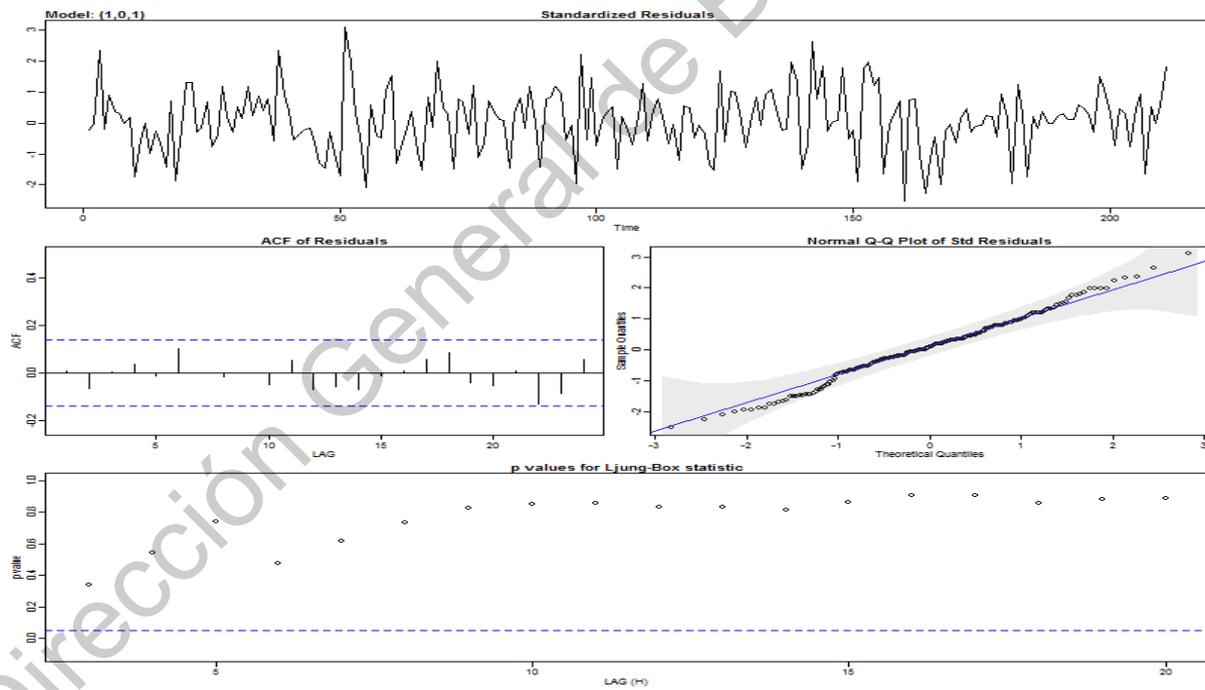


Figura H.6: Coliflor grande

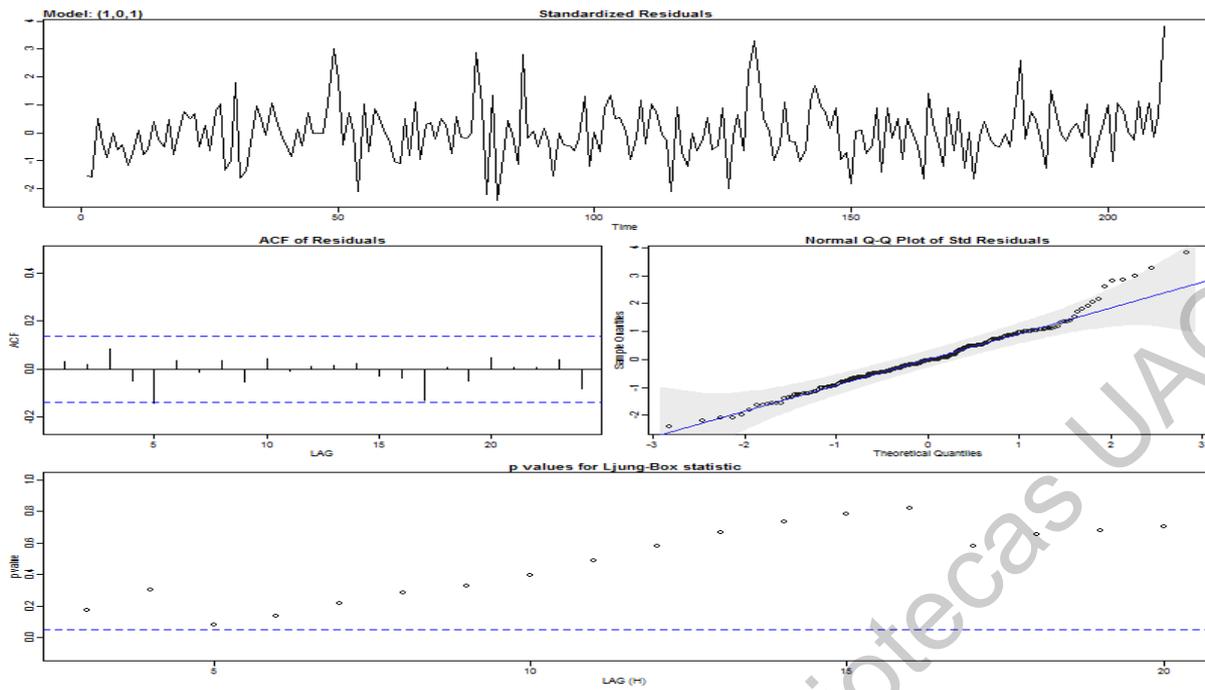


Figura H.7: Epazote

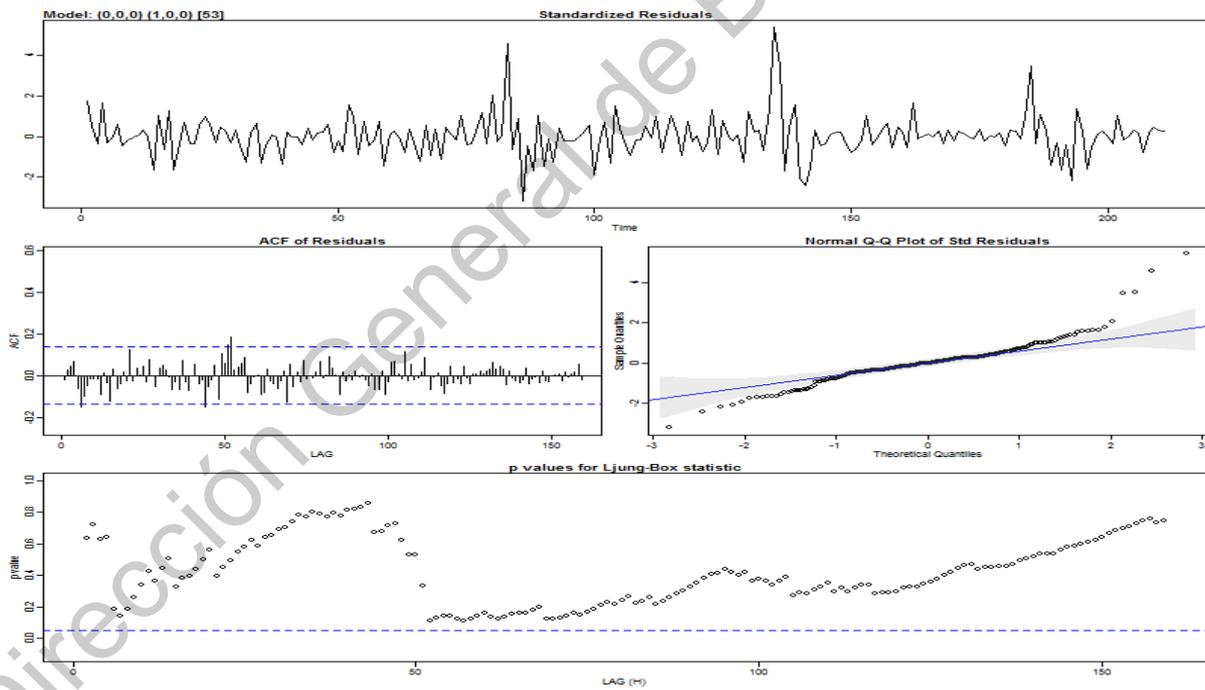


Figura H.8: Jícama

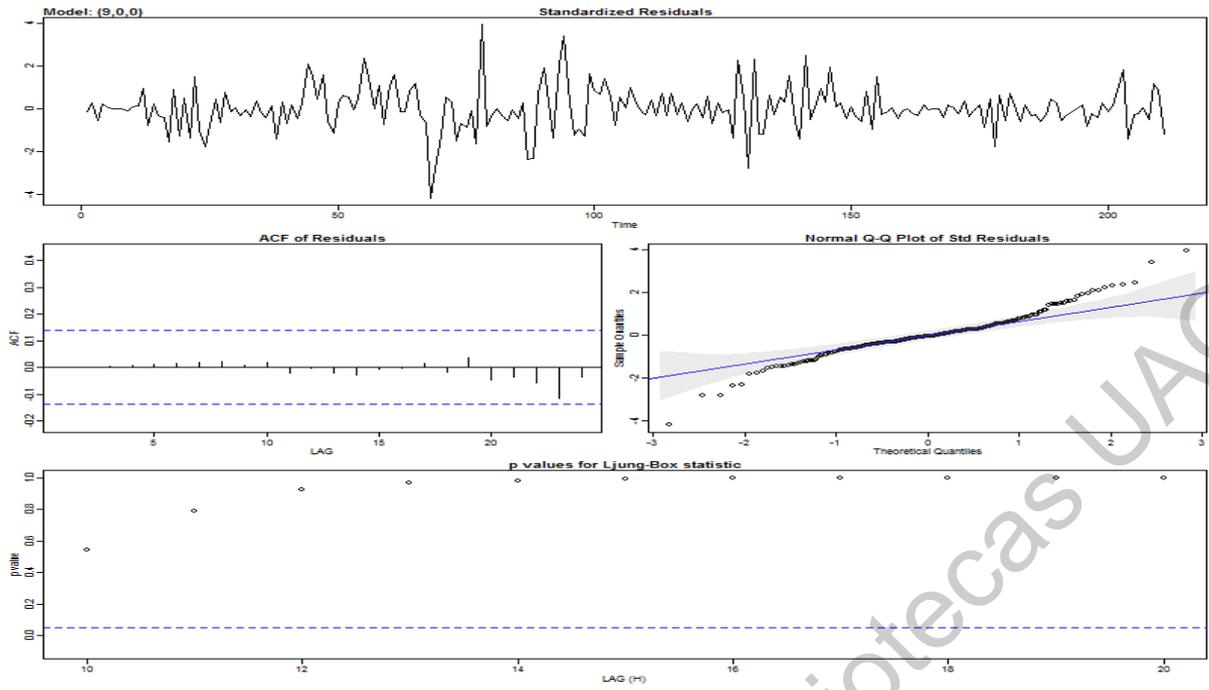


Figura H.9: Kiwi

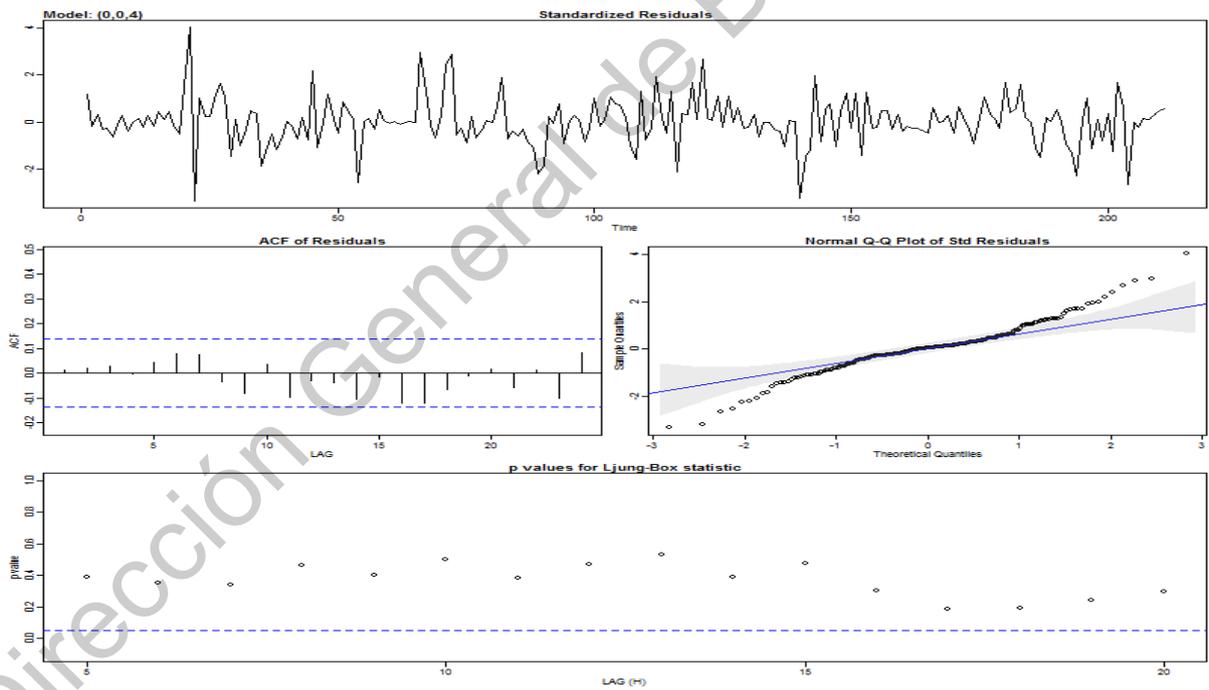


Figura H.10: Lima

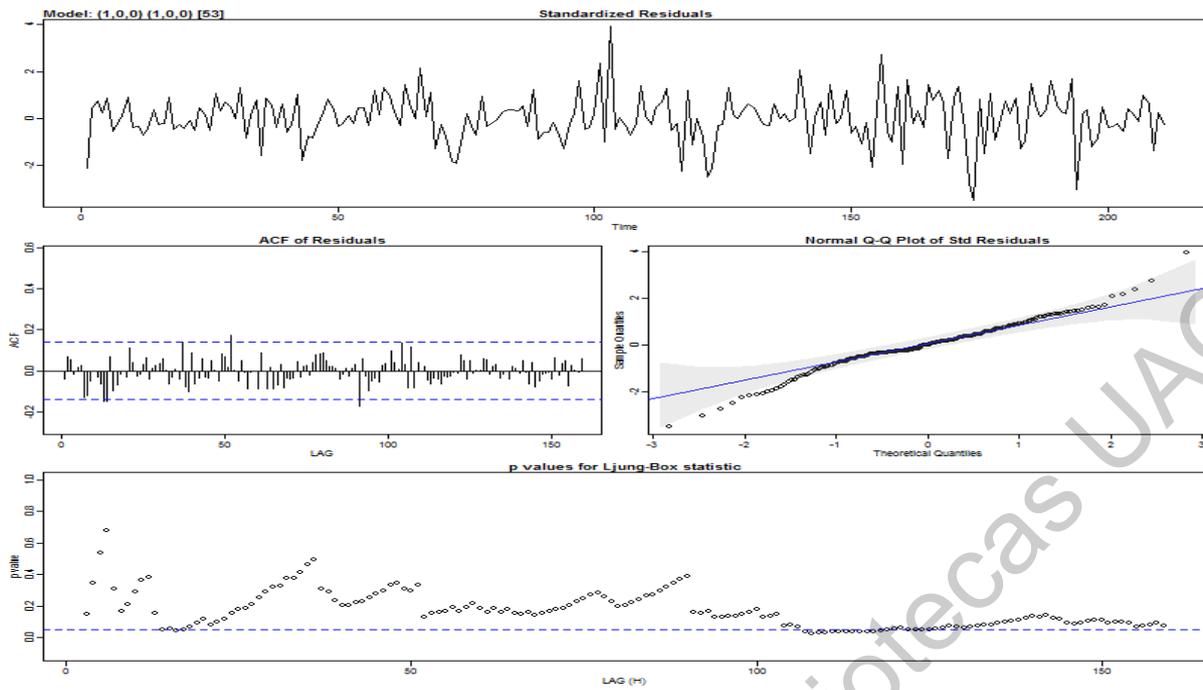


Figura H.11: Limón con semilla #5

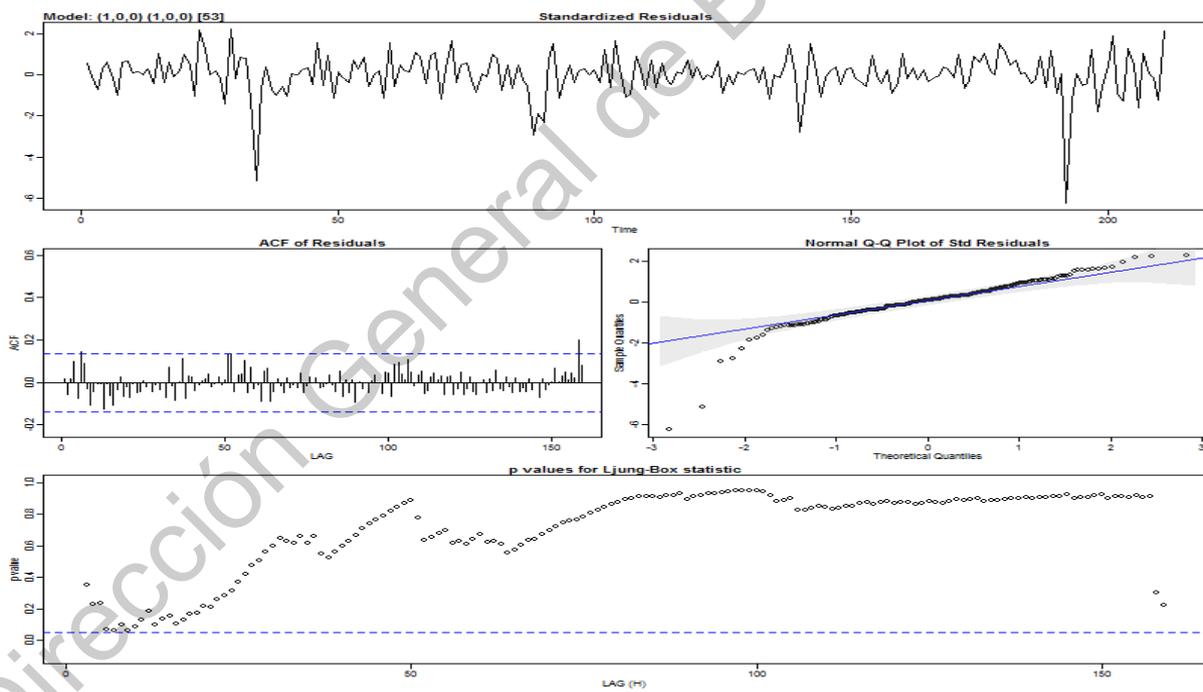


Figura H.12: Naranja Valencia grande

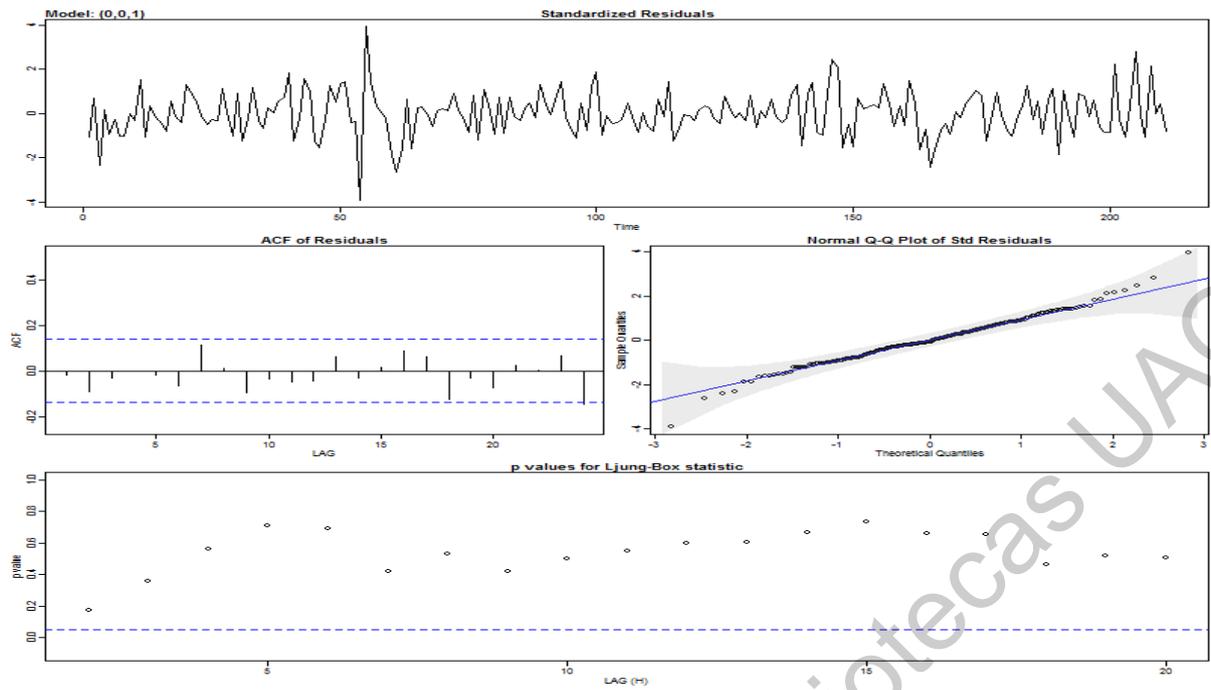


Figura H.13: Nopal grande

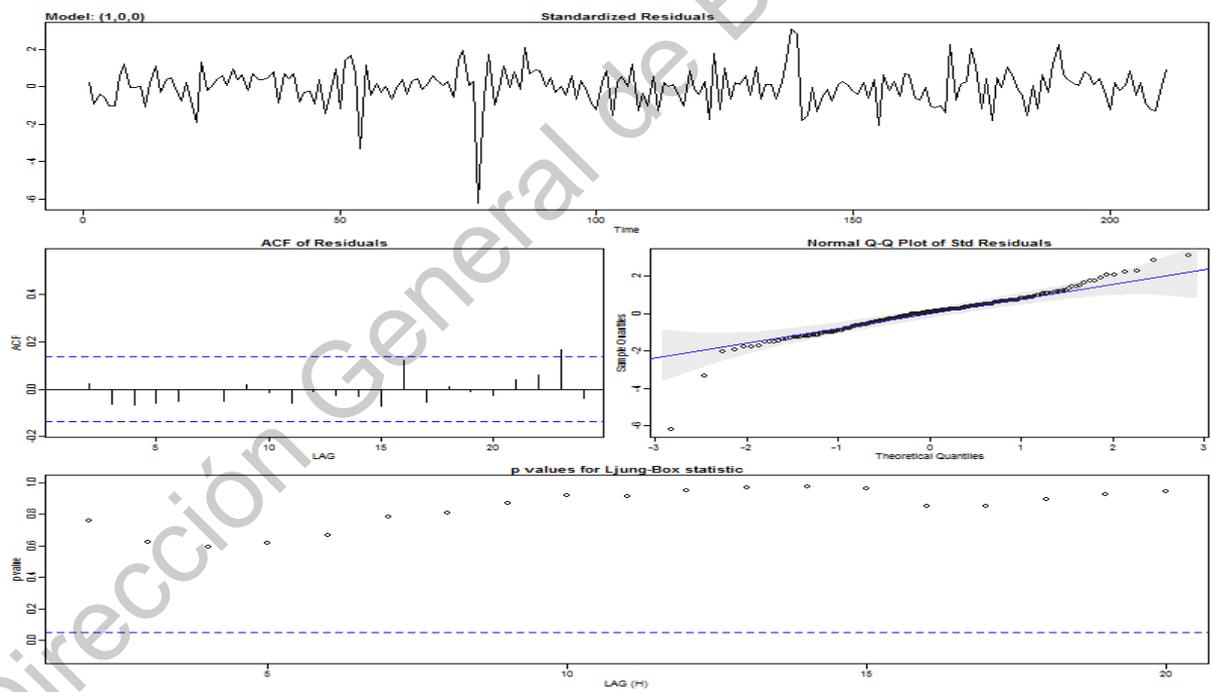


Figura H.14: Piña mediana

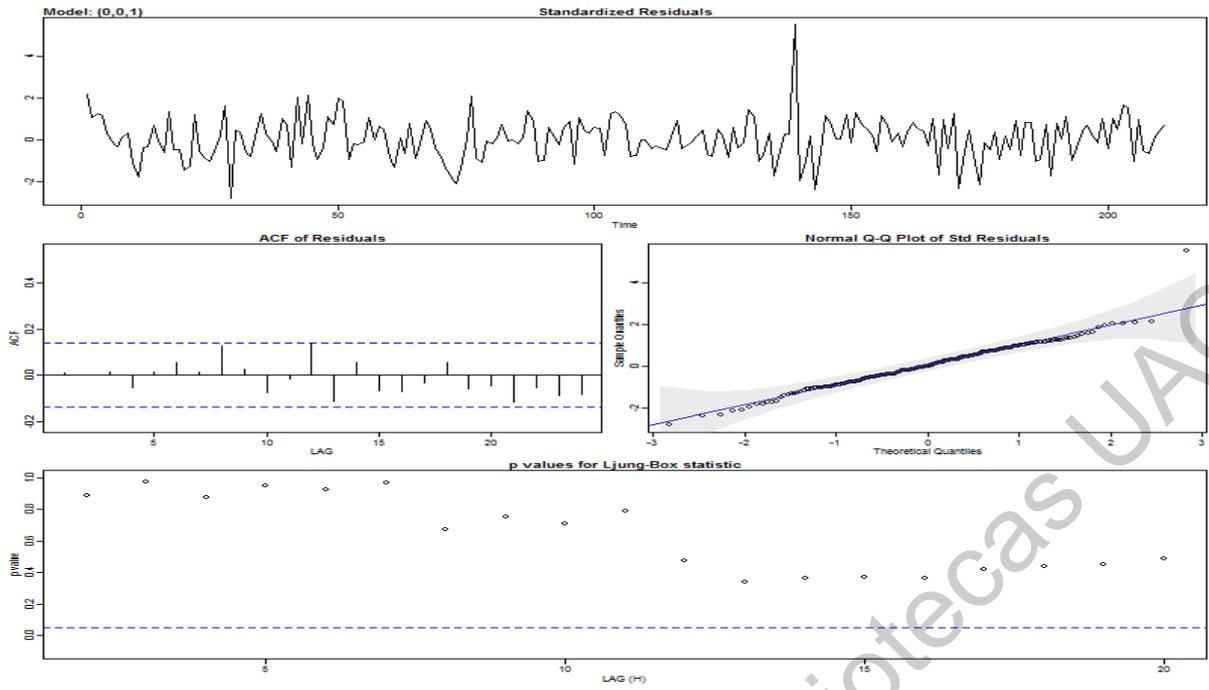


Figura H.15: Plátano macho

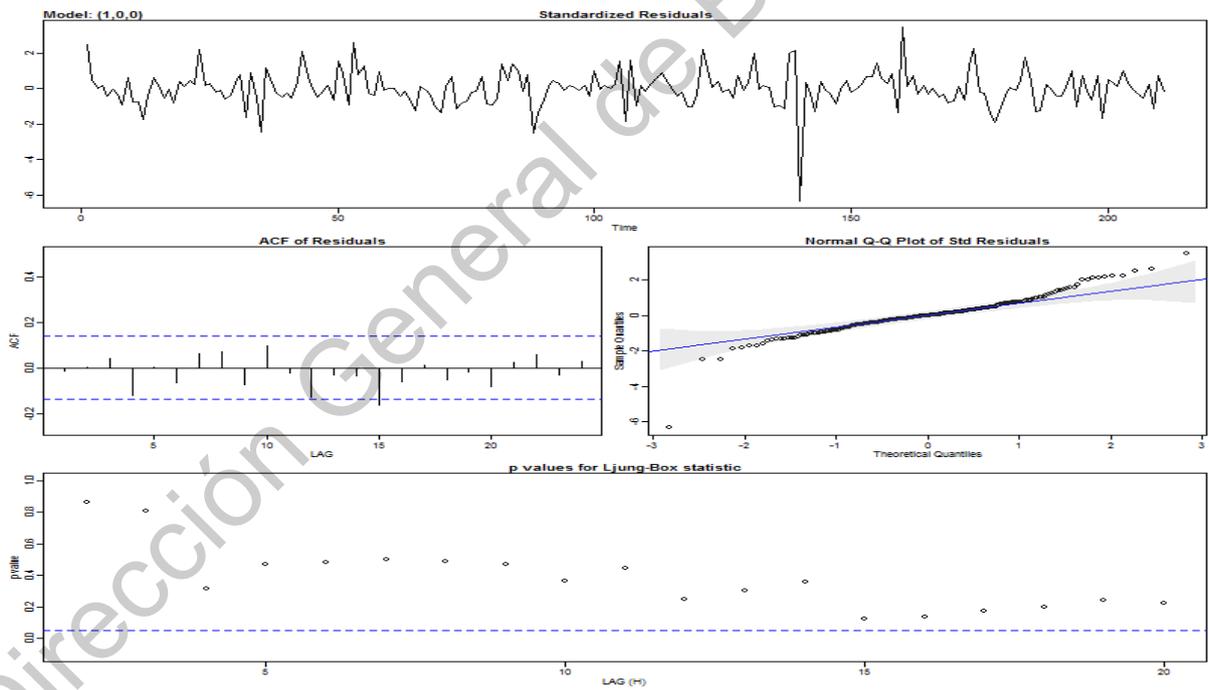


Figura H.16: Plátano Tabasco

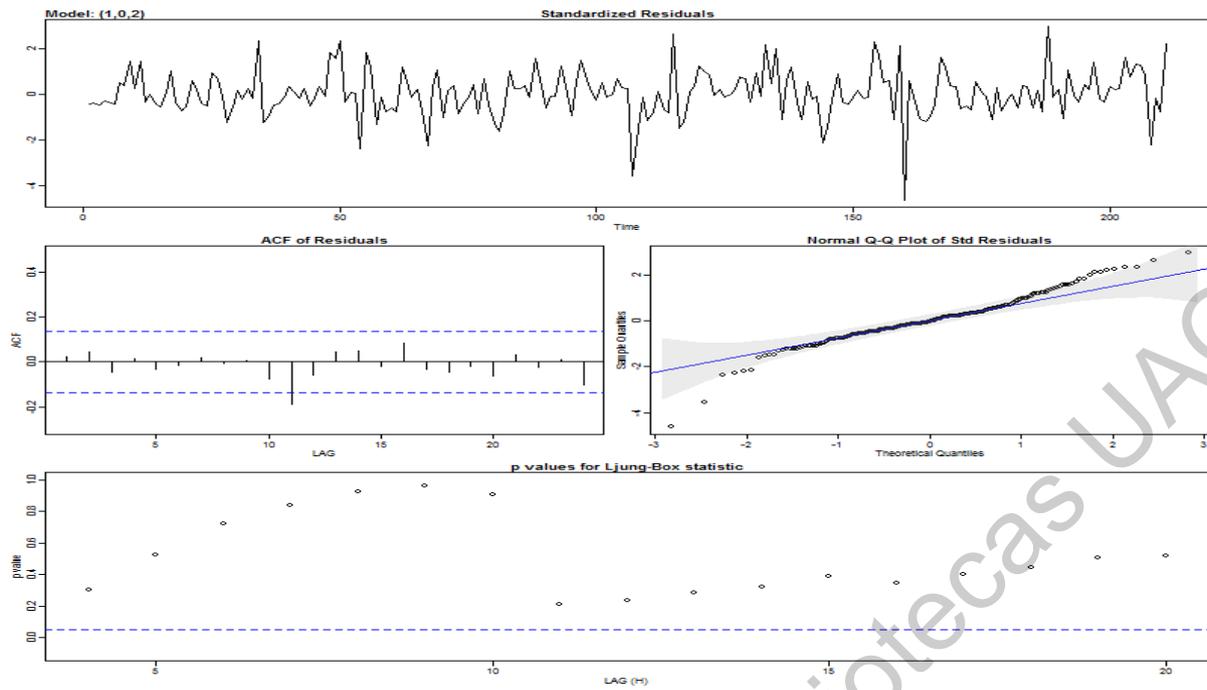


Figura H.17: Tomate saladette

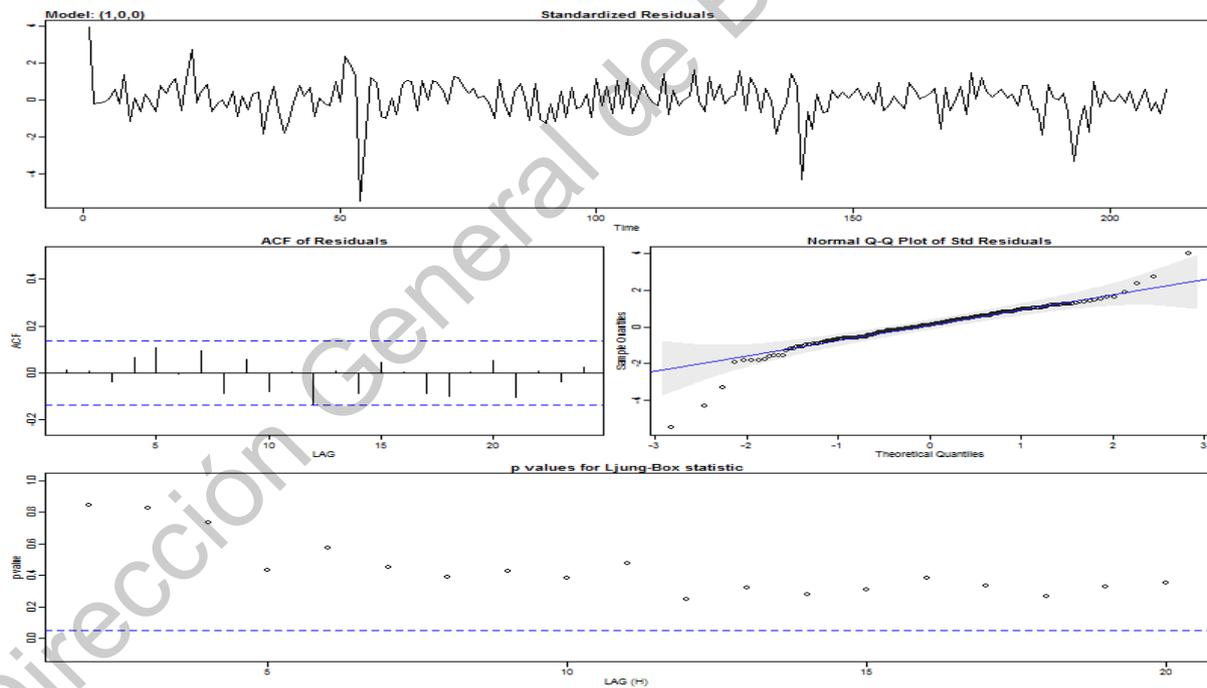
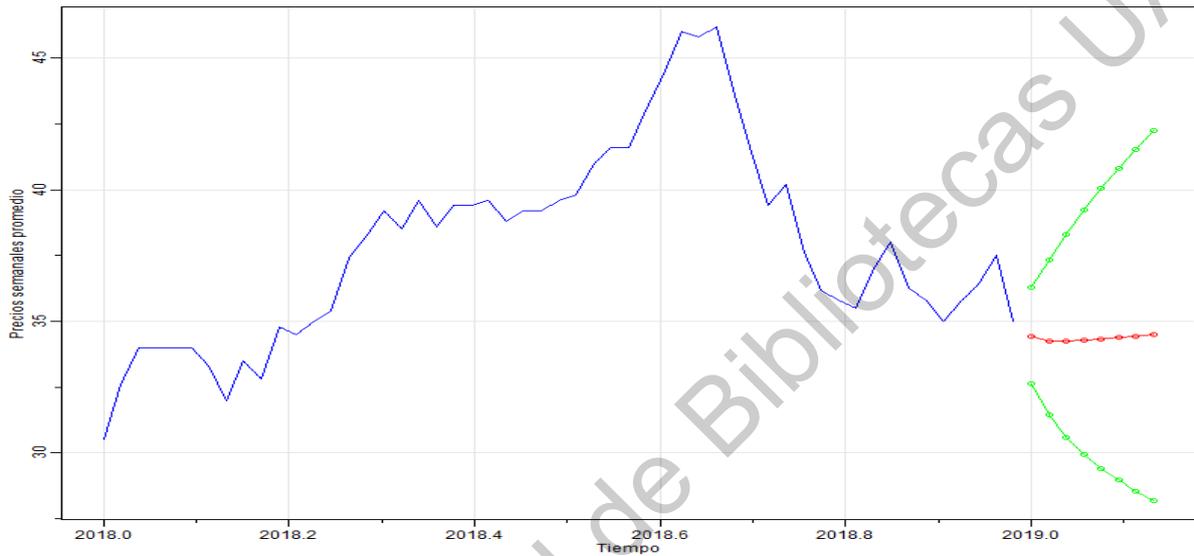
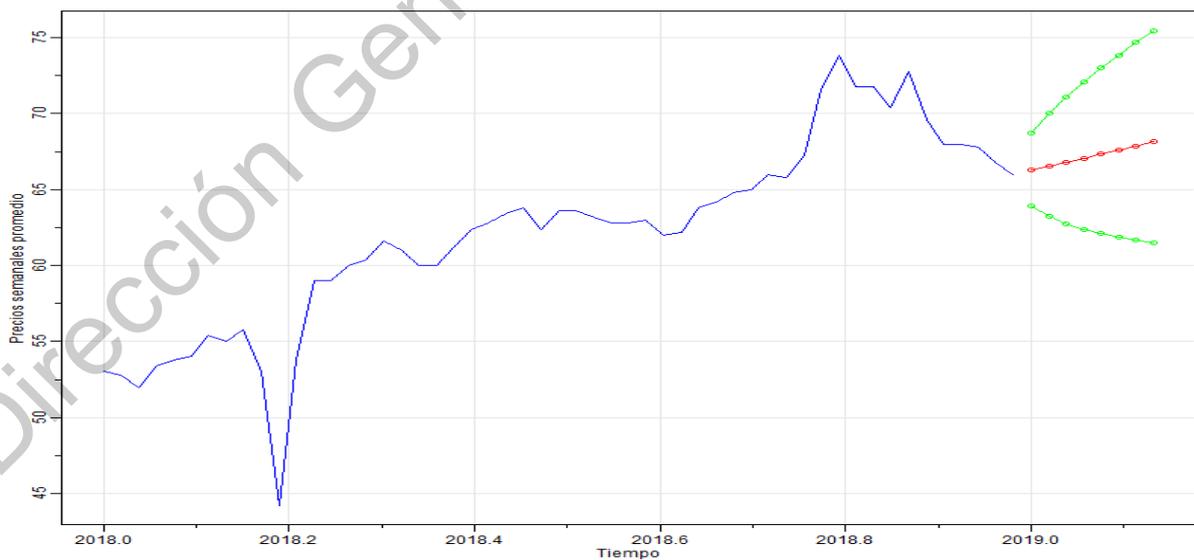


Figura H.18: Toronja roja

## APÉNDICE I

## Predicciones de las series de tiempo

*Figura I.1: Predicciones del aguacate Hass**Figura I.2: Predicciones del cacahuate*

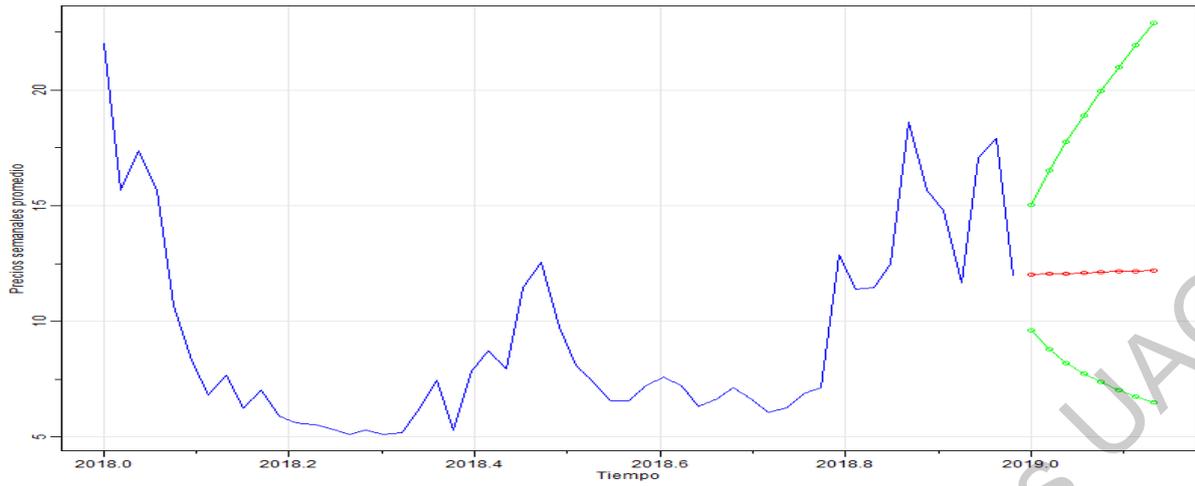


Figura I.3: Predicciones de la calabacita italiana

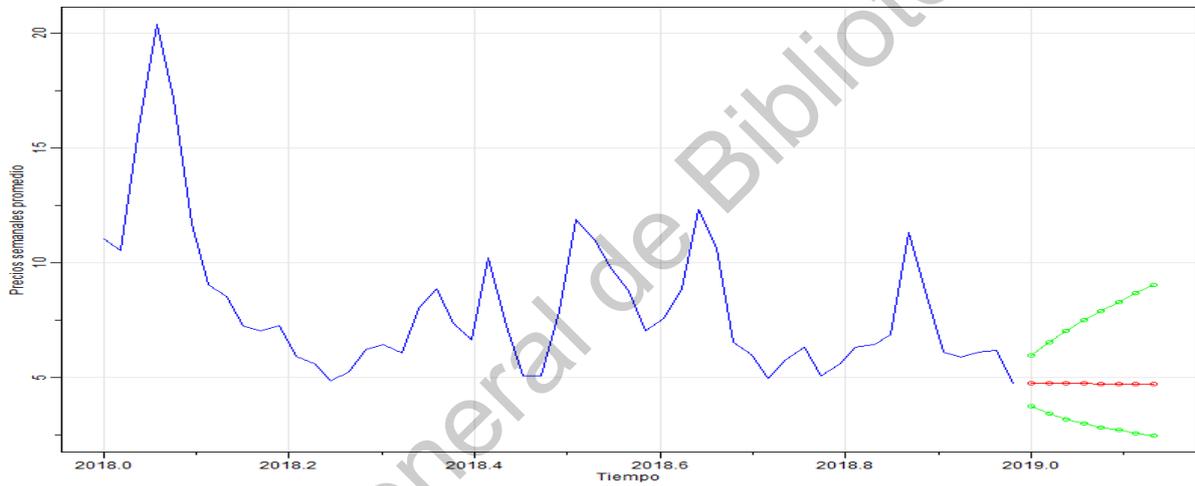


Figura I.4: Predicciones del chayote sin espinas

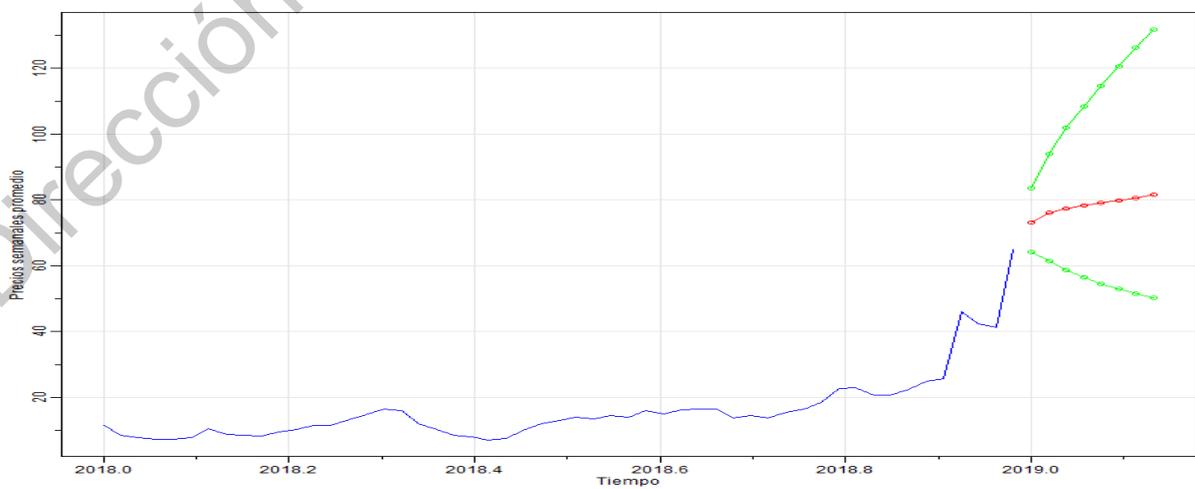


Figura I.5: Predicciones del chile serrano

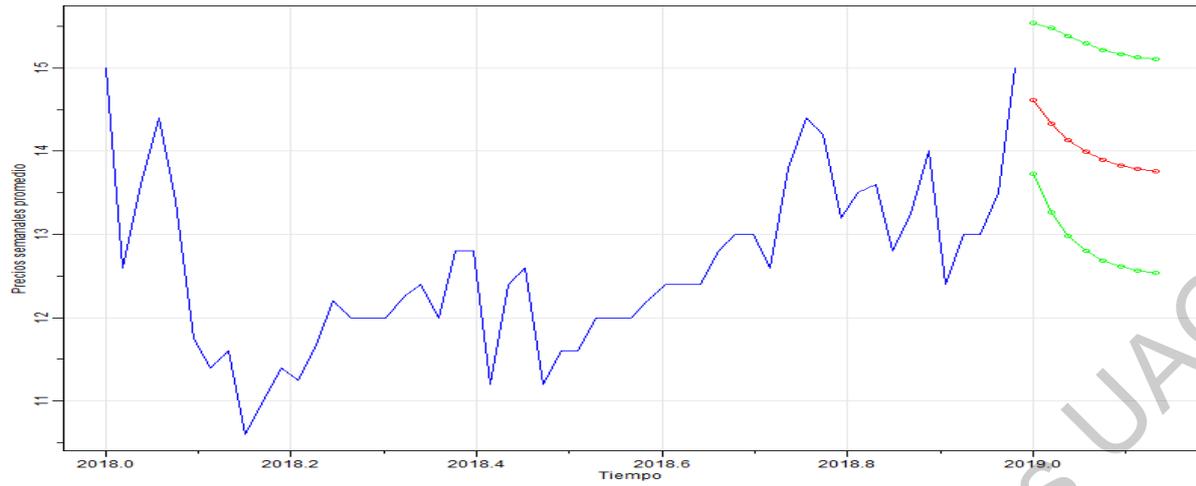


Figura I.6: Predicciones de la coliflor grande

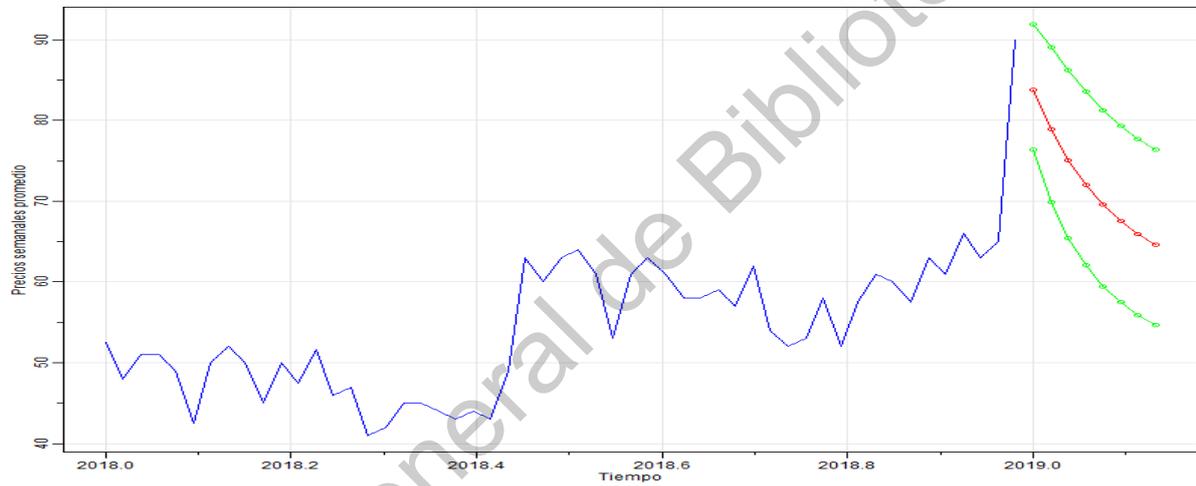


Figura I.7: Predicciones del epazote

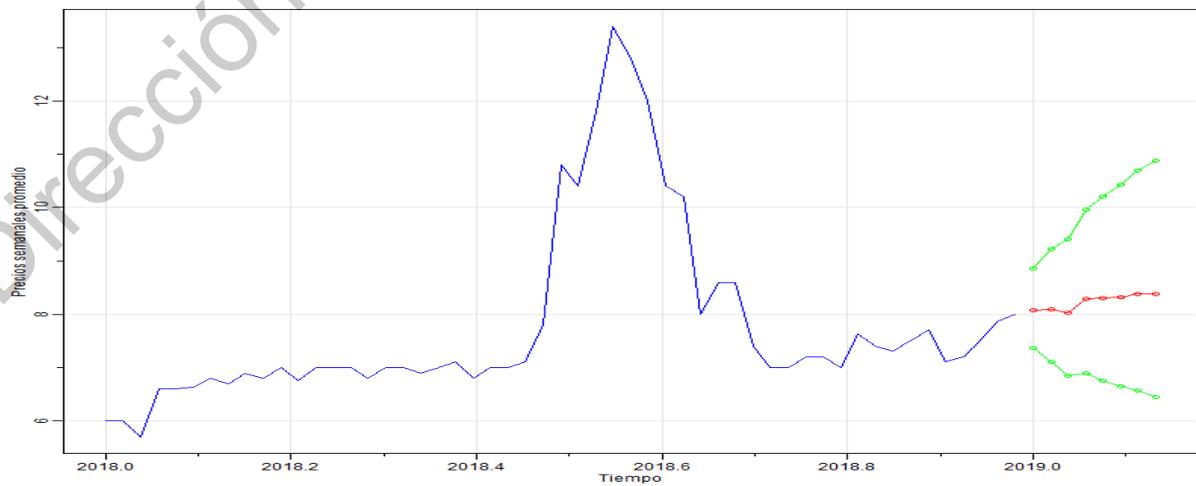


Figura I.8: Predicciones de la jícama

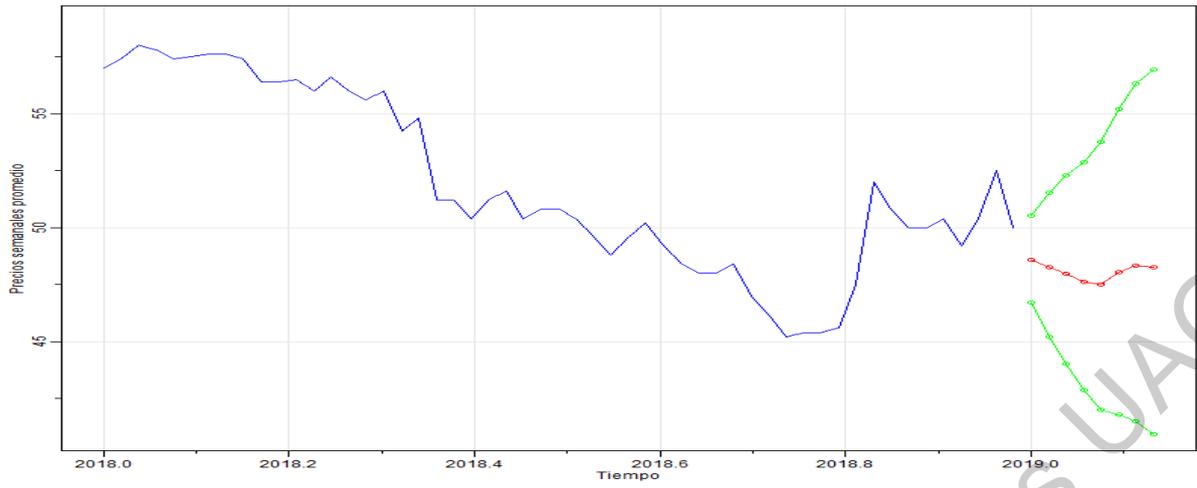


Figura I.9: Predicciones del kiwi

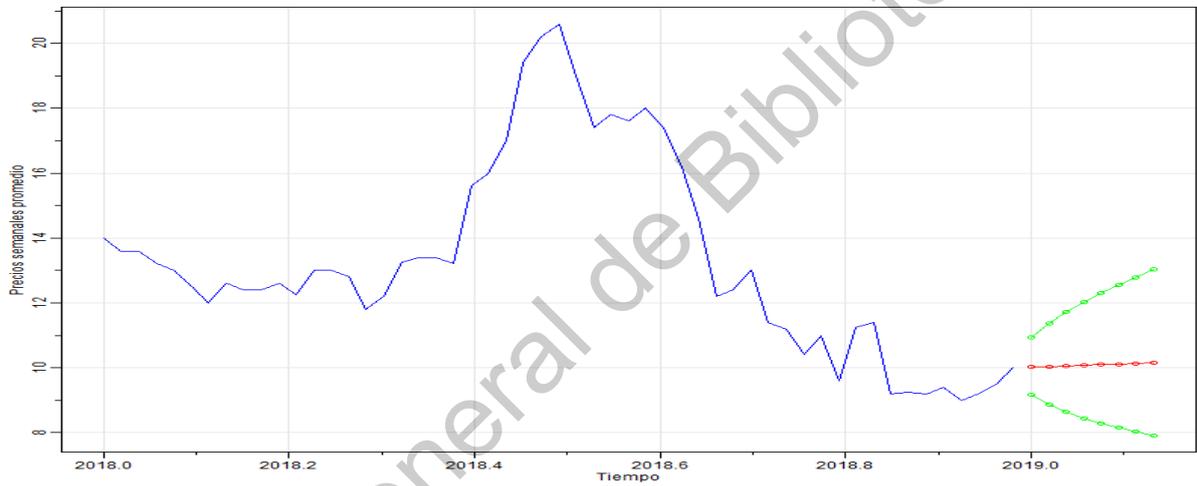


Figura I.10: Predicciones de la lima

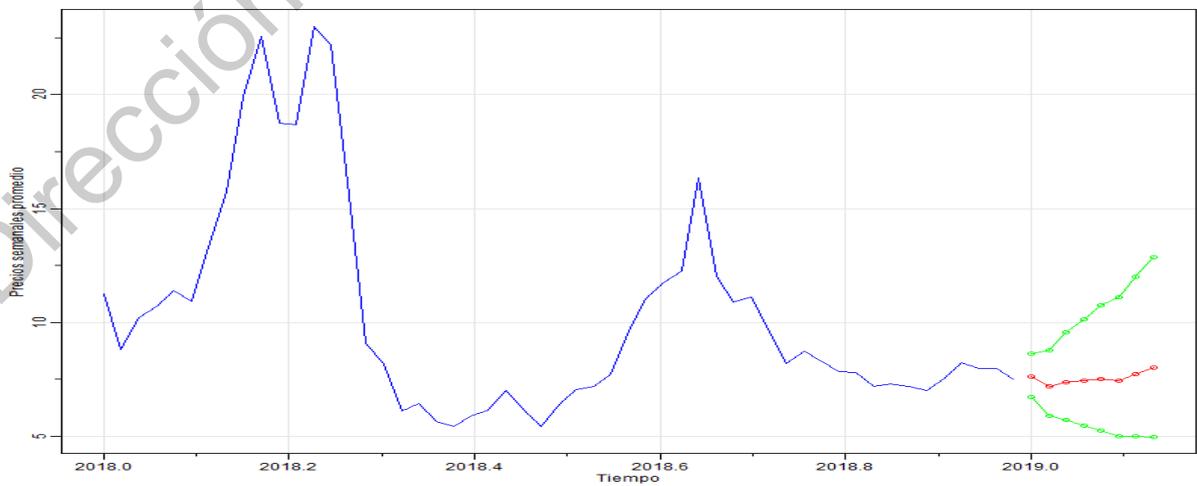


Figura I.11: Predicciones del limón #5

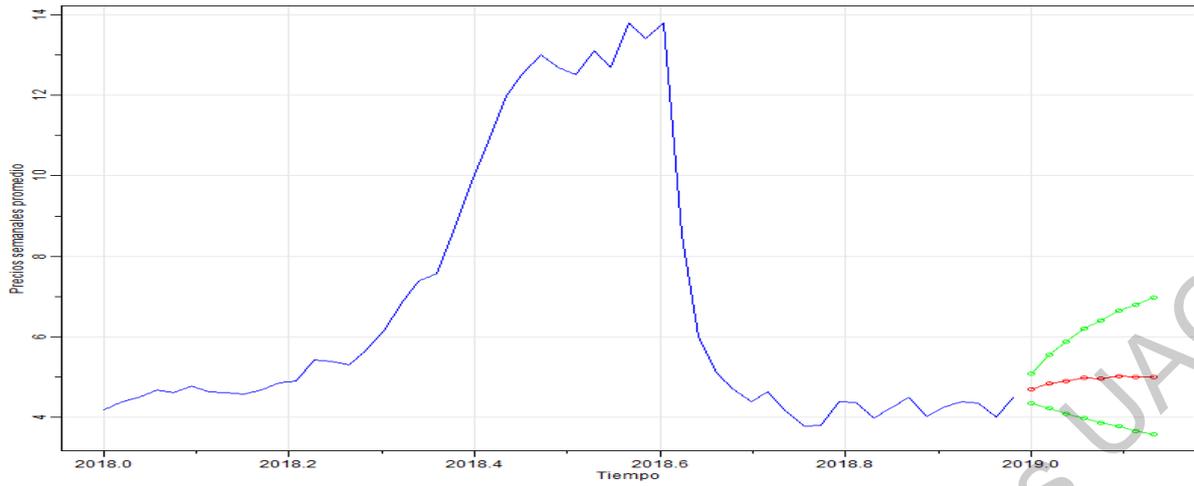


Figura I.12: Predicciones de la naranja Valencia grande

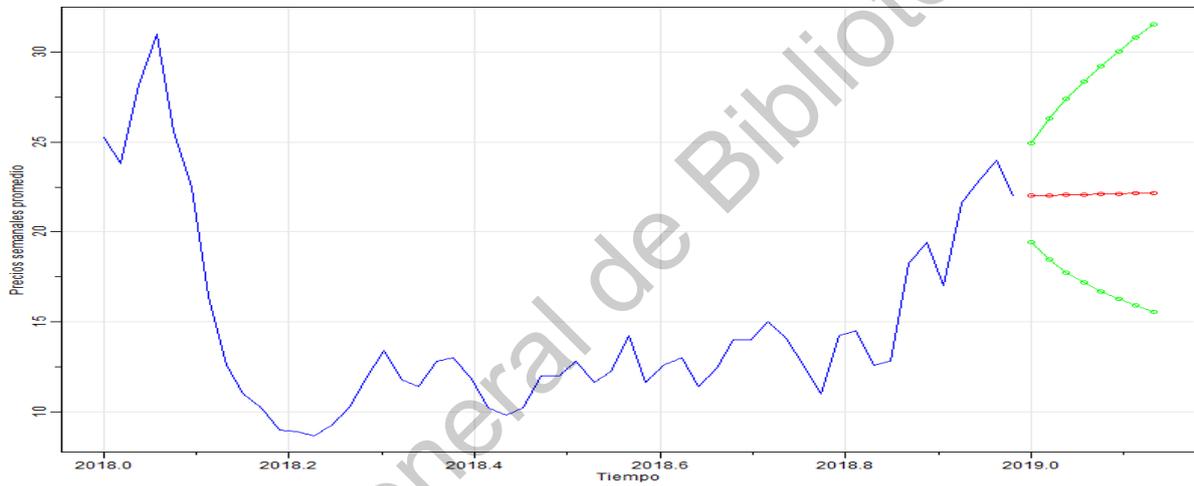


Figura I.13: Predicciones del nopal grande

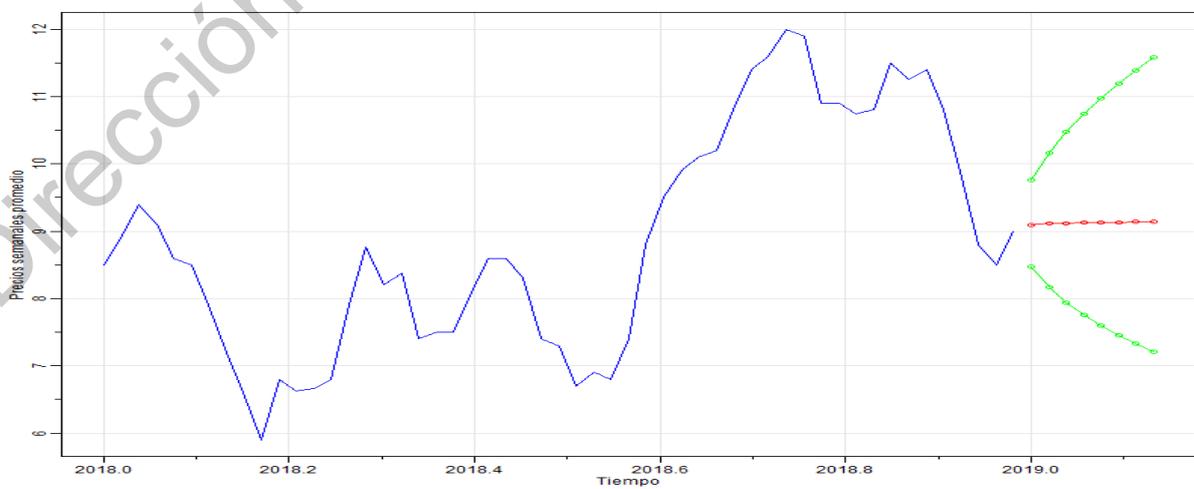


Figura I.14: Predicciones de la piña mediana

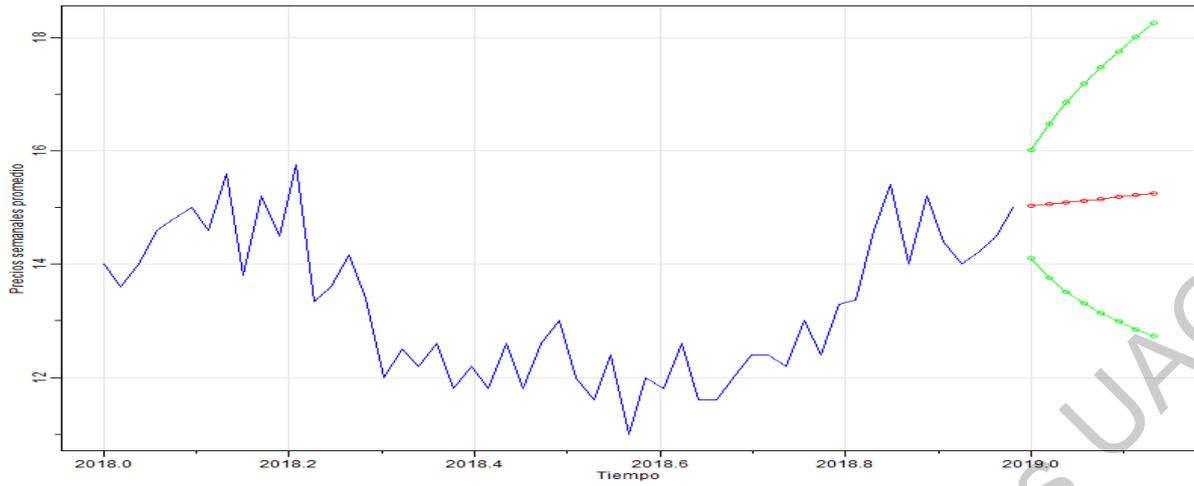


Figura I.15: Predicciones del plátano macho

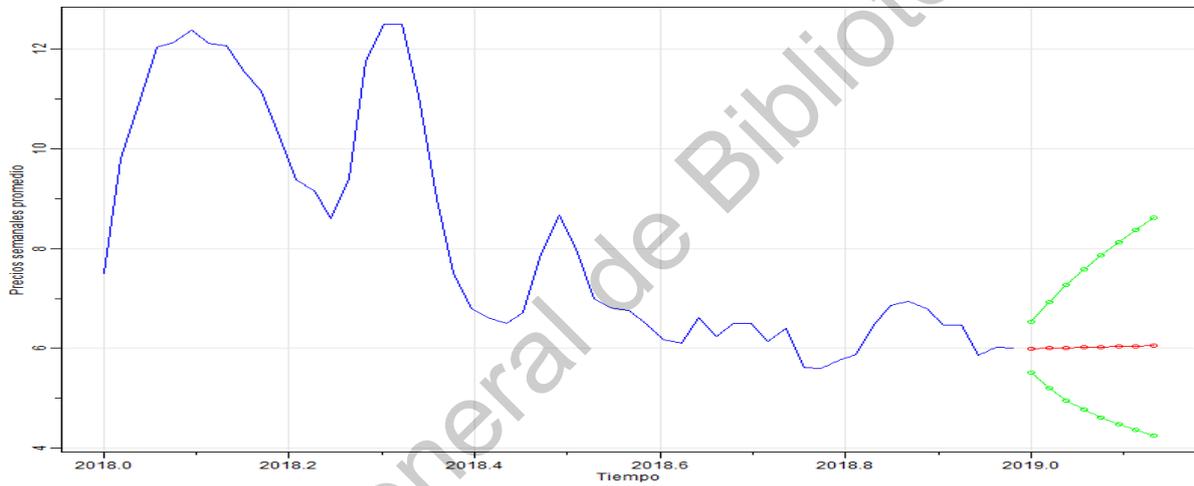


Figura I.16: Predicciones del plátano Tabasco

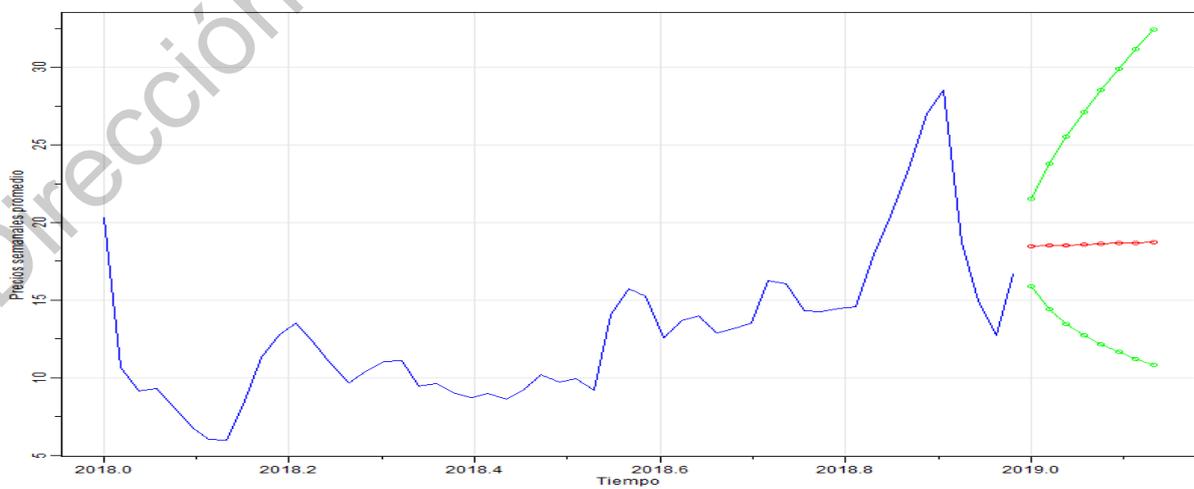


Figura I.17: Predicciones del tomate saladette

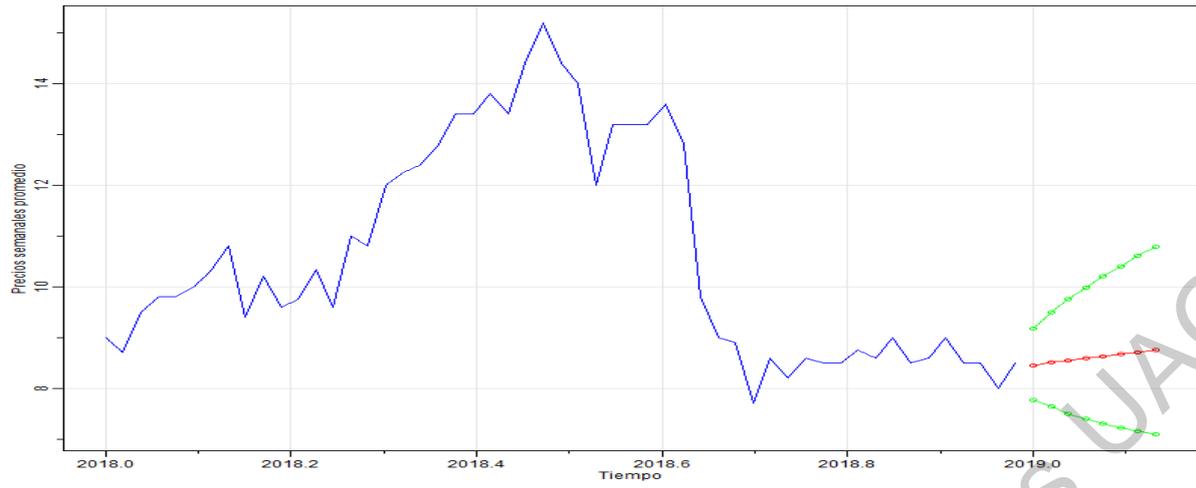
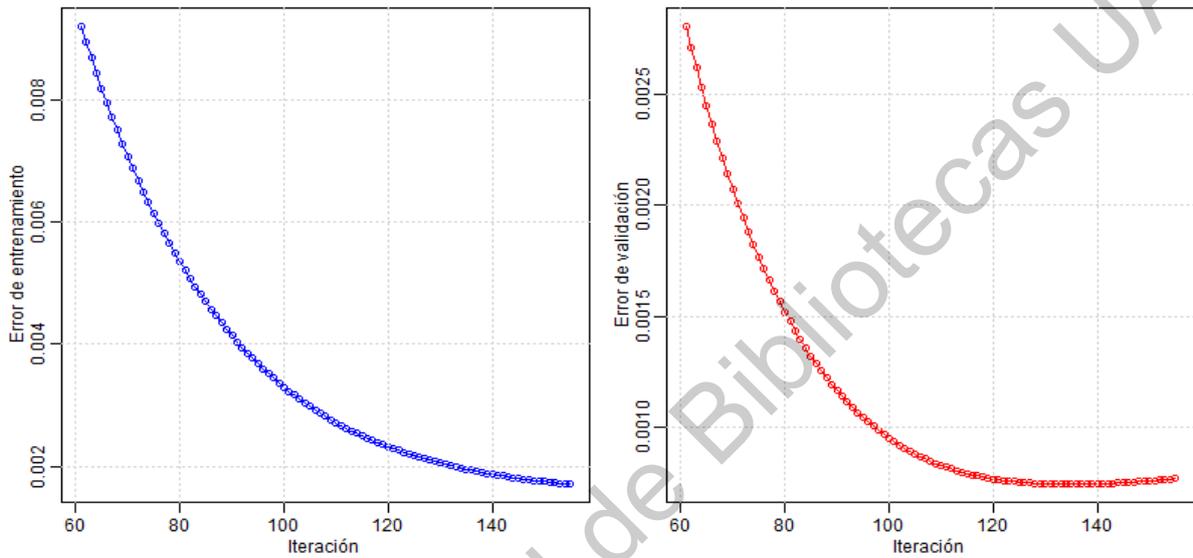
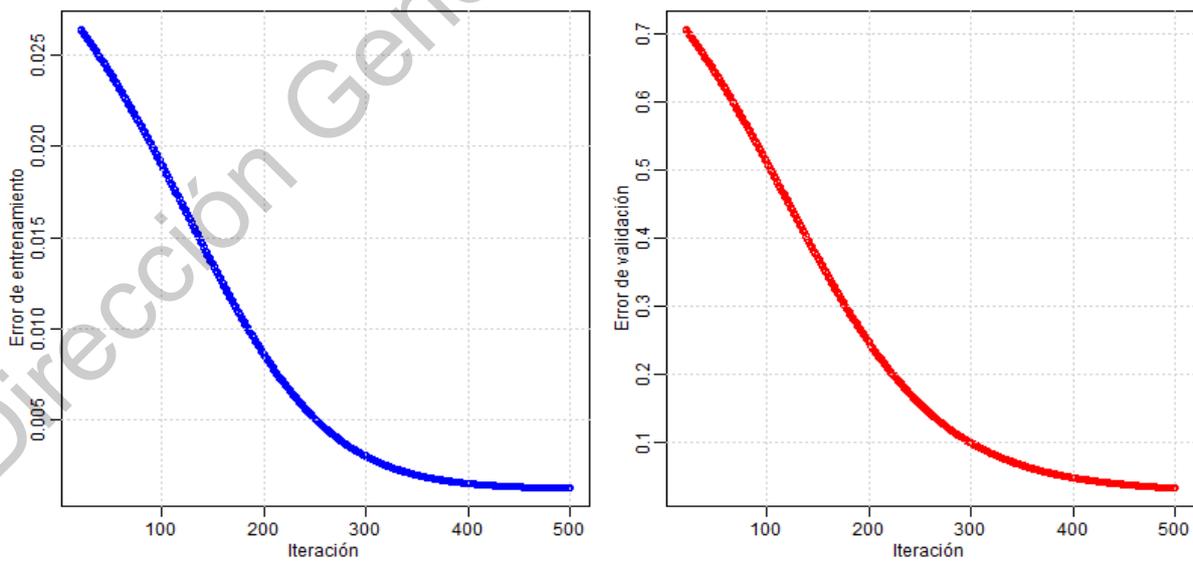


Figura I.18: Predicciones de la toronja roja

Dirección General de Bibliotecas UAQ

## APÉNDICE J

## Aprendizaje de la red neuronal

*Figura J.1: Aguacate Hass**Figura J.2: Cacahuete*

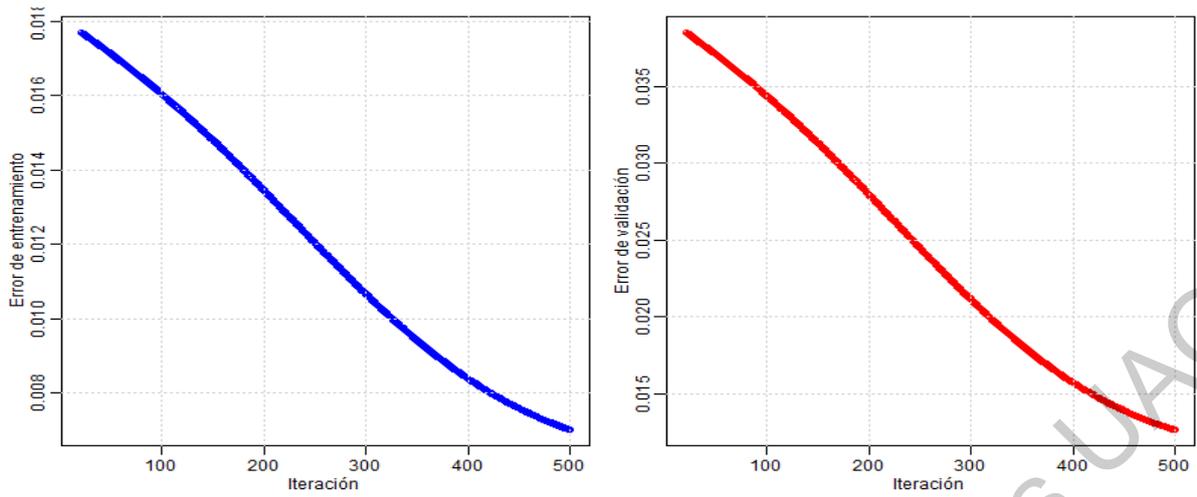


Figura J.3: Calabacita italiana

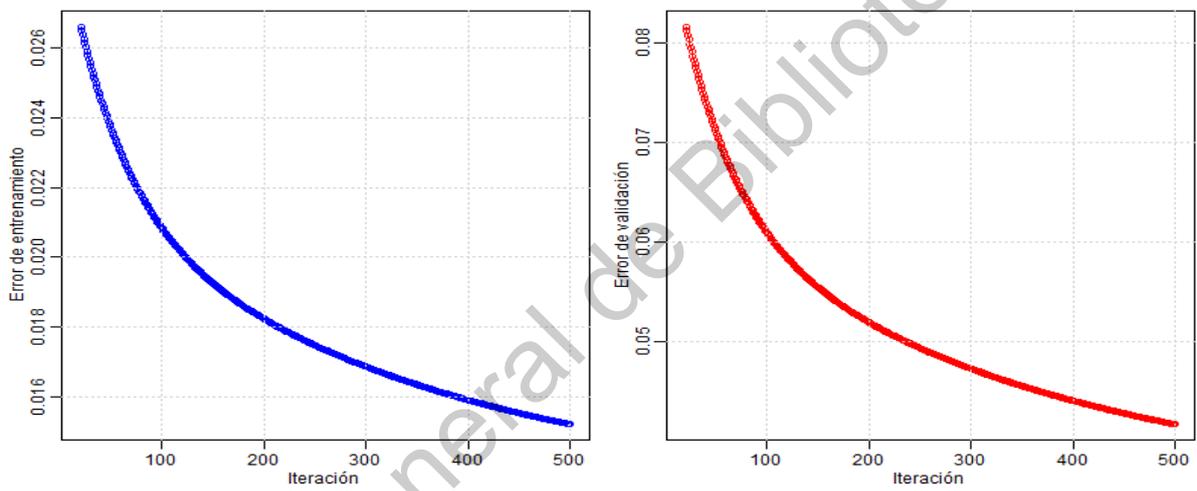


Figura J.4: Chayote sin espinas

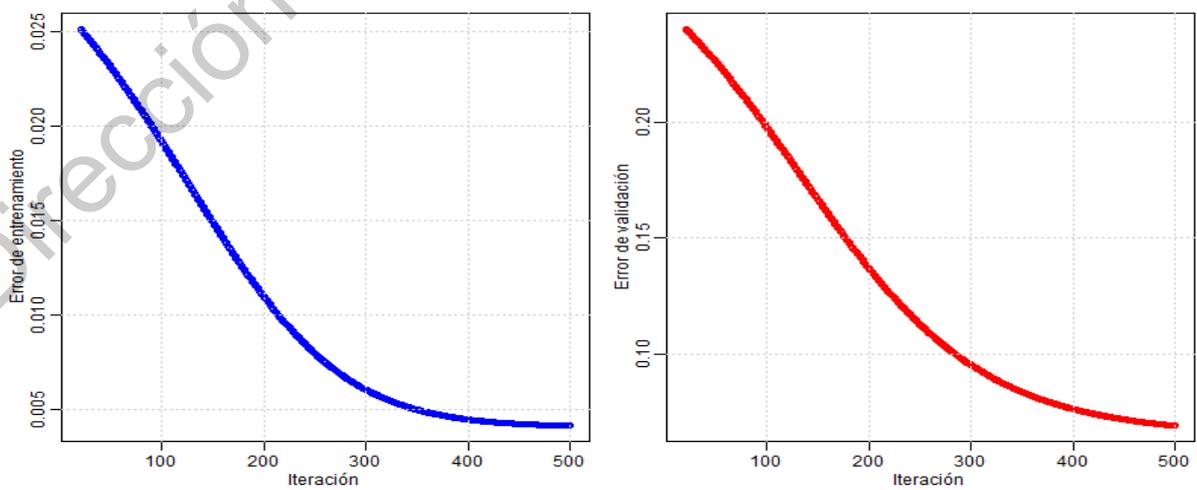


Figura J.5: Chile serrano

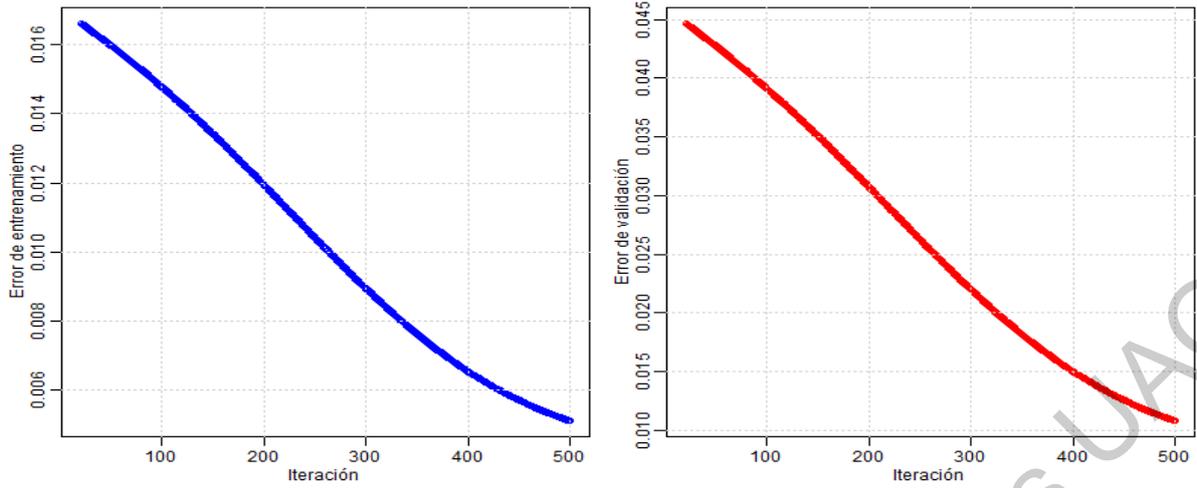


Figura J.6: Coliflor grande

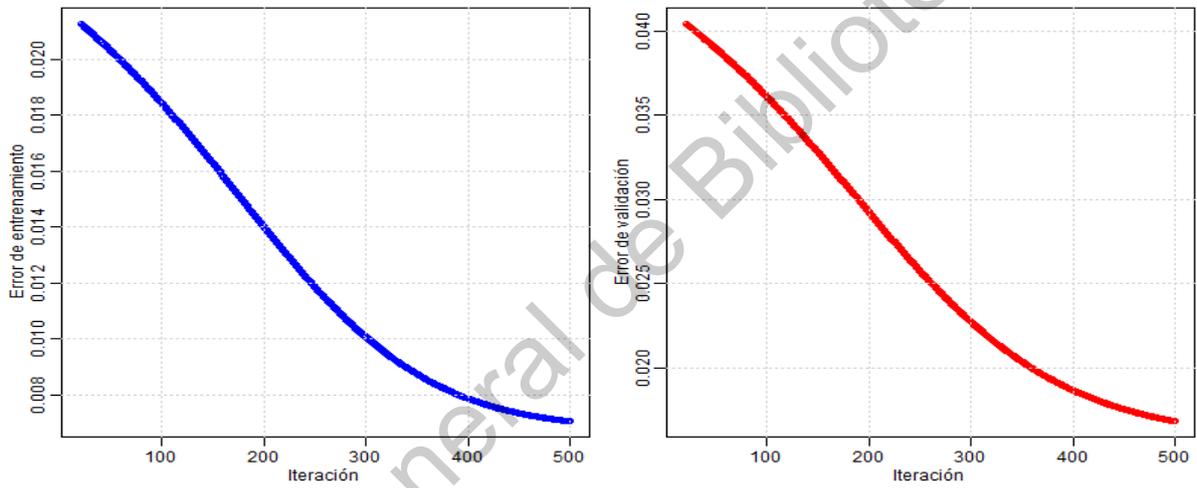


Figura J.7: Epazote

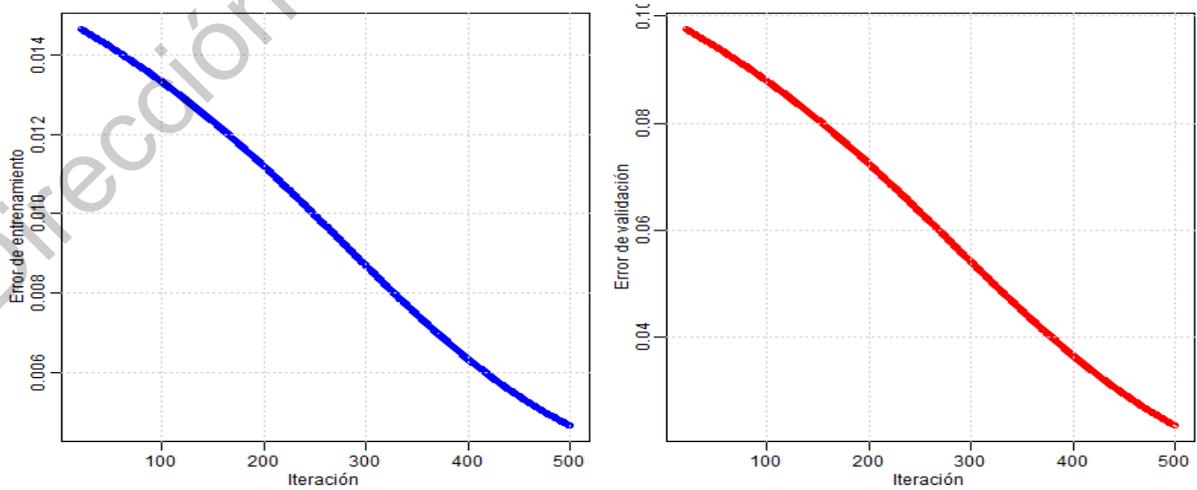


Figura J.8: Jícama

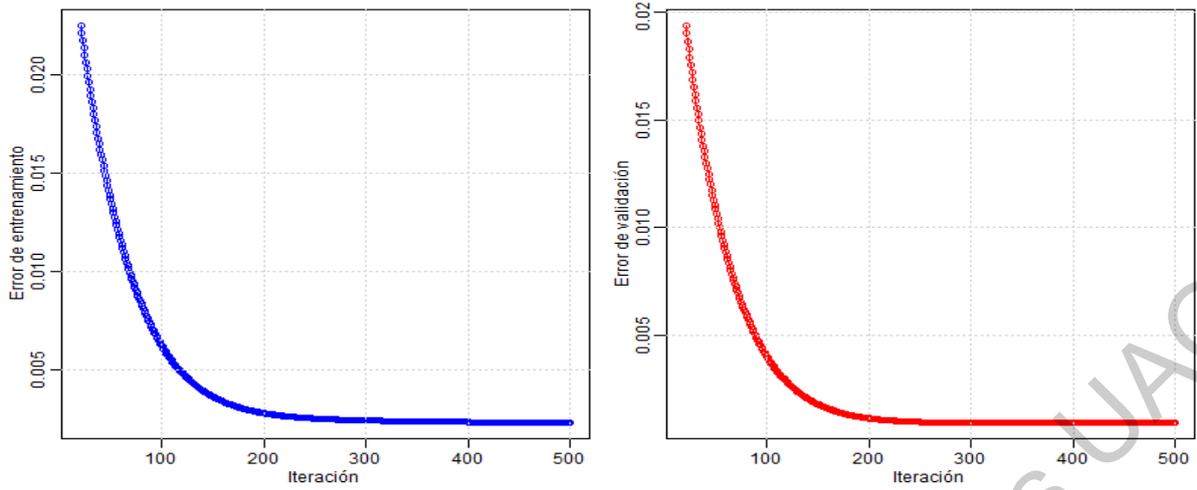


Figura J.9: Kiwi

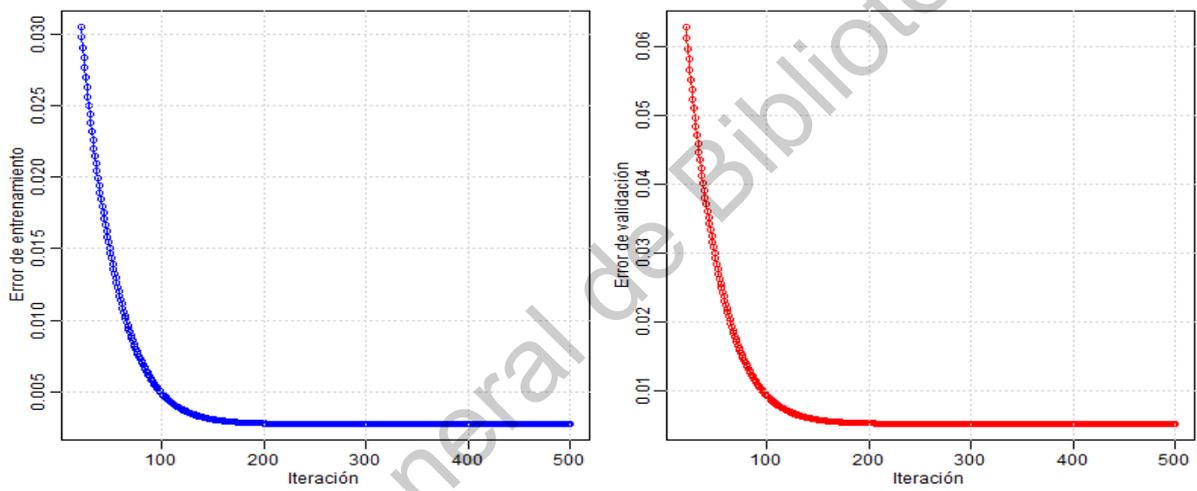


Figura J.10: Lima

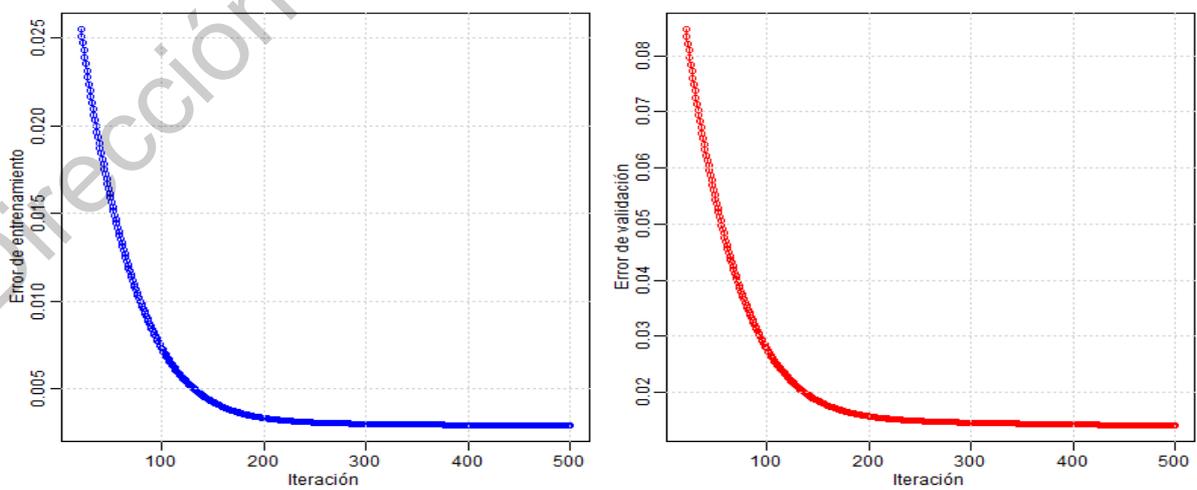


Figura J.11: Limón son semilla #5

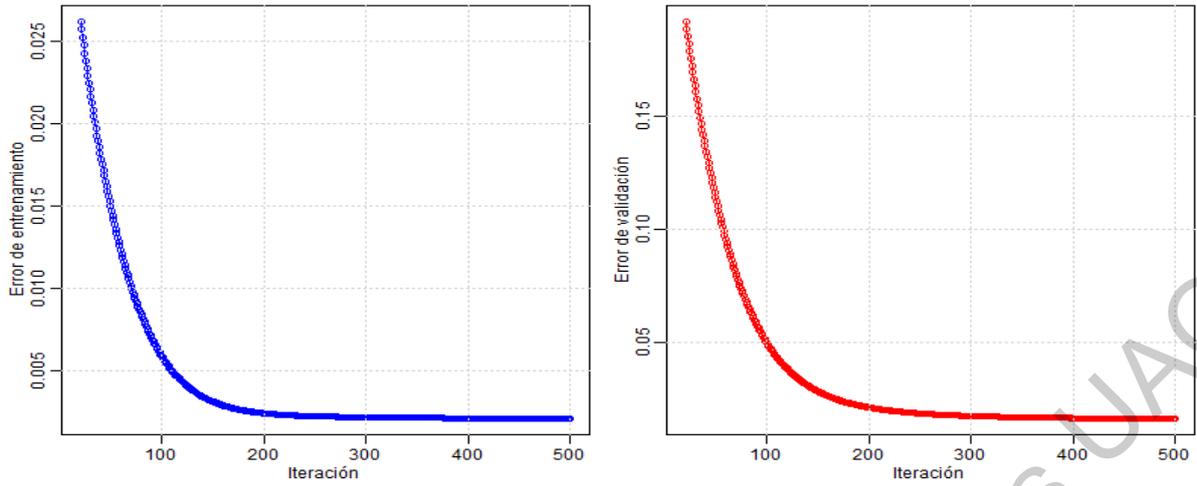


Figura J.12: Naranja Valencia grande

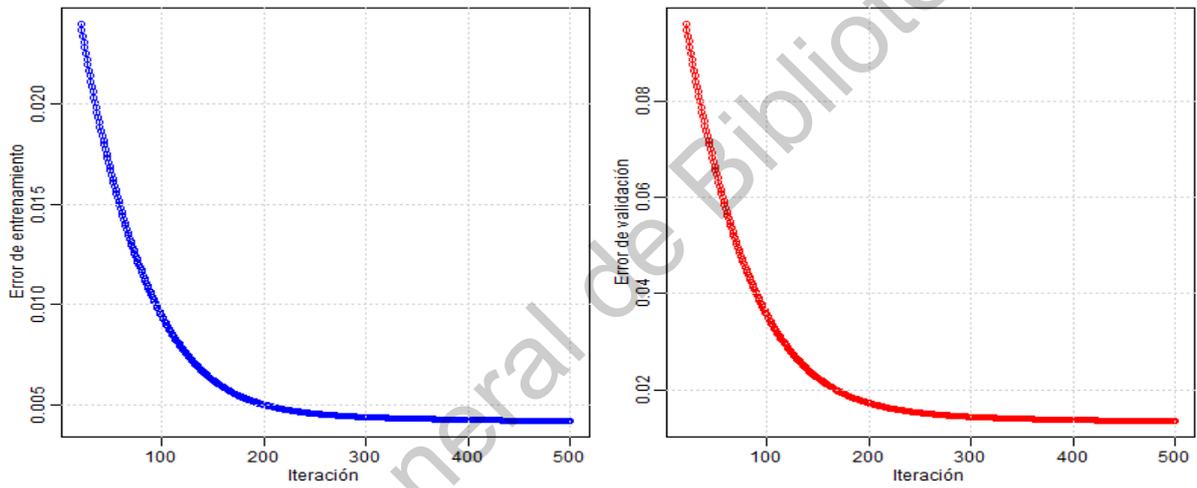


Figura J.13: Nopal grande

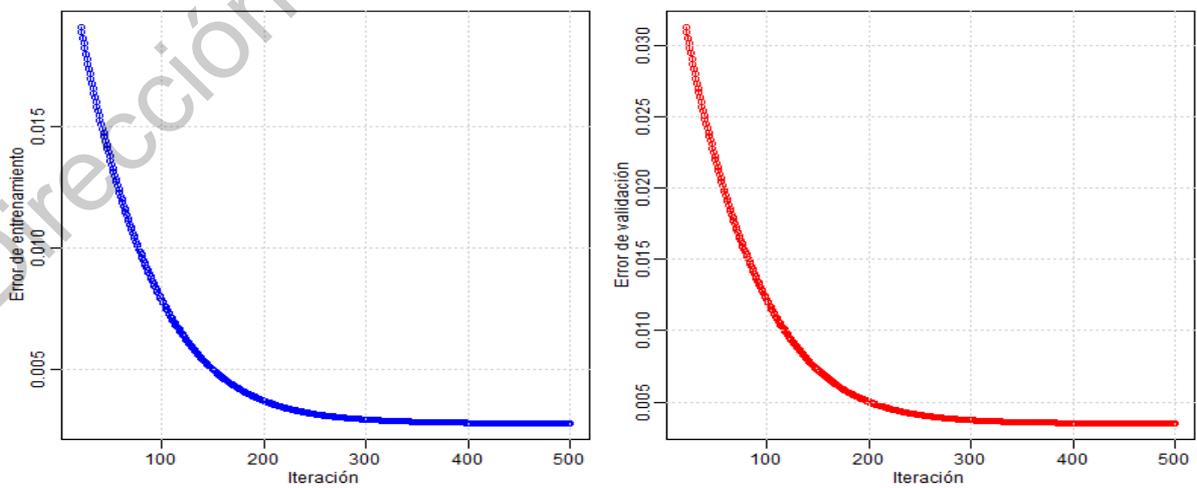


Figura J.14: Piña mediana

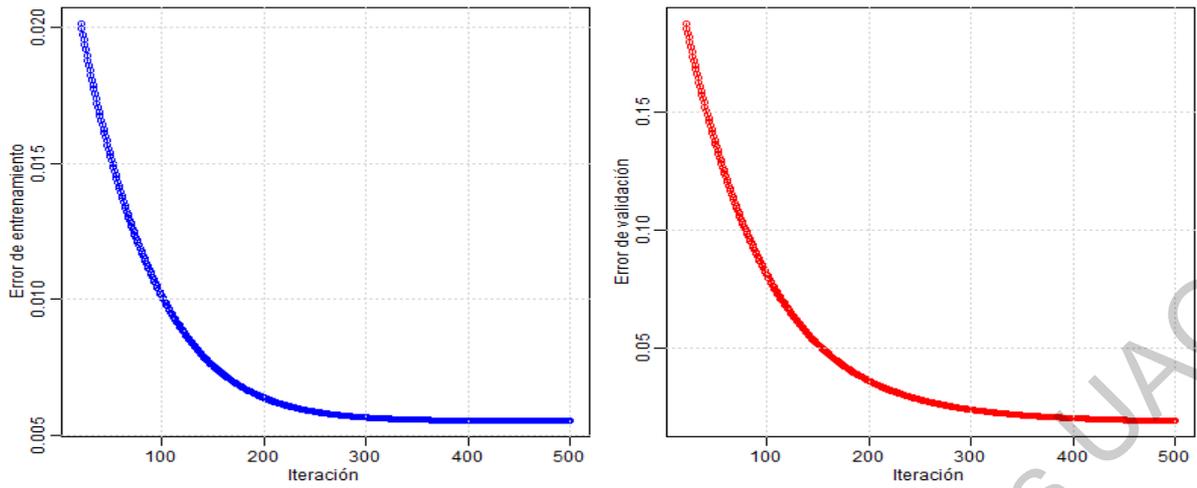


Figura J.15: Plátano macho

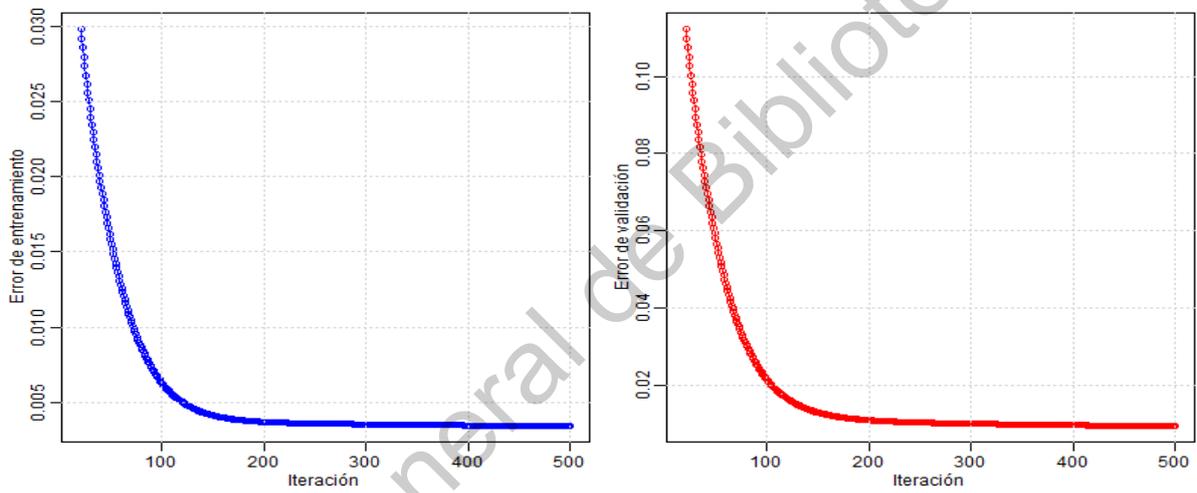


Figura J.16: Plátano Tabasco

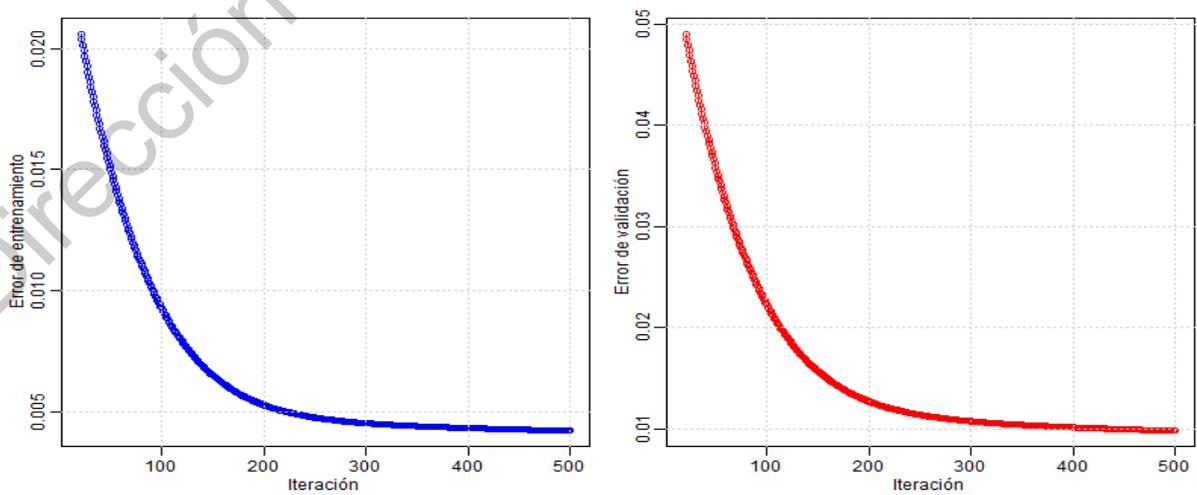


Figura J.17: Tomate saladette

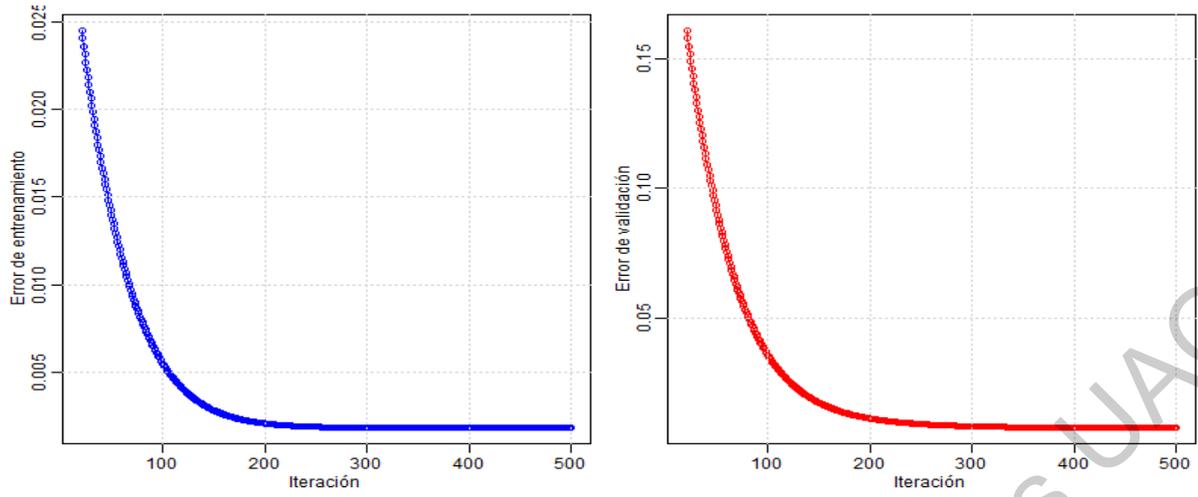


Figura J.18: Toronja roja

Dirección General de Bibliotecas UAQ

## APÉNDICE K

### Algoritmo: Construcción de base de datos

```

library(lubridate)
library(rvest)

ReadDB=function(){
  root="http://www.economia-sniim.gob.mx/nuevo/Consultas/
    MercadosNacionales/PreciosDeMercado/Agricolas/
    ResultadosConsultaFechaFrutasYHortalizas.aspx?fechaInicio="
  date_s=vector()
  date_e=vector()
  date_s=as.character.Date(seq(as.Date("01/01/2009",format = "%d
    /%m/%Y"), by="12┐months", length.out=11),format = "%d/%m/%Y"
  )
  date_e=c(as.character.Date(seq(as.Date("31/12/2009",format = "%
    d/%m/%Y"), by="12┐months", length.out=10),format = "%d/%m/%Y
    "),"28/02/2019")
  for(i in 1:11){
    path=paste0(root,date_s[i],"&fechaFinal=",date_e[i],"&
    ProductoId=-1&OrigenId=-1&Origen=Todos&DestinoId=220&
    Destino=Quer%c3%a9taro:%20Mercado%20de%20Abasto%20de%20
    Quer%c3%a9taro&PreciosPorId=1&RegistrosPorPagina=15000")
    page=read_html(path)
    tabla=html_nodes(page,"#tblResultados")
    database=html_table(tabla)
    database[[1]]=database[[1]][-1,-9]
  }
}

```

```

t1=which(database[[1]]$X1=="Frutas")
database[[1]]=database[[1]][-t1,]
t2=which(database[[1]]$X1=="Frutas_de_Temporada")
database[[1]]=database[[1]][-t2,]
t3=which(database[[1]]$X1=="Hortalizas")
database[[1]]=database[[1]][-t3,]
t4=which(database[[1]]$X1=="Chiles_Secos")
database[[1]]=database[[1]][-t4,]
if(i==1){
  bd=database[[1]]
}
else{
  bd=rbind(bd,database[[1]])
}
}
return(bd)
}

PriceKg=function(size,prices){
  numeros=sapply(size, function(x){regmatches(x,gregexpr("[[:digit:]]", x))})
  unidad=vectord(length = length(numeros))
  for (i in 1:length(numeros)){
    aux <- numeros[[i]]
    if (length(aux > 0)){
      aux2 <- character()
      for (j in 1:length(aux)){
        aux2 <- paste(aux2,aux[j],sep="")
      }
      unidad[i]=aux2
    }
  }
}

```

```

    else{unidad[i]= "1"}
}
unidad=as.numeric(unidad)
prices=prices/unidad
return(prices)
}

InfoProducts=function(db,n_prod){
  aux=vector("list")
  for (i in 1:n_prod){
    if((i==46||i==68)||i==56||i==57)||i==42){
      id=which((db$Producto==names_products[i])&(db$Modalidad
        =="I"))
    }
    else{
      if((i==7)||i==12||i==63)||i==64||i==69){
        if((i==7||i==63)||i==64||i==69){id=which((db$Producto==
          names_products[i])&(db$Modalidad=="N")&(db$
            Presentacion=="Kilogramo"))}
        else{id=which((db$Producto==names_products[i])&(db$
          Modalidad=="N")&(db$Presentacion!="Rollo"))}
      }
      else{
        id=which((db$Producto==names_products[i])&(db$Modalidad=="N
          "))
      }
    }
  }
  weekprice=tapply(db$PrecioKg[id],list(as.factor(year(db$Fecha
    [id])),as.factor(week(db$Fecha[id]))),function(x){mean(x,
    na.rm=T)})}
}

```

```

    aux[[i]]=list(names_products[i], weekprice)
  }
  return(aux)
}

```

```

ChooseProducts=function(names_products, dt, string){
  id=vector()
  for (i in 1:length(string)) {
    id=c(id, which(names_products==string[i]))
  }
  aux=vector("list")
  for(j in 1:length(string)){
    aux[[j]]=list(dt[[id[j]]][[1]], dt[[id[j]]][[2]])
  }
  return(aux)
}

```

```

WeekSerie=function(dt){
  a=which(names(dt[[2]][,1])==2015)
  price=as.vector(t(dt[[2]]))[((a-1)*53+1):((a+3)*53+8)]
  aux=which(is.na(price) == T)
  autocomplete=sapply(aux, function(x){mean(price[(x-3):(x+3)], na
    .rm = T)})
  price[aux]=autocomplete
  return(price)
}

```

```

db=ReadDB()
names(db)=c("Fecha", "Producto", "Calidad", "Presentacion",
  "Origen", "PrecioMin", "PrecioMax", "PrecioFrec")

```

```

db$Fecha=as.Date(db$Fecha,format="%d/%m/%Y")
db$Producto=as.factor(db$Producto)
db$Presentacion=as.factor(db$Presentacion)
db$PrecioFrec=as.numeric(db$PrecioFrec)

names_products=levels(db$Producto)
names_products=sort(names_products)
n_prod=length(names_products)

#Calculando precio por Kilogramo, ya que hay diferentes
  presentaciones
#Se usa el precio frecuente
size=levels(db$Presentacion)
size[30]="Ciento de 100 pzas"
size[32]="Docena de 12 pzas"
db$Presentacion=factor(as.numeric(db$Presentacion),levels=1:
  length(size),labels = size)
db=cbind(db,"PrecioKg"=PriceKg(db$Presentacion,db$PrecioFrec))

#Separando productos Nacionales y internacionales
db=cbind(db,"Modalidad"=unname(sapply(db$Origen, function(x){if(x
  == "óImportacin"){return("I")}else{return("N")}})))

#Separamos los datos por producto
datos=InfoProducts(db,n_prod)

string=sort(c("Coliflor grande","Nopal grande","Naranja Valencia
  grande",
  "Epazote","Chile Serrano","Aguacate Hass",
  "Calabacita Italiana",
  "íJcama","Kiwi","Lima","óLimn c/semilla #5",

```

```
        "áPltano_Macho",
        "ñPia_mediana", "Chayote_sin_espinas", "Cacahuate",
        "Toronja_Roja", "Tomate_Saladette", "áPltano_Tabasco"
    ))

#Elegimos un subconjunto de productos a estudiar
ANN_products=ChooseProducts(names_products,datos,string)
for(i in 1:length(ANN_products)){
    ANN_products[[i]][[2]]=WeekSerie(ANN_products[[i]])
}
```

Dirección General de Bibliotecas UAQ

## APÉNDICE L

### Algoritmo: Red neuronal recurrente ELMAN

```

neurona=setRefClass("neurona",
  fields = list(id="character",input="numeric",
    bias="numeric",
      num_nodos = "numeric",weight =
        "numeric",
      output="numeric" ),
  methods = list(initialize = function(posicion
    , num_conexiones,s){
    id<<-posicion
    num_nodos <<- num_conexiones
    set.seed(s)
    weight<<-runif(num_nodos,min = -1,max = 1)
    bias<<- 0
  },
  ComputeOutput=function(x){
    input<<-x
    output<<-FtAc(FtIn(input,weight,bias))
  },
  ComputeOutputNet=function(x){
    input<<-x
    output<<-FtIn(input,weight,bias)
  })
)

FtIn=function(x,w,b){

```

```

    return(sum(x*w)+b)
}

FtAc=function(x){
    return(1/(1+exp(-x)))
}

CreateELMAN=function(topology){
    n_capas=length(topology)
    red=vector("list")
    k=1
    for (i in 2:n_capas) {
        red[[i]]=vector("list")
        for (j in 1:topology[i]) {
            if(i==n_capas){
                id_n=paste0("y",i,j)
                red[[i]][[j]]=neurona$new(id_n,topology[i-1],k)
            }
            if(i==n_capas-1){
                id_n=paste0("u",i,j)
                red[[i]][[j]]=neurona$new(id_n,1,k)
            }
            if(i==2){
                id_n=paste0("z",i,j)
                red[[i]][[j]]=neurona$new(id_n,topology[i-1],k)
            }
            k=k+1
        }
    }
    for (l in 1:length(red[[3]])) {
        red[[3]][[l]]$output=0
    }
}

```

```

}
return(red)
}

FrontPropELMAN=function(red){
  input=vector()

  for(i in c(2,4)){
    aux=vector()
    if(i==2){
      net=vector()
      for (k in 1:length(red[[3]])) {
        net=c(net, red[[3]][[k]]$output*red[[3]][[k]]$weight)
      }
    }
    for (j in 1:length(red[[i]])) {

      if(i==2){
        red[[i]][[j]]$input=red[[1]]
        red[[i]][[j]]$output=FtAc(FtIn(red[[1]], red[[i]][[j]]$
          weight,0)+sum(net))
        aux=c(aux, red[[i]][[j]]$output)
      }else{
        red[[i]][[j]]$ComputeOutputNet(input)
        aux=c(aux, red[[i]][[j]]$output)
      }

    }

  }

  input=aux
  if(i==2){
    for (k in 1:length(red[[i]])) {

```

```

        red[[3]][[k]]$input=red[[3]][[k]]$output
        red[[3]][[k]]$output=aux[[k]]
    }
}
}
}
}

```

```

BackPropELMAN=function(red,result,stf){
    delta=vector("list")
    delta[[2]]=vector("list")
    delta[[3]]=vector("list")

    for(i in c(4,2)){
        if(i == length(red)){
            delta[[1]] = (result-red[[i]][[1]]$output)*stf* red[[i]][[1]]$input
            W_temp = red[[i]][[1]]$weight

        } else{
            for (j in 1:length(red[[i]])) {
                delta[[3]][[j]]=(result-red[[4]][[1]]$output)*stf*W_temp[j]*
                    (red[[i]][[j]]$output)*(1-red[[i]][[j]]$output)*red[[1]]
                delta[[2]][[j]]=(result-red[[4]][[1]]$output)*stf*W_temp[j]*
                    (red[[i]][[j]]$output)*(1-red[[i]][[j]]$output)*red[[3]][[j]]$input
            }
        }
    }

    return(delta)
}

```

```
}
```

```
Brownian_simulation=function(set){
  set.seed(0)
  dW=sqrt(1/set)*rnorm(set,0,1)
  W=c(0,cumsum(dW))
  return(W)
}
```

```
stf=function(u,s,w,N){
  bt=1
  fi=exp((sum(u)/N)+(s*(w[length(w)]-w[1]))) / bt
  return(fi)
}
```

```
ELMAN_Validation=function(prices,n,epoch,l_rate,patience){
  ANN=CreateELMAN(c(n,2*(n+1),2*(n+1),1))
  n_set=(prices-min(prices[1:159]))/(max(prices[1:159])-min(
    prices[1:159]))
  N=159
  t_set=n_set[1:N]
  v_set=n_set[(N+1):212]
  B=Brownian_simulation(N)
  u=1/(N-(0:(N-1)))^2
  s=sqrt(var(t_set))
  k=1
  train=0
  E_epoch=vector("list")
  E_epoch[[1]]=vector()
  E_epoch[[2]]=vector()
```

```

sample=vector("list")
for (j in 1:(N-n)) {
  sample[[j]]=t_set[0:n+j]
}

sample_eval=vector("list")
for (j in 1:(53-n)) {
  sample_eval[[j]]=v_set[0:n+j]
}

n_p=0
opt=0
error=1000
aux=vector("list")
while((k<=epoch)&(n_p<patience)){
  E_tn=vector()
  D_W=vector("list")
  for (i in 1:(2*n+2)) {
    D_W[[i]]=rep(0,n)
  }
  D_C=rep(0,2*n+2)
  D_V=rep(0,2*n+2)
  D_E_tn=vector("list")
  for (i in 1:(N-n)) {
    ANN[[1]]=sample[[i]][1:n]
    FrontPropELMAN(ANN)
    phi_tn=stf(u[0:(n-1)+i],s,B[0:(n)+i],N)
    E_tn=c(E_tn,0.5*phi_tn*(sample[[i]][n+1]-ANN[[4]][[1]]$
      output)^2)
    D_E_tn=BackPropELMAN(ANN,sample[[i]][n+1],phi_tn)
    D_V=D_V+D_E_tn[[1]]
    D_C=D_C+unlist(D_E_tn[[2]])
  }
  n_p=n_p+1
  error=error-E_tn
  if(error<opt) opt=error
}

```

```

    for (j in 1:(2*n+2)) {
        D_W[[j]]=D_W[[j]]+D_E_tn[[3]][[j]]
    }
}
train=sum(E_tn)/(N-n)
E_epoch[[1]]=c(E_epoch[[1]],train)

k=k+1

pred=0
for(i in 1:(53-n)){
    ANN[[1]]=sample_eval[[i]][1:n]
    FrontPropELMAN(ANN)
    pred=pred+((sample_eval[[i]][n+1]-ANN[[4]][[1]]$output)^2)
        /(53-n)
}
E_epoch[[2]]=c(E_epoch[[2]],pred)
if(pred<error){
    n_p=0
    opt=k-1
    error=pred
}
else{
    n_p=n_p+1
}

if((k<=epoch)&(n_p<patience)){
    ANN[[4]][[1]]$weight=ANN[[4]][[1]]$weight+(1_rate/(N-n))*D_
        V
    for (j in 1:(2*n+2)) {
        ANN[[3]][[j]]$weight=ANN[[3]][[j]]$weight+(1_rate/(N-n))*

```

```

        D_C[j]
        ANN[[2]][[j]]$weight=ANN[[2]][[j]]$weight+(1_rate/(N-n))*
        D_W[[j]]
    }
}

}

print(opt)
return(E_epoch)
}

```

```

ELMAN_Training=function(prices,n,epoch,l_rate){
  ANN=CreateELMAN(c(n,2*(n+1),2*(n+1),1))
  n_set=(prices-min(prices[1:212]))/(max(prices[1:212])-min(
    prices[1:212]))
  N=212
  t_set=n_set[1:N]
  B=Brownian_simulation(N)
  u=1/(N-(0:(N-1)))^2
  s=sqrt(var(t_set))
  k=0
  train=0
  sample=vector("list")
  for (j in 1:(N-n)) {
    sample[[j]]=t_set[0:n+j]
  }

  while(k<epoch) {

```

```

E_tn=vector()
D_W=vector("list")
for (i in 1:(2*n+2)) {
  D_W[[i]]=rep(0,n)
}
D_C=rep(0,2*n+2)
D_V=rep(0,2*n+2)
D_E_tn=vector("list")
for (i in 1:(N-n)) {
  ANN[[1]]=sample[[i]][1:n]
  FrontPropELMAN(ANN)
  phi_tn=stf(u[0:(n-1)+i],s,B[0:(n)+i],N)
  #E_tn=c(E_tn,0.5*phi_tn*(sample[[i]][n+1]-ANN[[4]][[1]]$
    output)^2)
  D_E_tn=BackPropELMAN(ANN,sample[[i]][n+1],phi_tn)
  D_V=D_V+D_E_tn[[1]]
  D_C=D_C+unlist(D_E_tn[[2]])
  for (j in 1:(2*n+2)) {
    D_W[[j]]=D_W[[j]]+D_E_tn[[3]][[j]]
  }
}

if(k<epoch){
  ANN[[4]][[1]]$weight=ANN[[4]][[1]]$weight+(1_rate/(N-n))*D_
    V
  for (j in 1:(2*n+2)) {
    ANN[[3]][[j]]$weight=ANN[[3]][[j]]$weight+(1_rate/(N-n))*
      D_C[j]
    ANN[[2]][[j]]$weight=ANN[[2]][[j]]$weight+(1_rate/(N-n))*
      D_W[[j]]
  }
}

```

```

    }
    k=k+1
}
return(ANN)
}
ELMAN_Forecasting=function(ANN,n,prices){
  N=212
  n_prices=(prices-min(prices[1:N]))/(max(prices[1:N])-min(prices
    [1:N]))

  f_price=vector()
  for(i in 1:8){
    ANN[[1]]=c(n_prices[(N-n+1):N],f_price)[0:(n-1)+i]
    FrontPropELMAN(ANN)
    f_price=c(f_price,ANN[[4]][[1]]$output)
  }
  f_price=f_price*(max(prices[1:N])-min(prices[1:N]))+min(prices[
    1:N])
  return(f_price)
}

```

## Bibliografía

- Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Canales, E., Andrango, G., and Williams, A. (2019). Mexico's agricultural sector: Production potential and implications for trade. *Choices*, 34(316-2019-3254):1–12.
- Catalão, J. P. d. S., Mariano, S. J. P. S., Mendes, V., and Ferreira, L. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *electric power systems research*, 77(10):1297–1304.
- Cruz, I. B., Martínez, S. S., Abed, A. R., Ábalo, R. G., and Lorenzo, M. M. G. (2007). Redes neuronales recurrentes para el análisis de secuencias. *Revista Cubana de Ciencias Informáticas*, 1(4):48–57.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3):525–546.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Kiernan, D. (2014). *Natural Resources Biometrics*. Open SUNY Textbooks, Milne Library, State University of New York at Geneseo.
- Larranaga, P., Inza, I., and Moujahid, A. (1997). Tema 8. redes neuronales. *Redes Neuronales, U. del P. Vasco*, 12:17.
- Li, G.-q., Xu, S.-w., and Li, Z.-m. (2010). Short-term price forecasting for agro-products using artificial neural networks. *Agriculture and Agricultural Science Procedia*, 1:278–287.

- López, R. F., Fernández, J. M., and Fernández, J. M. F. (2008). *Las redes neuronales artificiales*. Netbiblo.
- Matich, D. J. (2001). *Redes neuronales: Conceptos básicos y aplicaciones*. Universidad Tecnológica Nacional, México.
- Michalowicz, A., Provost, K., Laruelle, S., Mimouni, A., and Vlaic, G. (1999). F-test in exafs fitting of structural models. *Journal of synchrotron radiation*, 6(3):233–235.
- Mitchell, R. (2018). *Web scraping with Python: Collecting more data from the modern web*. O'Reilly Media, Inc.”.
- Paredes-Garcia, W. J., Ocampo-Velázquez, R. V., Torres-Pacheco, I., and Cedillo-Jiménez, C. A. (2019). Price forecasting and span commercialization opportunities for mexican agricultural products. *Agronomy*, 9(12):826.
- Rojas, R. (1996). The backpropagation algorithm. In *Neural networks*, pages 149–182. Springer.
- Shumway, R. H. and Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.
- Stoffer, D. and Stoffer, M. D. (2020). Package ‘astsa’. *blood*, 8:1.
- Wang, J. and Wang, J. (2016). Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations. *Energy*, 102:365–374.
- Wasserman, L. (2004). *All of statistics: a concise course in statistical inference*, volume 26. Springer.
- Wickham, H. and Wickham, M. H. (2016). Package ‘rvest’. URL: <https://cran.r-project.org/web/packages/rvest/rvest.pdf>.
- Xie, H. and Wang, B. (2017). An empirical analysis of the impact of agricultural product price fluctuations on china’s grain yield. *Sustainability*, 9(6):906.
- Zhao, B. (2017). Web scraping. *Encyclopedia of big data*, pages 1–3.