Journal of Physics: Conference Series **1448** (2020) 012023 doi:10.1088/1742-6596/1448/1/012023

Comparison of statistical forecasting techniques for Colombian coffee demand in South Korea

R D Rincón[1], W Palacios[1], and H O Paipa[2]
[1] Grupo de Investigación en Productividad y Competitividad, Universidad Francisco de Paula Santander, San José de Cúcuta, Colombia
[2] Grupo de Investigación en Innovación y Gestión Productiva, Universidad Francisco de Paula Santander, San José de Cúcuta, Colombia

E-mail: wlamyrpalacios@ufps.edu.co

Abstract. This article shows a comparison of various methods as a statistical technique for making applied forecasts of Colombian coffee demand in South Korea. The aim is to model the demand behavior in the most adjusted and efficient way possible. To do this, the correlation factor between demand and different macroeconomic variables was analyzed, the one with the greatest relationship is selected and the autocorrelation factor is evaluated. Later, different deterministic methods are used such as linear regression by least squares, simple moving average, weighted moving average, simple exponential smoothing, and exponential smoothing with trend. As a result, a multiple linear regression analysis is obtained, an evaluation of the predictive capacity of the regression model was made through analysis of variance, and the calculation of the standard error of multiple estimation, multiple determination coefficient and the adjusted determination coefficient.

1. Introduction

Colombia is a leading country recognized for growing high quality coffee. The climate and topography of the coffee land, the history, tradition and quality of the Colombian coffee, the different programs of guarantee of origin, the 100% Colombian program and the capacity to develop programs in benefit of clients and consumers in the main markets of the world, are without a doubt some of the strengths that have made of the Colombian coffee a world reference [1].

According to the coffee market report by [2] for the month of May 2018, world coffee exports in April 2018 were 10.18 million bags (60 kg), compared to 9.50 million in April 2017, driven by an increase of 14.1% in exports of robust and 6.8% in those of soft Colombians. World production in 2017-2018 is estimated to be 1.2% higher at 159.66 million bags and is expected to increase in eight of the top 10 coffee producers.

The demand for coffee with special conditions is on the rise, according to the report by [3,4] describe South Korea, a country subject to this research study, as a potential market, in which market dynamics have led to new consumption trends, which are conducive to the export of specialty coffee, a strong product in Norte de Santander, which can meet part of the existing demand.

The world coffee market can be measured through indicators, specifically the trade balance of countries, import and export movements of producers and consumers. [5] describes coffee year 2016-2017 as a period that closed with a production of 156.2 million bags, equivalent to a contraction of -1.9% per year. Due to a deterioration of production in Brazil (main world producer), reaching 51.4

million bags (-8% annually). There, Arabica production was reduced from 43.9 million bags to 38.4 million bags, in line with the year of lower production of the biannual cycle in that country.

Colombia, thanks to the renewal of its coffee plantations during 2010-2014, has managed to reach 14.3 million bags (+2% annually) in its coffee year 2016-2017, despite the delay in flowering in the first half of the year. Particularly for the case of the department of Norte de Santander records an annual production of 22 million kilos of coffee, which are harvested on the slopes of the eastern mountain range, in an area of 24 565 hectares of 36 municipalities.

Toledo, municipality of Norte de Santander, Colombia, located 66 kilometers from the capital of Norte de Santander, included in the area of production of specialty coffees by part [6], has cultivated about 1660 hectares of coffee. In his report [7] presented his commitment to improve the economic and social welfare of more than 18000 coffee producers in the department. The reconversion of traditional and aged coffee cultivation, the improvement of coffee quality, the strengthening of the trade union, the education of coffee growers and new generations, as well as the strengthening of the smart card network, the improvement of coffee social infrastructure, preventive health and income diversification are among the main projects developed by the departmental committee to fulfill the mission and vision of the "Federación Nacional de Cafeteros" of Colombia.

Exports of coffee from this sector are characterized by being green coffee directly by the "Federación Nacional de Cafeteros" which has led in recent years the creation of various associations and coffee cooperatives for processing coffee [8,9].

It is for this reason that this research study is developed as a contribution to the coffee sector of Norte de Santander, Colombia, since the prediction of demand is a problem of great importance for decision-making at the level of production, finance, marketing, among others. and from the results, the most appropriate decisions for their work are made.

2. Materials and methods

2.1. Materials

The process begins with an exploratory analysis of the time series described as a set of historical data on Colombian coffee consumption in South Korea. These data range from 1995 to 2017, provided by the economic complexity observatory, the world's leading visualization engine for international trade data. Figure 1 shows the behavior of Colombian coffee consumption in South Korea in the period described above.

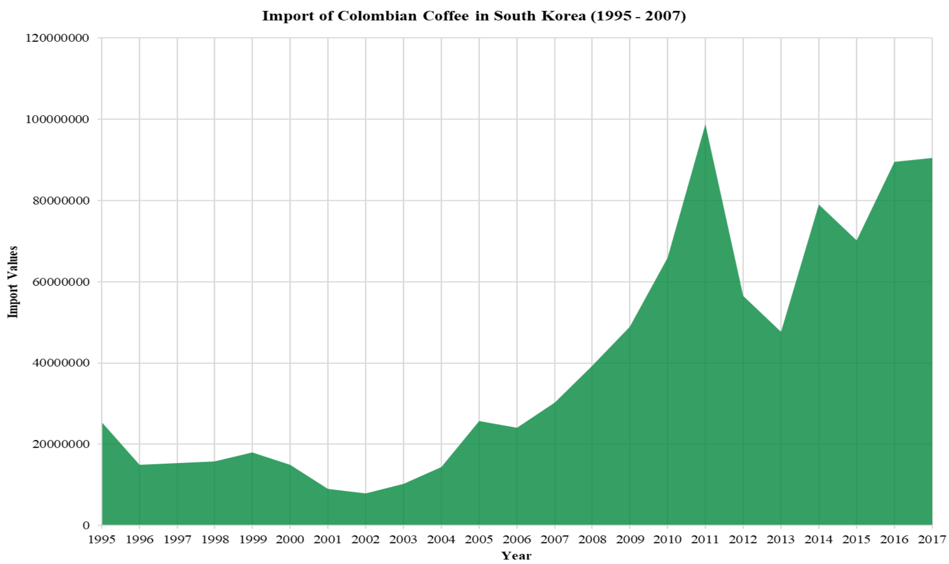

Figure 1. Behavior of Colombian coffee consumption in South Korea from 1995 to 2017.

2.2. Theoretical and methodological framework
The selection of a forecasting method is a process that requires parameters to choose the best method, as proposed [10] in Table 1.

Table 1. Guide to selecting an appropriate forecasting method.

Forecast method	Amount of historical data	Data pattern	Forecast horizon
Simple moving average	6 to 12 months, often weekly data used.	The data must be stationary (*i.e.* no trend or temporality).	Short to medium
Weighted moving average and simple exponential smoothing	To begin with, you need 5 to 10 observations.	Data must be stationary	Short
Exponential smoothing with trend	To begin with, you need 5 to 10 observations.	Stationary and trends	Short
Linear regression	From 10 to 20 observations; for seasonality, at least 5 observations per season.	Stationary, trends and temporality	Short to medium

This study presents a [11] descriptive field research design where Colombian coffee consumption in South Korea is considered as an endogenous z_t variable, i.e. the variable dependent on exogenous or explanatory variables. To provide compliance with the objective of forecasting and modeling the behavior of Colombian coffee consumption in South Korea in the most adjusted and effective way possible [12].

3. Results and discussion
In order to carry out a complete study of the time series, the analyses were grouped into two stages as follows: Qualitative and quantitative analysis of the series, comparison of the forecasting method. Each one with a defined objective, from which are derived the mathematical, statistical and analytical understanding of the series. Next, the processes considered in each of the stages of the study are presented.

3.1. First stage. Qualitative and quantitative analysis
In this first phase, the qualitative and quantitative analysis of the series was carried out, identifying the components of the series such as trend and seasonality. The descriptive analysis of the series allows us to know the road to travel to determine the model and its estimates.

3.1.1. Description of the series. The series presented in Figure 1 shows a strong upward trend between 2006 and 2011 corresponding to the exponential growth of coffee consumption in South Korea, due to the opportunities for entry of this product to the market of the Asian country due to the free trade agreement after a mild decline between 2000 and 2005. A clear component of positive trend can be seen. In the following section, each component can be seen separately [13].

3.1.2. Identification of the components of the series. According to Figure 2 and Figure 3, for the determination of the seasonal component, a first approximation was made eliminating heteroscedasticity by means of a logarithmic transformation, applying natural logarithm, since the series presents a behavior similar to a multiplicative seasonal scheme that requires deseasonalization of the data.

To define the function of simple correlation or correlogram, and according to Figure 4, it was practiced the methodology of [14] who expresses that, the coefficient of linear correlation between the variables Z_t and Z_{t-1}, is called the coefficient of linear autocorrelation of order 1 (r_1) whose calculation needs to obtain the time series offset by a unit of time. The calculation of the value of the coefficient of linear autocorrelation was obtained by means of Equation (1), sample correlation.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (1)$$

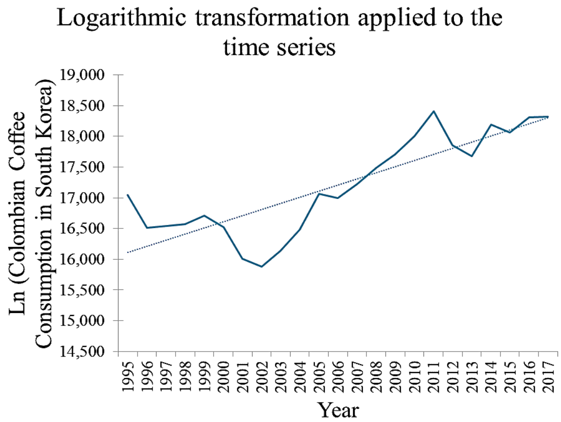

Figure 2. Seasonal component of the time series.

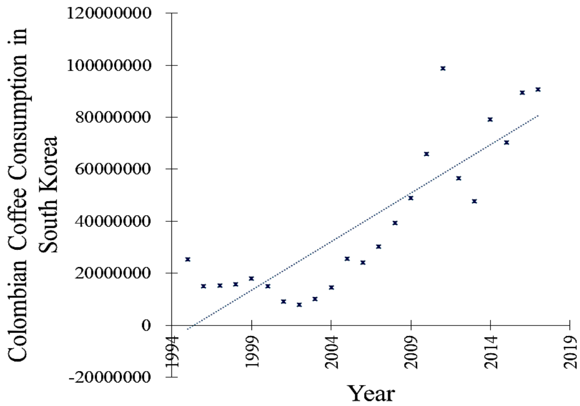

Figure 3. Component of the trend.

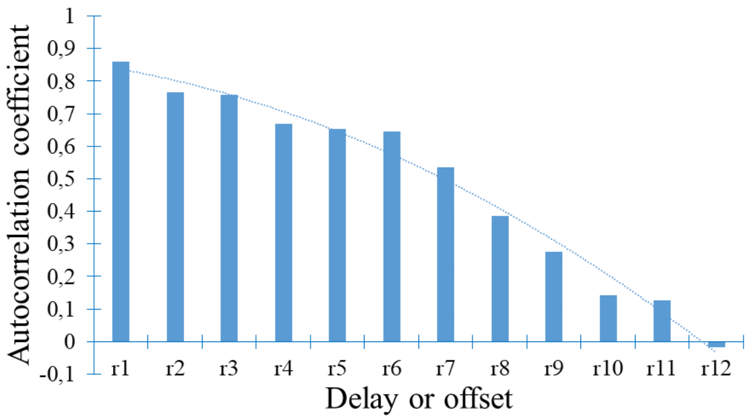

Figure 4. Linear autocorrelation component of order 20 (r_{20}).The Z_t series has been decalibrated up to 20 time units.

According Figure 2, the time series presents a pattern of behavior similar to a multiplicative seasonal scheme that requires deseasonalization of the data, therefore, was applied natural logarithm ($\ln Z_t$) in order to eliminate heteroscedasticity, thereby reducing seasonality and evidence the trend behavior. To Figure 3, The series presents a clear trend. Therefore, a linear deterministic trend seems suitable for the Z_t series. And to Figure 4, with the help of Microsoft Excel ® the autocorrelation coefficients of order 1 (r_1) between the Z_t and Z_{t-1} variables have been obtained, up to order 20 between the Z_t and Z_{t-20} variables. Where it is observed in Figure 4 that the correlogram shows that the autocorrelation structure is decreasing between more and more separated instants of time which means that the dependence of the series with time is less and less.

3.1.3. Quantitative analysis. A quantitative analysis was developed applying the guide presented by [8] to select a prognostic method where Colombian coffee consumption in South Korea would be the dependent variable and the time or course of the years would act as the independent variable, by the linear regression method, as well.

When eliminating heteroscedasticity by means of a logarithmic transformation in the series of Colombian coffee consumption in South Korea, one can observe a behavior of the trend pattern, behavior that was generated once the seasonality was partially eliminated in the temporal series, in

Figure 3. It is evident that the trend is of the directly proportional type, with a pronounced ascent in the demand, therefore it was proceeded to make the calculations with the method of linear regression according to [15] determine the linear equation, as well, least squares estimators for the simple linear regression model. Equation (2) indicates the linear base mathematical model, Equation (3) indicates how the estimators are calculated, and Equation (4) is the union of the base equation and the estimators.

$$\widehat{y_i} = \widehat{\beta_0} + \hat{\beta}_1 x_i \quad (2)$$

$$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}, \text{donde } S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \text{ y } S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (3)$$

$$\widehat{\beta_0} = \bar{y} + \hat{\beta}_1 \bar{x} \quad (4)$$

The estimates of least square for the slope and points of intersection of the adjusted line were calculated, being then that, the linear regression equation is obtained, with $\widehat{\beta_1} = -182.446$ and $\widehat{\beta_0} = 0.0995$ therefore, Equation (5):

$$\widehat{y_i} = -182.446 + 0.0995 x_i \quad (5)$$

The predictive capacity of Equation (5) needs to be evaluated, using an inference or significance test to establish the relationship between the two variables. Significance test of the parameter called slope of the line. The regression Equation (5) is then analyzed using a hypothesis test to see if the slope of the regression line is other than zero [16,17]. According Equation (6) to the procedure expressed by [9] hypothesis test the null hypothesis and alternative are:

$$t = \frac{\beta_1 - 0}{S_b} \quad (6)$$

It is established that,

$$H_0: \beta_1 \leq 0 \quad (7)$$

$$H_1: \beta_1 > 0 \quad (8)$$

Knowing that, Equation (6) should be applied with $n - 2$ degrees of freedom. With the help of Microsoft Excel® Software you get that $t = 7.11008$ The distribution t is the test statistic, there are 21 degrees of freedom, determined by $n - 2 = 23 - 2 = 21$. So, for Equation (7) and Equation (8), the level of significance of 0.05 is used, according to the distribution of t Student for 21 degrees of freedom with significance of 0.05, the critical value is 1.721. The decision rule is to reject the null hypothesis if the value calculated with Equation (6) is greater than 1.721. The calculated value of 7.110088 exceeds the critical value of 1.721, so the null hypothesis is rejected and the alternative hypothesis is accepted, in conclusion, the slope of the line is greater than zero. The independent variable, which refers to time, is useful to obtain a better estimate of Colombian coffee consumption in South Korea.

3.2. Forecast method comparison

Below are the estimates of the statistical models used, with the respective characterization of Colombian coffee consumption in South Korea, according to Table 2.

According Figure 5, the results of the forecasts are shown by means of statistical techniques such as: simple moving average with $n = 3$ and $n = 5$, in the same way the averages weighted with $w_1 = 0.5$ $w_2 = 0{,}3$ $w_3 = 0.2$ and with $w_1 = 0.4$ $w_2 = 0.25$ $w_3 = 0.15$ $w_4 = 0.12$ $w_5 = 0.08$. Where the best adjustment according to the mean absolute percentage error (MAPE) is the simple moving average with $n = 5$. And according Figure 6, the adjustments of forecasts are shown by means of statistical techniques such as: simple exponential smoothing and exponential smoothing with trend. The best adjustment according to the MAPE is the simple exponential smoothing with $\alpha = 0.20$.

Considering the characteristics of the series, which presents a multiplicative seasonal behavior, and with a linear tendency, it is advisable in practice to adjust it with another additional exogenous variable, this can be a macroeconomic variable, the international price of coffee, the employment rate in South Korea, the growth of exports in Colombia, among others.

Table 2. Comparison of the forecast methods used. Synthesis of results.

Forecast method	Parameters	MAPE forecast (%)	Forecast for 2018
Simple moving average	$n = 3$	2.080	18.233
	$n = 5$	2.615	18.113
Weighted moving average	$w_1 = 0.50$ $w_2 = 0.30$ $w_3 = 0.20$	1.899	18.267
	$w_1 = 0.40$ $w_2 = 0.25$ $w_3 = 0.15$ $w_4 = 0.12$ $w_5 = 0.08$	2.192	18.213
Simple exponential smoothing	$\alpha = 0.20$	2.687	17.97
	$\alpha = 0.30$	2.247	18.130
	$\alpha = 0.40$	2.013	18.205
Exponential smoothing with trend	$\alpha = 0.40$ $\delta = 0.30$ $T_i = -0.16$	1.805	18.414
	$\alpha = 0.30$ $\delta = 0.30$ $T_i = -0.16$	2.205	18.530
	$\alpha = 0.20$ $\delta = 0.40$ $T_i = -0.16$	2.626	18.730
	$\alpha = 0.10$ $\delta = 0.25$ $T_i = -0.16$	3.843	18.563

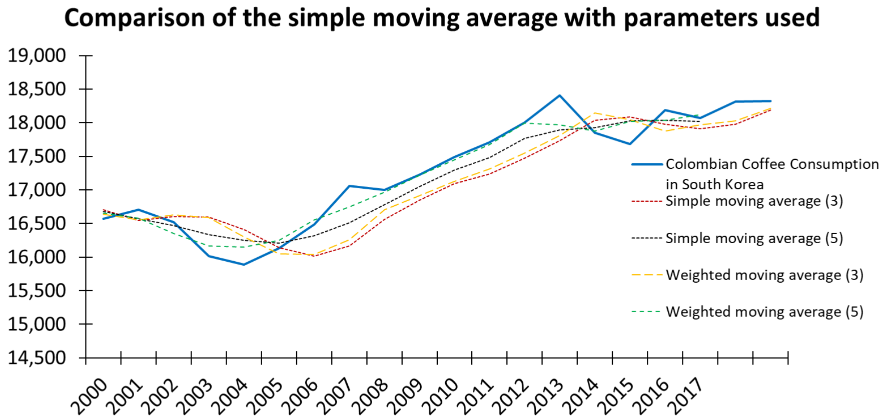

Figure 5. Behavior of Colombian coffee consumption in South Korea and its forecasts.

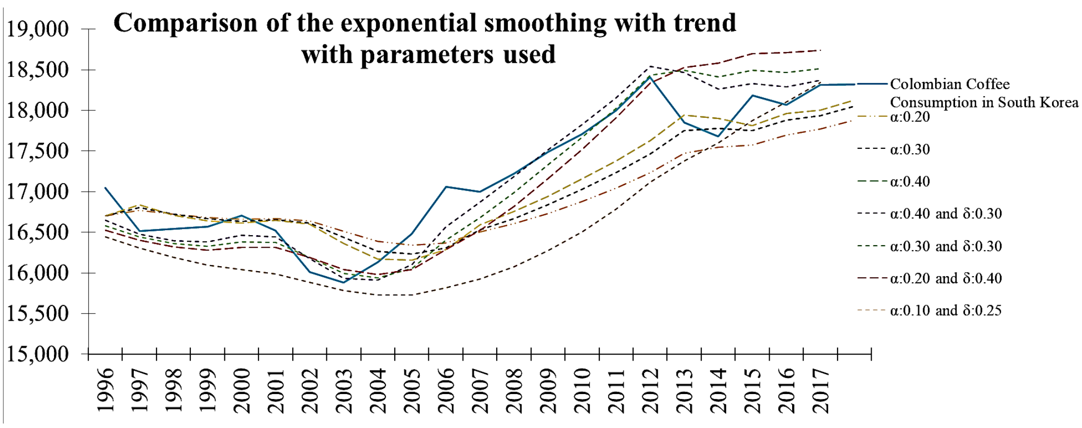

Figure 6. Behavior of Colombian coffee consumption in South Korea and its trend forecasts.

4. Conclusions

Of all the models estimated and analyzed, it was found that the simple exponential smoothing model with a parameter of $\alpha = 0.20$ was the most optimal to achieve the lowest forecast error, adjusting more to the reality of Colombian coffee consumption in South Korea. This is due to the adjustment of the parameter and to the reduction of seasonality by means of the application of the natural logarithm, since the true forecast for 2018 is not 17.97 because this value must be applied to the antilogarithm and will show the real value in monetary units being this $ 63520959.

The presented temporal series evidences the existence of a non-additive seasonal component, therefore an approximation was made eliminating heteroscedasticity by means of a logarithmic transformation, applying natural logarithm, since the series presents a behavior similar to a multiplicative seasonal scheme that requires deseasonalization of the data.

The autocorrelation component of the series was analyzed shifted the same linearly in an order of 12 units of time, which indicates that the structure of autocorrelation and the correlogram is concluded that this decreases between instants of time increasingly separated what is interpreted as the dependence of the series with time is less and less.

According to the significance test and the evaluation of the predictive capacity of the linear regression, the applied parameters and adjustments, it has that, the independent variable, that refers to the time, and the relation with the consumption of Colombian coffee in South Korea, obtain a good estimate and turns out to be useful for the forecast of later periods in the short term.

References

[1] Federación Nacional de Cafeteros 2018 *Comportamiento de la industria cafetera colombiana* (Bogotá: Federación Nacional de Cafeteros) p 4
[2] Federación Nacional de Cafeteros 2018 *Informe de sostenibilidad 2015 – 2018* (Bogotá: Federación Nacional de Cafeteros) p 41
[3] Comisión de Promoción del Perú para la Exportación y el Turismo 2017. *Informe ejecutivo de feria café show* (Lima: Comisión de promoción del Perú para la exportación y el turismo) p 4
[4] E Baena, J J Sánchez y O Montoya Suárez 2003 El entorno empresarial y la teoría de las cinco fuerzas competitivas *Scientia et Technica* **9(23)** p 61
[5] Federación Nacional de Cafeteros 2014 *Ensayos sobre economía cafetera* (Bogotá: Federación Nacional de Cafeteros) p 37
[6] Federación Nacional de Cafeteros 2009 *Informe Ejecutivo Comité Departamental* (Bogotá: Comité Departamental de Cafeteros Norte de Santander)
[7] R B Chase y F R Jacobs 2014 *Administración de operaciones, producción y logística, 13ª Edición* (Ciudad de México: McGraw Hill Editores) p 485
[9] D A Lind, W G Marchal y S A Wathen 2015 *Estadística aplicada a los negocios y la economía, 16ª Edición* (Ciudad de México: McGrawHill Editores) p 250
[10] A G Pérez 2015 *Guía metodológica para la presentación de anteproyectos de investigación, 1ª Edición* (Caracas: Universidad Pedagógica Experimental Libertador) p 23
[11] J C García Díaz 2011 *Series temporales, análisis, predicción. Ejercicios prácticos*, *1ª Edición* (Madrid: Universitat Politécnica de València) p 105
[12] D D Wackerly, W Mendenhall III y R L Scheaffer 2008 *Estadística matemática con aplicaciones*, *7ª Edición* (Ciudad de México: Cengage Learning Editores S.A.) p 302
[13] R Prada Nuñez y C A Hernández Suárez 2015 Análisis de una serie de tiempo utilizando diseño de experimentos como herramienta de calibración *Ecomatemático* **6(1)** 50
[14] A M Alonso y C García Martos 2012 *Análisis de series de tiempo* (Madrid: Universidad Carlos III de Madrid) p 20
[15] C Stephen N 2006 *Planificación y control de la producción, 1ª Edición* (Ciudad de México: Pearson Educación) p 55
[16] L J Krajweski y L P Ritzman 2000 *Administración de operaciones. Estrategia y análisis*, *5ª Edición* (Ciudad de México: Pearson Educación) p 491
[17] D Sippper y R L Bulfin Jr 1999 *Planeación y control de la producción, 1ª Edición* (Ciudad México: McGraw Hill Editores) p 96