

## APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS PARA PRONÓSTICO DE PRECIOS EN JITOMATE\*

### APPLICATION OF BOX-JENKINS METHODOLOGY FOR FORECASTING PRICES IN TOMATOES

Gaspar Marroquín Martínez<sup>1§</sup> y Luis Eduardo Chalita Tovar<sup>1</sup>

<sup>1</sup>Posgrado de Economía. Colegio de Postgraduados. Carretera México-Texcoco, km 36.5. Montecillo, Texcoco, Estado de México. C. P. 56230. (chalita@colpos.mx).

<sup>§</sup>Autor para correspondencia: marroquin@colpos.mx.

#### RESUMEN

Los productos del sector agroalimentario tienen como características económicas distintivas, la alta variabilidad en sus precios. Teniendo en cuenta la incertidumbre de los precios, una posible forma de planificar racionalmente la toma de decisiones, que consiste en elaborar pronósticos confiables del comportamiento futuro de esa variable. En este trabajo se usó la metodología Box-Jenkins, para identificar un modelo econométrico autoregresivo integrado de media móvil (ARIMA), que se ajusta al comportamiento de la serie de tiempo de precios nominales en venta al mayoreo de jitomate bola en México. Se concluye de acuerdo a los resultados, que la serie de tiempo objeto de estudio, se ajusta a un modelo ARIMA (23, 0, 1), dicho modelo posee dos factores autoregresivos y uno de media móvil. Con el modelo se hicieron pronósticos para 12 meses, los cuales comprenden de diciembre de 2008 a noviembre de 2009.

**Palabras clave:** ARIMA, incertidumbre, serie de tiempo.

Los productos del sector agroalimentario tienen como una de sus características económicas distintivas la alta variabilidad en sus precios. Ésta se origina, por una parte en los factores biológicos y climáticos que definen su productividad, y por otra el productor no puede fijar el precio de sus productos.

#### ABSTRACT

Agri-food products have as a distinctive economic characteristic, its high variability in prices. Given the uncertainty of prices, a possible way of rational planning decisions, is to develop reliable forecasts of that variable's future behavior. In this paper we used the Box-Jenkins methodology to identify an econometric autoregressive integrated moving average model (ARIMA), which fits the behavior of time series of nominal prices for beef tomato wholesaling in Mexico. According to the results we concluded that the time series under consideration, fits to an ARIMA model (23, 0, 1), this model has two autoregressive factors and a moving average. With this model there were made forecasts for 12 months, which are from December 2008 to November 2009.

**Key words:** ARIMA, uncertainty, time series.

Agri-food products have as a distinctive economic characteristic, its high variability in prices. It is originated, first due to biological and climatic factors that define their productivity and, secondly, because the producers cannot set the price of their products. Given the uncertainty of prices, a possible way of

\* Recibido: octubre de 2010  
Aceptado: julio de 2011

Dada la incertidumbre de los precios, una posible forma de planificar racionalmente la toma de decisiones consiste en elaborar pronósticos confiables del comportamiento futuro de esa variable (Cartes *et al.*, 2008).

En el presente trabajo se aplica la metodología de los modelos Box-Jenkins, usando datos mensuales de una serie de tiempo de precios nominales, por la venta al mayoreo de jitomate bola de primera calidad de enero 1998 a noviembre 2008, para conocer el comportamiento de dichos precios y realizar pronósticos. Se utilizó una serie de tiempo de los precios nominales de jitomate bola de primera calidad reportada en la Central de Abasto de Iztapalapa de la Ciudad de México de enero de 1998 a noviembre de 2008. La información fue obtenida del sistema nacional de información e integración de mercados (SNIIM) de la Secretaría de Economía (SE) y fue procesada en el paquete computacional SAS. La metodología seguida para la construcción del modelo ARIMA, se ajustó a los pasos sugeridos por Box-Jenkins: identificación del modelo, estimación de parámetros, verificación de supuestos (Greene, 2003).

### Identificación del modelo

Antes de identificar el modelo al cual se ajustan los datos, se procedió a verificar la estacionalidad o no estacionalidad de la serie de tiempo. Como se muestra en el Cuadro 1, se aplicó la prueba de raíz unitaria de Dickey Fuller aumentada (Damodovar, 2004).

**Cuadro 1. Prueba de Dickey Fuller aumentada para la serie de tiempo del precio de jitomate bola de primera calidad. Table 1. Augmented Dickey-Fuller test for time series of premium beef tomato prices.**

Tipo	Retardos	Rho	Pr< Rho	Tau	Pr< Tau	F	Pr> F
Media simple	0	-49.1274	0.0001	-5.39	<0.0001	14.58	0.001
	1	-72.586	0.0011	-5.93	<0.0001	17.64	0.001
	2	-68.3513	0.0011	-5.17	<0.0001	13.44	0.001
Tendencia	0	-62.1614	0.0004	-6.28	<0.0001	19.7	0.001
	1	-106.473	0.0001	-7.16	<0.0001	25.62	0.001
	2	-130.584	0.0001	-6.46	<0.0001	20.88	0.001

La prueba de hipótesis se puede escribir de la siguiente manera:  $H_0: \rho = 1$  y  $H_1: \rho < 1$ ; y el estadístico de prueba es  $\tau$  (tau). La regla de decisión es: rechazar  $H_0$  si  $\tau$  (tau) calculada  $< \tau$  (tau) de tablas.

La hipótesis nula es que la serie es estacionaria. Si no rechazamos la hipótesis  $\rho = 1$ , entonces la serie no es estacionaria. Dado que la tau calculada es menor (más negativa) que la tau de tablas se rechaza la hipótesis nula; por lo tanto, la serie de tiempo es estacionaria (Pankratz, 1983).

rational planning decisions, is to develop reliable forecasts of that variable's future behavior. (Cartes *et al.*, 2008).

In this paper, the Box Jenkins model methodology was applied, using monthly data from a time series of nominal prices for premium beef tomato whole sale from January 1998 to November 2008, to understand those prices behavior and forecasting. We used a time series of nominal prices of premium beef tomatoes reported for the Iztapalapa's Central Supply in Mexico City, from January 1998 to November 2008. Information was obtained from the Information and Market Integration National System (IMINS) of the Ministry of Economy (ME) and was processed in the SAS software package. Methodology used to construct the ARIMA model was adjusted to the steps suggested by Box-Jenkins: model identification, parameter estimation, verification of assumptions (Greene, 2003).

### Model identification

Before identifying the model which fits the data, the stationary and no stationary of time series was verified. As shown in Table 1, the augmented Dickey-Fuller test of unit root was applied (Damodovar, 2004).

Hypothesis test can be written as follows:  $H_0: \rho = 1$  and  $H_1: \rho < 1$ , test statistic is  $\tau$  (tau). The decision rule is: reject  $H_0$  if  $\tau$  (tau) calculated  $< \tau$  (tau) of tables.

Null hypothesis is that the series is stationary. If we do not reject the hypothesis  $\rho = 1$ , then the series is not stationary. Since the calculated tau is lower (more negative) than the tau tables, the null hypothesis is rejected, therefore, time series is stationary (Pankratz, 1983).

Once verified that the analyzed series is stationary, the model is identified by analyzing the correlograms. Autocorrelations decrease up to lag 3, and then only

Una vez que se verificó que la serie objeto de análisis es estacionario, se procede a identificar el modelo, analizando los correlogramas. Las autocorrelaciones decrecen hasta el rezago 3, luego sólo los rezagos 12 y 23 son significativos. Las autocorrelaciones parciales muestran picos en los rezagos 1, 11 y 23 mismos que parecen estadísticamente significativos.

### Estimación de parámetros

En la etapa de identificación se sugiere que el proceso que generó la serie de tiempo es como máximo un proceso AR (23), por lo que se procedió a estimar dicho modelo sumándole un factor de media móvil MA(1) (Cuadro 2).

12 and 23 lags are significant. Partial autocorrelations show peaks at lags 1, 11 and 23, which seem statistically significant.

### Parameter estimation

At the identification stage, it is suggested that the process that generated the time series is at most an AR process (23), which is why the model was estimated by adding a moving average factor MA (1) (Table 2).

Considering the values of  $t$  and the value of  $P$ , both the two autoregressive coefficients and the moving average are significant (SAS, 1998).

### Cuadro 2. Estimación del modelo para la serie de tiempo del precio de jitomate bola de primera calidad.

Table 2. Estimation of the model for time series of premium beef tomato prices.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimación	Error estándar	Valor $t$	Aprox. $Pr > t$	Retardo
MU	9.63269	0.96146	10.02	<0.0001	0
MA 1,1	-0.28653	0.12607	-2.27	0.0247	1
AR 1,1	0.39435	0.11578	3.41	0.0009	1
AR 1,2	0.30032	0.08436	3.56	0.0005	23

Considerando los valores de  $t$  y el valor de  $P$ , tanto los dos coeficientes auto regresivos como la media móvil son significativos (SAS, 1998).

### Diagnóstico comparativo

En esta etapa se analiza y se decide si nuestro modelo es estadísticamente adecuado. La prueba más importante del modelo es el ajuste razonable a los datos, y los residuos estimados sean puramente aleatorios. Por lo que se aplican dos pruebas, una de autocorrelación de los residuales y otra de ruido blanco. La hipótesis nula de la correlación entre los residuales es cero, y la regla de decisión es rechazar  $H_0$  si Chi cuadrada calculada  $>$  Chi cuadrada de tablas.

El estadístico de prueba es Chi-cuadrada, y a un nivel de significancia de 5% no se rechaza la hipótesis nula. En los diferentes grupos, las correlaciones son cercanas a cero (Cuadro 3).

### Comparative diagnosis

At this stage we analyze and decide if our model is statistically adequate. The most important test of the model is the reasonable fit to data and that estimated residuals are purely random. Two tests were applied, a residuals autocorrelation and a white noise test. The null hypothesis of correlation between the residuals is zero and the decision rule is to reject  $H_0$  if calculated Chi square  $>$  Chi square tables.

The Chi-square statistic test at 5% significance level, null hypothesis cannot be rejected. In different groups, correlations are close to zero (Table 3).

For null hypothesis, residuals are white noise. Comparing the statistic test values calculated using tables, at a 5% significance level; null hypothesis cannot be rejected. Residuals behave as white noise (Table 4).

**Cuadro 3. Prueba de autocorrelación de residuales.**  
**Table 3. Test of residuals autocorrelation.**

Retardo	Chi-cuadrada	DF	Pr> Chi cuadrada	Autocorrelación					
6	7.54	3	0.0565	-0.015	-0.035	0.005	-0.122	0.19	-0.046
12	14.59	9	0.1028	0.066	0.102	-0.076	-0.012	0.039	0.163
18	28.87	15	0.0167	-0.02	0.085	0.147	0.019	0.243	-0.071
24	33.55	21	0.0405	0.023	-0.004	0.059	0.155	0.008	0.036
30	37.92	27	0.0791	0.017	0.018	0.056	-0.097	0.106	-0.039

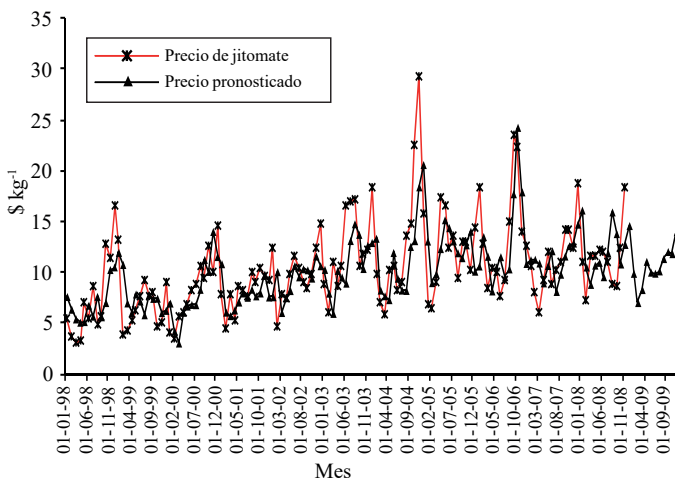
Para la hipótesis nula los residuales son ruido blanco. Comparando los valores del estadístico de prueba calculados con los de tablas, a un nivel de significancia de 5% no se rechaza la hipótesis nula. Los residuales tienen un comportamiento de ruido blanco (Cuadro 4).

**Cuadro 4. Prueba de ruido blanco.**  
**Table 4. Test of white noise.**

Retardo	Chi cuadrada	DF	Pr> Chi cuadrada	Autocorrelaciones					
6	7.76	6	0.2561	-0.022	-0.043	-0.004	-0.134	0.181	-0.058
12	14.17	12	0.2898	0.056	0.091	-0.088	-0.023	0.027	0.155
18	27.45	18	0.0709	-0.03	0.074	0.138	0.008	0.234	-0.084
24	31.14	24	0.1498	0.012	-0.016	0.046	0.141	-0.007	0.022
30	35.5	30	0.2250	0.002	0.003	0.042	-0.111	0.093	-0.054

**Pronóstico**

En la Figura 1 se llevó a cabo la última etapa de la metodología; en el cual, se identificó, estimó y verificó el modelo con respecto al modelo estimado, se realizaron pronósticos para 12 periodos mensuales, obteniéndose los siguientes valores (Cuadro 5).



**Figura 1. Pronóstico del precio de jitomate.**  
**Figure 1. Forecast of beef tomatoes prices.**

**Forecast**

Figure 1 holds the last stage of methodology, in which were identified, estimated and verified the model respect to estimated model. Forecasts were made for 12 monthly periods, obtaining values shown in Table 5.

**Cuadro 5. Pronóstico.**  
**Table 5. Forecast.**

Mes	Año	Precio pronosticado (\$ kg <sup>-1</sup> )
Diciembre	2008	14.56
Enero	2009	9.88
Febrero	2009	7.02
Marzo	2009	8.29
Abril	2009	11.05
Mayo	2009	9.99
Junio	2009	9.89
Julio	2009	10.12
Agosto	2009	11.33
Septiembre	2009	12.02
Octubre	2009	11.81
Noviembre	2009	13.96

**CONCLUSIONS**

This paper concludes according to the results, that time series under study fits an ARIMA model, this model has two autoregressive factors and one

## CONCLUSIONES

En el presente trabajo se concluye de acuerdo a los resultados, que la serie de tiempo objeto de estudio se ajusta a un modelo ARIMA, dicho modelo posee dos factores auto regresivos y uno de media móvil. Los precios actuales y futuros de esta hortaliza se pueden explicar por sus precios en el pasado.

Con el modelo se hicieron pronósticos para 12 meses de diciembre de 2008 a noviembre de 2009.

## LITERATURA CITADA

- Carter, H. R.; Griffiths, W.; Guay, E. and Lim, C. 2008. Principles of econometrics. Third edition. John Wiley & Sons. USA. 125 p.
- Damodovar, N. G. 2004. Econometría. Cuarta edición. Mc. Graw Hill. D. F., México. 350 p.

moving average. Current and future prices of this vegetable can be explained by their prices in the past.

With the model, forecasts were made for 12 months from December 2008 to November 2009.

*End of the English version*



- Greene, W. H. 2003. Econometric analysis. Fifth edition. Prentice Hall. USA. 215 p.
- Pankratz, A. 1983. Forecasting with univariate Box-Jenkins models. John Wiley & Sons. USA. 125 p.
- Secretaría de Economía (SE). Sistema nacional de información e integración de mercados. SNIIM. 2008. URL: <http://www.secofi-sniim.gob.mx/nuevo/>.
- SAS. 1998. SAS User's guide: Statistics. Release 6.03 Edition. SAS Institute, Inc. Gary, N.C., USA. 1028 p.