AN ECONOMETRIC MODEL OF CANADA'S NEWSPRINT EXPORTS TO THE UNITED STATES OF AMERICA

by

Junchang Liu ©

A Graduate Thesis Submitted in
Partial fulfillment of the Requirement
for the Degree of Master of Science in Forestry

School of Forestry

Lakehead University

Thunder Bay, Ontario, May 1990

ISBN 0-315-63203-8

DEDICATION

To the College of Forest Economic Management
Beijing Forestry University, People's Republic of China

ABSTRACT

Liu, Junchang 1990. An econometric model of Canada's newsprint exports to the United States of America. Master of Science (Forestry). 124 pp. Major advisor: Dr. G. Hazenberg.

Key words: demand, econometric model, forecasting power, ordinary least-squares methods (OLS), simultaneous-equation model, supply, two-stage least-squares methods (2SLS).

An econometric model was built for predictive purposes and for understanding the relationship between the US market and Canadian newsprint producers. In this study, a simultaneous-equation model was developed, which consists of four equations, one each for the supply, export, consumption and price of newsprint. The period of 1955-1986 was covered by this model. The data for the years 1985 and 1986 were reserved to test the predictive power of the model. In fitting the four equations for the period of 1955-1984, the coefficients of determinations, the R-square values, between observed and predicted values were higher than 99 per cent. The test results of the forecasting power showed that there was no difference between predicted and observed values at the 5 per cent level of significance. Sources of forecasting error are expressed as three partial inequality coefficients associated with bias, variance and covariance of predicted and observed values. In the present forecast, the major source of error came from the covariance between the predicted and observed values. This could not be further reduced. The model can be used to make annual predictions and provides an accurate means of predicting demand, supply, consumption and price of Canadian newsprint exports to the United States of America.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I am greatly indebted to Beijing Forestry University, People's Republic of China, for financing my studies at Lakehead University, and the School of Forestry Graduate Studies Committee, Lakehead University, for accepting me as a graduate student.

My special thanks are due to Dr. G. Hazenberg, School of Forestry, Lakehead University, whose constant advice, supervision and direction ensured the successful completion of this study. I am also indebted to Dr. K.M. Brown and Professor K.C. Yang, School of Forestry, Lakehead University, who together with my supervisor constituted my advisory committee.

I would also like to thank Dr. B. Moazzami and Professor F.J. Anderson, Department of Economics, Lakehead University, for their advice during data selection and estimating the model.

Last but not least I would like to express my heartfelt gratitude to my father and mother-in-law for looking after my daughter Jingran Liu while I was away from home during the course of this study.

Finally I would like to thank my colleagues and all the people who in one way or another contributed to the successful completion of this study.

J. C. Liu

1 INTRODUCTION

Canada's newsprint industry has developed very rapidly in this century. In 1913, Canada only produced ninety thousand tonnes of newsprint. In 1988 Canada produced about ten million tonnes of newsprint, which accounted for one-third of the world's total supply of newsprint (CPPA, 1988). This growth is due to the existence of foreign markets. Canada's domestic market is relatively small; most newsprint is exported to other countries. It was reported that 86 per cent of newsprint was exported to over 50 countries in 1987. This export earned almost one-half the nation's net trade surplus (CPPA, 1988).

Among all of the foreign countries which purchased Canada's newsprint, the United States of America (USA or US) purchased about three-quarters of the Canadian newsprint production (CPPA, 1988). Therefore, Canada's newsprint industry has a strong relationship with the US market. If the newsprint market of the USA were to change, it would have a strong influence on Canada's newsprint makers. This can be seen from the history of Canada's newsprint industry which largely developed after 1913, the year in which the US government eliminated its tariff on newsprint imports. Since the demand by US consumers has been decreasing, Canada's newsprint makers face, at present, a decreasing market trend. It is expected that Canada will have more idle capacity in the entire industry and that about 10 per cent of available capacity will be shut down because of slumping US demand*. Therefore, investigating the influence of the US market on Canada's newsprint industry is necessary for understanding the development of the newsprint industry and its possible future direction.

This study investigated the factors of the US market which influence Canada's newsprint industry. By using a number of econometric test methods, the most important factors have been isolated. An econometric model was built which consists of four equations, one each for the supply, export, consumption and price of newsprint. The period of 1955-1986 was covered.

The objectives of this study were as follows:

(1) to find the most important variables which influence Canada's newsprint industry;

(2) to develop an econometric model that would predict the demand, supply and price of Canadian newsprint exports to the USA.

After the introduction, this study is presented in six

---

* Atkinson. S., analyst, McNeil Mantha Inc. cited in a CP report 'Bleak future forecast for newsprint makers' The Chronicle Journal 14 July 1989, p.1.

chapters. These chapters consist of a historical background, a literature review, chapters on methodology, results, discussion and the last chapter contains recommendations. Chapter 2 provides a short history of Canada's newsprint industry and the newsprint market structure both inside and outside Canada. The chapter on literature review summarizes the available literature related to this study and reviews some economic theory used in building the model. The chapter on methodology deals with the econometric method used to specify the model, to test the specification error, to estimate the parameters and to test the forecasting power of the model. The discussion chapter will consider some limitations in building the model. The recommendations chapter provides a brief summary of this study and some recommendations for using this model and suggestions for further study.

2 BACKGROUND

2.1 History of the Canadian newsprint industry

The expansion of the Canadian newsprint industry is closely connected with the US market. Before the US government eliminated its tariff on newsprint imports in 1913, Canada's newsprint industry developed very slowly. The production of newsprint accounted for a very small part of pulp and paper production during that time. After 1913, the newsprint industry expanded quickly. This growth of pulp and paper between 1900 and 1920 can be seen from Table 2.1.

Table 2.1. Canadian pulp and paper industry: summary statistics 1900 and 1920

Type of production	1900		1920	
	('000 t)	(%)	('000 t)	(%)
Newsprint	–	–	876	43
Market pulps	75	38	820	40
Other papers & paperboard	125	62	339	17
Total	200	100	2035	100

Source: Davis *et al*, 1957.

Table 2.1 shows the difference of pulp and paper production before and after removal of the import duty for newsprint into the USA for the two selected years. In 1900, the newsprint production was so little that it was ignored in statistical reports. In 1920, the newsprint production reached 876 thousand t, which accounted for 43 per cent of the output of the pulp and paper industry.

Table 2.2 indicates the Canadian newsprint production, capacity and operating rate for the period of 1919 - 1984 in intervals of 5 years. It shows that the newsprint industry continued to develop rapidly in the 1920's. By the end of that decade the newsprint capacity in Canada reached 3.5 million t. In 1929, newsprint production was about 3 million t (CPPA, 1963). But in the 1930's, the newsprint industry underwent a difficult period. It was reported that "more than one-half of the productive capacity went into receivership as the depression progressed" (Davis *et al*, 1957). The operating rate was 69 per cent in 1939, which had decreased 16.5 per cent from 1929 to 1939. By the end of the 1930's, production was about 3.2 million t (CPPA, 1963). Newsprint production only increased by 230 thousand t during this ten year interval.

Since World War II the Canadian newsprint industry expanded rapidly again because of the postwar economic development in the world. As Table 2.2 shows, the newsprint capacity totalled 9.8 million t in 1984. Production that year reached 9 million t. Canada still remains the world's

Table 2.2: Canadian newsprint capacity, production and operating rates: 1919 - 1984

Year	Capacity ('000 t)	Production ('000 t)	Operating Rate (%)
1919	905	849	93.8
1924	1638	1418	86.5
1929	3521	2984	85.0
1934	4182	2911	69.6
1939	4633	3175	68.5
1944	4726	3265	69.1
1949	5113	5176	101.2
1954	5920	5984	101.1
1959	7521	6394	85.0
1964	7506	6623	88.2
1969	8342	7500	89.9
1974	8899	8661	96.3
1979	9063	8756	96.6
1984	9820	9015	91.8

Source: CPPA, 1957, 1964, 1985.

largest newsprint producer and exporter. It accounts for over 60 per cent of the newsprint traded internationally, exporting more than five times as much as Sweden, its closest competitor (CPPA, 1988).

2.2 The newsprint market

2.2.1 The domestic market

The demand for newsprint in the domestic market has increased in both absolute and relative terms since 1955. In 1955, shipments within Canada were only 390 thousand t, which was 7 per cent of the production in that year. In 1986, shipments reached 1.1 million t and accounted for 12 per cent of the production in that year. The domestic market plays at present a more important role for Canadian newsprint makers than before. While it is true that domestic consumption increased considerably, it is still relatively small compared with foreign demand for Canadian newsprint. Therefore, the domestic market affects the prospective over-all demand outlook rather less. This does not mean that the domestic market is not important for Canadian newsprint makers. It has a strong attraction for Canadian newsprint producers because of its relatively greater stability to them. Table 2.3 indicates the development of the domestic market for the period from 1913 to 1984.

Table 2.3 shows that domestic consumption increased quite rapidly from 1913 to 1929. During the 1930's, the consumption fell from about 188 to 172 thousand t. By 1944, the total consumption had nearly returned to the 1929 level, but the per capita consumption was still lower than in 1929. From 1944 to 1984, Canadian consumption increased almost

Table 2.3: Domestic consumption of newsprint: selected years 1913-1984

Year	Total ('000 t)	Per capita (kg)	Year	Total ('000 t)	Per capita (kg)
1913	28	10.6	1949	304	22.5
1915	64	7.9	1954	382	24.9
1919	86	10.3	1959	473	24.3
1924	114	12.4	1964	550	25.7
1929	188	18.6	1969	709	30.6
1934	157	14.5	1974	886	35.8
1939	172	15.1	1979	971	41.0
1944	170	14.1	1984	1070	42.5

Source: Davis *et al* (1957) and CPPA (1978, 1987).

without abatement. Since 1959, the domestic consumption growth continued at a slow but steady rate and increased 126 per cent over the 1959 - 1984 period. It is predicted that the consumption will be between 1.34 and 1.43 million t by the year 2000 (Roberts and Luck, 1985).

2.2.2 Foreign markets

Foreign countries are the main markets for Canadian newsprint producers. More than 80 per cent of Canadian newsprint production is exported. In 1987, exports accounted for 86 per cent. This quantity of export makes Canada the largest newsprint supplier in the world. On a world basis, Canada accounted for 61 per cent of the world exports of newsprint in 1986 (CPPA, 1987).

Table 2.4 summarizes the main trends for the period 1971-1984; these suggest the overriding importance of the USA in the global newsprint consumption and Canada's role in satisfying this demand in recent years.

For the period 1971 to 1984 world newsprint demand increased by almost 8.3 million t or 43 per cent. The supply from Canada increased by 1.9 million t or 27 per cent. In 1984, Canada supplied almost 33 per cent of the world's need. Since the rate of demand increased faster than the growth rate of Canadian newsprint supply, the relative proportion of Canada's supply in the world decreased from 37 per cent in 1971 to 33 per cent in 1984. The main reason of the decline is the decreasing share of the US market. This situation will be discussed in section 2.4.

Table 2.4 shows that the USA is the major market for newsprint. It already occupied this leading position in the 1920's. During that time the USA accounted for more than half of the world demand. And now it is still the largest newsprint consumer in the world. In 1984, it required 11.6 million t or about 42 per cent of the world demand.

The current structure of the world newsprint market is shown in Tables 2.5 and 2.6. Table 2.5 shows Canada's newsprint

Table 2.4: World and USA newsprint demand and Canadian total shipments: 1971-1984

Year	Demand			Canadian total shipments	
	World ('000 t)	USA ('000 t)	(%)	('000 t)	% of world shipment
1971	19229	8309	43.2	7112	37.0
1972	20017	8766	43.8	7570	37.8
1973	21319	9487	44.5	8012	37.6
1974	23295	9634	41.4	8706	37.4
1975	21052	8245	39.2	7010	33.3
1976	22295	9018	40.4	7903	35.4
1977	22716	9141	40.2	8169	36.0
1978	24001	9776	40.7	8913	37.1
1979	25096	10106	40.3	8779	35.0
1980	25836	10301	39.9	8622	33.4
1981	26660	10547	39.6	8913	33.4
1982	24441	9889	40.5	8054	33.0
1983	25557	10510	41.1	8439	33.0
1984	27539	11581	42.1	9019	32.7

Source: CPPA, 1981, 1987.

industry as a leading producer in the world. Table 2.5 also shows the importance of the USA as a consumer country. The newsprint consumption by the USA accounted for 43 per cent of the world's newsprint consumption. Table 2.6 indicates that Canada is a major exporter of newsprint in the world.

Table 2.5: The major newsprint producer and consumer countries in the world: 1984

Country	Production		Country	Consumption	
	('000 t)	(%)		('000 t)	(%)
Canada	9015	33.0	Canada	1070	3.9
USA	5030	18.4	USA	11441	42.8
Finland	1543	5.6	Finland	589	2.2
West Germany	714	2.6	West Germany	1266	4.7
Norway	799	2.9	UK	1407	5.2
Sweden	1533	5.6	Indonesia	2660	9.8
Japan	2550	9.3	Australia	600	2.2
USSR	1400	5.1	USSR	1057	3.9
Sum	22584	82.5	Sum	20090	74.7
World	27344	100.0	World	27126	100.0

Source: CPPA, 1985a and b.

As mentioned before, by far the greatest amount of Canada's newsprint export went to the USA. If export to the USA were excluded, the remainder, i.e. 1.4 million t in 1984, still makes Canada the leading exporter to other foreign markets.

Table 2.6 also shows the USA as the largest importer of newsprint in the world. In 1984, its imported quantity accounted for 50.8 per cent of total world import. It is far ahead of other importing countries. It is interesting to note from Table 2.5, that the 8 leading producers and 8 top

Table 2.6: The major newsprint exporting and importing countries in the world: 1984

Country	Export		Country	Import	
	('000 t)	(%)		('000 t)	(%)
Canada	7965	62.5	USA	6474	50.8
USA	260	2.0	France	332	2.6
Finland	1317	10.3	West Germany	725	5.7
West Germany	172	1.3	The Netherlands	296	2.3
Norway	655	5.1	Nigeria	220	1.7
Sweden	1155	9.1	India	196	1.5
New Zealand	174	1.4	Japan	215	1.7
USSR	350	2.7	Australia	223	1.7
Sum	12048	94.4	Sum	8681	68.0
World	12744	100.0	World	12743	100.0

Source: CPPA, 1985a and b.

consumer countries accounted for more than 70 per cent in their respective categories. If Tables 2.5 and 2.6 were combined, 16 countries accounted for more than 65 per cent of all newsprint trade in 1984. The two tables have been set up along the lines suggested by Davis *et al* (1957). In Tables 2.4, 2.5 and 2.6, the world newsprint market has been outlined. It is clear that the demand and supply have been rising without apparent abatement. Among only a few newsprint producer and consumer countries, Canada stands out as the major producer country and the USA as the major consumer country. The trade in newsprint between them is the subject of this study.

2.3 Market structure of newsprint in North America

The market of newsprint in North America resembles a competitive market. As Schaefer (1979) pointed out, that although the oligopsonistic element may be significant, the market structure of newsprint in North America approximated more closely a competitive industry than an oligopolistic one in 1970's. This is still true. Today both sides (producer and buyer) have some power to decide the newsprint price.

In the period from 1930 to 1950, the market resembled an oligopsonistic market structure. The newspaper publishers did not have much power to influence price. Schaefer (1979) pointed out:

"In 1958, US publishers controlled 8.7% of Canadian and 28.8% of US newsprint capacity for a total of 13.3% of North American capacity. ... In terms of concentration ratios, the four largest firms in North American held 38.9% of capacity in 1958 and the eight-firm ratio rises to about 50%."

This situation showed that the market was an oligopoly. But after 1970, the situation has changed since new capacity, technical change and a slower growth of demand gave purchasers an advantage. Some buyers could force producers to assume part of the storage costs usually borne by the purchaser.

2.4 The United States demand for Canadian newsprint

As stated previously, the USA purchases most of the Canadian newsprint exported. Among export of Canadian newsprint, almost three-quarters goes to the USA, which accounted for 57 per cent of their newsprint requirements in 1988 (CPPA,1988).

This heavy demand of the USA for Canadian newsprint has brought about a great advance in the Canadian newsprint industry. On the other hand, the US customers have obtained more and more newsprint from Canada. Table 2.7 summarizes the historical trends of US newsprint supply by source since 1913.

Table 2.7 shows that most of the US newsprint supply comes from Canada and from its own domestic producers. Other countries only play a minor role. Since 1930, Canada has been supplying more than half of the US total market each year. Before 1950, the part played by Canadian mills in the supply of the USA increased not only in terms of tonnage but also in terms of their proportion of total supply, because of the decline of USA production. After 1950, the situation changed. The production of newsprint in the USA has

Table 2.7: USA newsprint supply by sources: selected years 1913 - 1984

Year	Canada	USA	Europe	Total	Canada	USA	Europe
	(←--------'000 t------------→)				(←---------%---------→)		
1913	218	1255	-	1473	14.8	85.2	0.0
1915	329	1176	4	1509	21.8	77.9	0.3
1920	679	1466	51	2196	30.9	66.8	2.3
1925	1315	1508	133	2956	44.5	51.0	4.5
1930	2145	1272	134	3551	60.4	35.8	3.8
1935	2122	911	197	3230	65.7	28.2	6.1
1940	2741	998	339	4078	67.2	24.5	8.3
1945	2666	707	0	3373	79.0	21.0	0.0
1950	4307	909	155	5371	80.2	16.9	2.9
1955	4599	1246	132	5978	76.9	20.8	2.2
1960	4789	1773	133	6695	71.5	26.5	2.0
1965	5528	1980	231	7739	71.4	25.6	3.0
1970	5591	3008	285	8884	62.9	33.9	3.2
1975	4980	3165	32	8177	60.9	38.7	0.4
1980	6118	4058	125	10301	59.4	39.4	1.2
1984	6586	4807	189	11582	56.9	41.4	1.7

Source: CPPA, 1956, 1987.

increased. Therefore, the percentage of Canadian newsprint supply has been falling although the supply in absolute terms is still increasing.

There is little doubt that supplies from other countries will not become of importance to the USA. Hence, the proportion of Canada's newsprint supply to the USA depends on US production. If the growth of newsprint production in the USA does not exceed the growth of demand and other variables remain constant, the ratio of Canada's newsprint supply can be expected to increase. But since 1965 the relative share of the Canadian newsprint industry of the US market has decreased.

"The decline of the Canadian market share in the United States is due to the expansion of U.S. capacity especially in the southern states. The rapid growth of U.S. capacity since World War II was made possible by technological advances in the 1930's which made it feasible to use southern pine in the manufacture of newsprint." (Schaefer, 1979).

It is to be expected that if the market for newsprint continues its growth, Canada will again increase its capacity substantially by the year 2000. Exports of newsprint to the USA are forecast to be 8.7 million t in the year 2000 (Manning and Grinnell, 1971).

3 LITERATURE REVIEW

Econometrics is one of the most important tools in economic research. It "deals with the measurement of economic relationships" (Koutsoyiannis, 1977). As Gujarati (1988) pointed out: "although measurement is an important part of econometrics, the scope of econometrics is much broader". It "may be considered as the integration of economics, mathematics and statistics for the purpose of providing (for example, elasticities, propensities, marginal values) and verifying economic theories" (Koutsoyiannis, 1977).

In economic studies, hypotheses are formulated based on economic theory. Econometric methods are used to test the hypotheses and express economic relationships in quantitative terms. Through econometric studies, not only relationships between economic variables, but also the degree of the relationships, can be found. From statistical functions, forecasts are possible for economic phenomena. Therefore, econometric studies have become quite popular in recent economic research. Wallis (1980) dealt with consumption, production and investment functions and provided some ideas for building such econometric models. Choudhry *et al* (1972), for example, used econometric methods to study the Canadian economy and set up a non linear econometric model of the Canadian economy.

3.1 General econometric studies

3.1.1 Econometric models

"An econometric model is a set of equations to provide a quantitative explanation of the changes in economic variables" (Buongiorno and Gilless, 1987). Models represent the major characteristics of an economic problem and ignore others not relevant to a particular problem. Through a model, a theory can be tested so that a complex economic situation can be clearly understood. As Neal and Shone (1976) pointed out:

> "a theory cannot be tested directly since it involves theoretical concepts which are not observable. However, it is possible to test the theory indirectly through a model (or models). The predictions yielded by such a model will then provide an indirect test of that theory."

Econometric models can be divided into two categories according to the data used in the model. One category consists of short-run models, the other of long-run models. Short-run models usually use monthly or quarterly data. Long-run models use annual observations (Mills and Manthy, 1974). An example of a short-run model was given by Davidson *et al* (1978). Long-run models were provided, among others, by Davis (1952) and Brown and De Cani (1963).

Econometric models can also be distinguished by the kind of problem which is studied. If the problem belongs to macro economics, a macro econometric model should be formed. If the study problem is one of micro economics, a micro econometric model should be specified.

As well, econometric models can be divided, on technical grounds, into single-equation and simultaneous-equation models. The single-equation model contains only one equation in each model such as a demand or supply equation. In the equation, there is a one-way causation between the dependent variable y and a set of explanatory variables X_i, i=1---p [$y=f(\underline{X})$]. Simultaneous-equation models contain more than one equation; for example, a market model may contain demand, supply and price equations. In this kind of model, there is a two-way causation [not only $\underline{y}=f(\underline{X})$ but also $\underline{X}=f(\underline{y})$] in a function. An example of a simulataneous equation model was given by Suits (1955).

3.1.2 Econometric techniques

There are a number of econometric methods used to deal with different problems. For single equation models, ordinary least-squares methods (OLS) and generalized least-squares methods (GLS) are often used. For simultaneous-equation models, reduced-form techniques, two-stage least-squares (2SLS) and three-stage least-squares (3SLS) methods are usually applied.

3.2 Forest econometric studies

3.2.1 General forest econometric studies

Econometrics, as an important economic technique, is widely applied in forest economic research. In 1967, Hair used regression methods to study the trends in demand for paper and board in the USA. As a result of that study, regression models were built for the US paper and board industry. McKillop (1969) investigated the short-run market structure for redwood lumber in the USA by means of econometric techniques. Through his research, he established an econometric market model, for redwood lumber, which provided a short-run explanation of the market and its constituent relationships. Meyer (1979) applied econometric methods for studying the timber markets of Switzerland. As the result of his study, a supply, a demand, an export and a production equation for Swiss roundwood were obtained.

Canadian forest economists also carried out a number of econometric studies and provided models for different products. Manning (1975) provided a market model of the Canadian softwood lumber industry. In this study, single equation methods were applied. Bulger (1986) set up a simultaneous-equation model for the softwood lumber industry in British Columbia (BC). In his study, he used simultaneous-equation methods to study the relationship between the US market and the BC softwood lumber industry and provided demand, consumption, supply and price equations. Jacques (1988) used Statistics Canada input-output models to assess the links between the forest sector and other sectors of the economy.

3.2.2 Econometric studies of the newsprint industry

Econometric models of the newsprint industry were built by several researchers. Schaefer (1979) studied the Canadian newsprint industry. In his study, a set of equations were developed which dealt with both competitive and oligopolistic market structures. Roberts and Luck (1985) built a single equation econometric model of newsprint consumption for a prediction of Canadian newsprint consumption for the year 2000. Johnson (1985) published a research report, in which a model of the US newsprint consumption was given. Ghebremichael (1989) published a study about the demand equation of the USA for Ontario's newsprint. In this study an econometric analysis for Ontario's newsprint industry was carried out. In all of the studies mentioned, single equation methods were applied.

4 METHODOLOGY

As mentioned earlier, econometric models can be divided into single-equation and simultaneous-equation models according to the relationship between the set of dependent variables ($\underline{y}$) and explanatory variables ($\underline{X}$). If there is one-way causation between $\underline{X}$ and $\underline{y}$, single equation models can be built. If there is two-way causation between $\underline{y}$ and $\underline{X}$, simultaneous-equation models should be employed.

A simultaneous-equation model is a multi-equation model, which consists of a number of separate equations. Both $\underline{y}$ and $\underline{X}$ can appear as dependent variables, although they might appear as explanatory variables in other equations of the model.

In this study, the market model for newsprint contains four equations (supply, export, consumption and price equations), in which X's are exogenous* and y's are endogenous** variables. In their implicit form, they are:

$$y_1 = f(y_2, x_{13}, x_{50}, x_{12}, y_3, x_{24}, x_{26}, x_{36})$$

---

* An exogenous variable is a variable whose value is predetermined.

** An endogenous variable is a variable whose values is determined within the model.

$$y_2 = f(y_4, X_{32}, X_5, y_3, X_{22}, y_1, X_{17}, X_{18}, X_{35}, X_{36})$$

$$y_3 = f(X_5, X_{21}, X_{22}, X_{12}, X_{17}, X_{18}, y_1, y_2, y_4)$$

$$y_4 = f(X_5, X_{21}, y_1, X_{17}, X_{18}, y_2, X_{35}, X_{12}, X_{151}, X_{33}, X_{36})$$

The meaning of the variables X and y are explained in Table 4.1.

The model indicates there is two-way causation between y and X. Therefore, this model is a simultaneous-equation model. There are two groups of econometric techniques to deal with simultaneous-equation models. One is a single-equation method. The other is a system method. The single-equation method is applied to one equation of the system at a time. The system method is applied to the whole system at once. All equations are solved simultaneously.

Since single-equation methods are easy to use and less sensitive to specification error, they are commonly used in economic research. In this study, single-equation methods are applied to estimate coefficients of the market model.

In order to keep a clear distinction between models and methods, a chart (Figure 4.1) is presented.

The methodology for the model developed in this study follows the five stages which are the standard methodology

Table 4.1: Variables used in the analysis

Category	Abbreviation
(1) Endogenous variables	
total shipments of newsprint in Canada	y_1
export of newsprint from Canada to the USA	y_2
total newsprint consumption in the USA	y_3
average price of Canadian newsprint in New York in US dollars	y_4
(2) Exogenous variables	
total shipments of US newsprint	x_5
cost of pulpwood	x_{12}
GNP of Canada	x_{13}
index of US advertising expenditures in newspapers	x_{17}
circulation of US newspapers	x_{18}
US federal reserve discount rate	x_{21}
GNP of the USA	x_{22}
capacity of Canada's newsprint industry	x_{24}
operating rate of Canada's newsprint industry	x_{26}
exchange rate (US/Can)	x_{32}
Bank of Canada prime interest rate	x_{33}
newsprint consumer stock in the USA	x_{35}
technical change variable	X36
price of newsprint in Canada	x_{50}
capital expenditure by Canada's paper and board industry	x_{151}

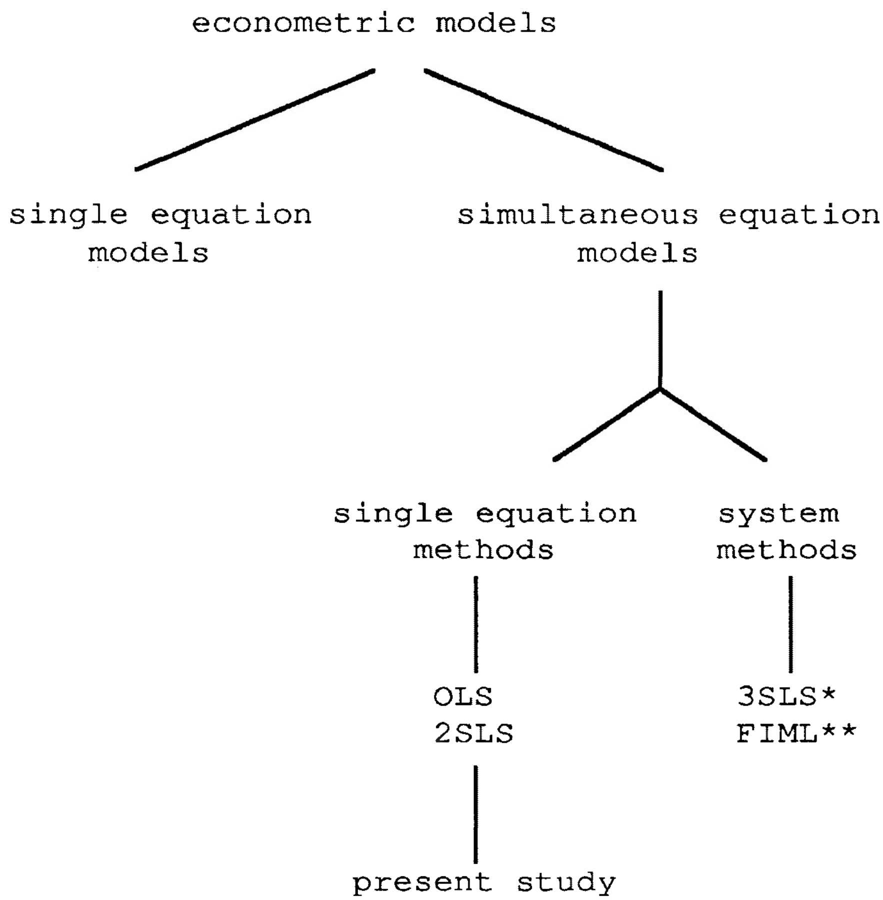

Figure 4.1. Flowchart for model and methods selection.

used in econometric studies and adapted from Koutsoyiannis (1977).

Stage A: the first step is the specification of the model, i.e. formulation of the model. [section 4.1]

Stage B: the second step is to search for data and to collect data. [section 4.2]

---

* Three-stage least squares method (3SLS).
** Full-information maximum likelihood method.

Stage C: the next step is to test the specification error and to test the formulation for identification problems; as well, the choice of econometric technique and the estimates of its parameters should be carried out. [section 4.3]

Stage D: once the model has been estimated, evaluation of the estimates is carried out based on both economic theory and econometric assumptions. [section 4.4]

Stage E: the final step of this study is concerned with the evaluation of the forecasting power of the model. [section 4.5]

4.1 Model specification

Based on an examination of other econometric studies and economic theory, the endogenous variables and exogenous variables were presented in Table 4.1, and a simultaneous-equation market model was proposed (p.22 & 23). The model took the following general mathematical form:

$$\underline{y} = \underline{Y}^*\underline{b} + \underline{X}\,\underline{c} + \underline{u} \qquad [4.1]$$

where:

$\underline{y}$ = the n by 1 vector of observations on the endogenous variable;

$\underline{Y}^*$ = the n by q matrix of endogenous variables

---

Equation [4.1] is obtained from Johnston (1972).

on the right side of an equation;

$\underline{b}$ = the q by 1 vector of structural coefficients attached to the equation to be estimated;

$\underline{X}$ = the n by k matrix of observations on the exogenous variables appearing in the equation;

$\underline{c}$ = the k by 1 vector of coefficients associated with $\underline{X}$;

$\underline{u}$ = the n by 1 vector of disturbances in the equation;

n = number of observations;

q = number of endogenous variables on the right side of an equation;

k = number of exogenous variables plus 1.

From the formulated model the sign of the coefficients to be estimated are given in Table 4.2. These signs are derived from theoretical considerations.

4.2 Data sources

Using the proposed market model, data were obtained for the postulated variables. The present study covers the period between 1955 and 1986. Numerous sources were consulted to obtain the data. Quantities of newsprint export and production were obtained from Reference Tables (1981, 1987, 1989), published by the Canadian Pulp and Paper Association

Table 4.2: The theoretical sign of parameters of explanatory variables in the specified model.

left hand variable	right hand variable	Sign of predictor variables in the model
y_1	y_2	+
	x_{13}	+
	x_{50}	+
	x_{12}	−
	y_3	+
	x_{24}	+
	x_{26}	+
	x_{36}	−
y_2	y_4	−
	x_{32}	+
	x_5	−
	y_3	+
	x_{22}	+
	y_1	+
	x_{17}	+
	x_{18}	+
	x_{35}	−
	x_{36}	−
y_3	x_5	+
	x_{21}	−
	x_{22}	+
	x_{12}	−
	x_{17}	+
	x_{18}	+
	y_1	+
	y_2	+
	y_4	−

Table 4.2: (Continued)

left hand variable	right hand variable	Sign of predictor variables in the model
y_4	x_5	-
	x_{21}	+
	y_1	-
	x_{17}	+
	x_{18}	-
	y_2	-
	x_{35}	-
	x_{12}	+
	x_{151}	-
	x_{33}	+
	x_{36}	+

(CPPA). The GNP of the USA and the federal discount rate were acquired from Hoffman *et al* (1987, 1988) and the U.S Bureau of the Census (1959, 1964, 1969, 1974, 1979, 1984, 1986). The three-month treasury bill market interest rate was collected from the Bank of Canada (1955, 1961, 1963, 1968, 1972, 1976, 1980, 1984, 1988) and Statistics Canada publications (1956, 1957, 1958, 1959, 1960). The reason for using this interest rate is that it is a market interest rate which is sensitive to market changes. Cost of pulpwood were obtained from various Statistics Canada publications (1964, 1966, 1968, 1969, 1986, 1987). The other cost of newsprint production in Canada was also obtained from the CPPA (Newsprint Data). US newsprint consumption, advertising expenditures in US newspaper, circulation of the US newspaper and the price of newsprint in the USA were provided by various CPPA publications. Technical change is explicitly recognized by the dummy variable (X_{36}) which represents this change throughout the period covered by this study. Nineteen variables, as presented in Table 4.1 and further defined in Appendix 1, were used in this study. It is to be noted that X_{35}, newsprint stock in the USA and X_{33}, the bank rate, are lagged by one year and that X_{151}, capital expenditure of Canada's paper and board industry, is lagged by two years.

4.3 Data analysis

The stage of data analysis was divided into 3 parts:

(1) test of the specification error;

(2) examination of the identification conditions of the model;

(3) choice of the appropriate econometric technique for the estimation of the function in order to estimate the model.

These stages are described in the following three sections.

4.3.1 Test of the specification error

The first step of the analysis, i.e. test of the specification error, was completed using the SHAZAM econometric computer package on the Microvax II at Lakehead

University, using VMS version 5.1. The test was divided into two subparts:

i) detection of the presence of unnecessary variables;

ii) test for omitted variables.

After model specification, the exogenous and endogenous variables were selected for this study. It was not certain at this point whether all these variables needed to be in the model. Where some variables were highly correlated with others, these variables were superfluous. In order to determine the need for their presence, t and F tests were used. When the necessity of only one variable at the time was tested, t tests were used. When more than one variable was tested, an F test was carried out. The t test is the standard Student t test of a regression coefficient divided by its standard error to test the H_0: $b_i=0$. The formula which was used to do the t test is as follows:

$$t = b_i/s_{bi} \quad \text{with } df = n-k \qquad [4.2]$$

where:

t = calculated t value;

b_i = the ith estimated regression coefficient;

s_{bi} = standard error of the ith estimated regression coefficient;

n = number of observations;

k = number of coefficients (including intercept term).

The formula used for the F test in this study was as follows.

$$F = [(R_1-R_2)/W]/[(1-R_1)/(n-M)] \qquad [4.3]$$

$$\text{with df} = [W, (n-M)]$$

where:

F = calculated F value;

R_1 = new R-square;

R_2 = old R-square;

W = number of new regressors;

n = number of observations;

M = number of estimated regression coefficients in the new model.

If the t or F value was significant at the 5 per cent level, the null hypothesis that the equation contained unnecessary variable(s) or $b_i=0$ ($b_i=b_j=0$, $i\neq j$) would be rejected and the variables retained in the model. Otherwise the null hypothesis would be accepted.

After the tests, described above, were carried out, it was certain that the variables remaining in the model were required and should be retained. But at this stage, it still was not clear whether all the necessary variables were indeed included in the model. In other words, were there any necessary variable omitted? To answer this question, the test for omitted variables was carried out. In this test, Ramsey's RESET test was used.

Gujarati (1988) introduced Ramsey's method for a general test of the specification error called RESET (regression specification error test). The steps involved in RESET are as follows:

(1) obtain the estimated dependent variable $\hat{Y}_i$ from the chosen model;

(2) rerun the model introducing $\hat{Y}_i$ as an additional regressor;

(3) let the R^2 obtained from step (2) be the new R^2 and that obtained from step (1) the old R^2. Then the F-test to be used is shown in equation [4.3];

(4) if the computed F value were significant at the 5 per cent level, the alternative hypothesis, that the chosen model is misspecified, would be accepted. Then a remedial method should be used in order to solve the problem. Otherwise the null hypothesis would be accepted.

4.3.2 The identification problem

The second part of the data analysis stage involved examination of the identification condition of the model. The identification problem is concerned with the coefficients of an equation that might be obtained from the estimated reduced-form* coefficients. If the coefficients

---

* A reduced-form equation is one in which the endogenous variable is expressed only by exogenous variables and the error term.

can be obtained, the equation is said to be identified. Otherwise it is considered unidentified.

An identified equation can be classified by two categories. One category is that of an exactly identified equation, which means that a unique value can be obtained for each coefficient in the equation. The other category deals with the condition of overidentification in which more than one numerical value can be obtained for each coefficient in the equation.

To solve the identification problem, the so-called order and rank conditions of identification were used in this study. Gujarati (1988) provided two definitions for the order condition of identification. These are:

"Definition 1: In a model of M simultaneous equations, in order for an equation to be identified, it must exclude at least M-1 variables (endogenous as well as predetermined) appearing in the model. If it excludes exactly M-1 variables, the equation is just identified. If it excludes more than M-1 variables, it is overidentified."

"Definition 2: In a model of M simultaneous equations, in order for an equation to be identified, the number of predetermined variables excluded from the equation must not be less than the number of endogenous variables included in that equation less 1; that is: $K - k \geq m - 1$. If $K-k = m-1$, the equation is just identified; but if $K-k > m-1$, it is overidentified.

Here:

m = number of endogenous variables in a given equation;

M = number of endogenous variables in the model;

k = number of predetermined variables in a given equation;

K = number of predetermined variables* in the model."

These parameters apply to all identification statements in section 4.3.2. The order condition may be understood from the following hypothetical example.

Consider the model:

demand $Q = a_0 + a_1 P + a_2 I + u_1$ [4.4]

supply $Q = b_0 + b_1 P + b_2 I + u_2$ [4.5]

This model has M=2 endogenous variables Q (quantity) and P (price) and K=1 predetermined variable I (income). To be identified, each equation must exclude at least one variable, i.e. M-1=1. Since this is not the case, neither equation is identified.

Consider the following change in the model:

demand $Q = a_0 + a_1 P + a_2 I + u_1$ [4.4]

supply $Q = b_0 + b_1 P + u_2$ [4.6]

Applying the order condition ($K-k > m-1$), the demand function is still unidentified because $K-k < m-1$ ($K-k=0$; $m-1=1$); but the supply function is identified because it excludes M-1=1 variable (variable I).

The order condition is a necessary but not sufficient

---

* Predetermined variables contain exogenous and lagged endogenous variable.

condition for identification; this means even if it were satisfied, it may happen that an equation is not identified. Therefore, both a necessary and a sufficient condition for identification are needed. This is provided by the rank condition of identification, which may be stated as follows:

"In a model containing M equations in M endogenous variables, an equation is identified if and only if at least one nonzero determinant of order (M-1) can be constructed from the coefficients of the variables (both endogenous and predetermined) excluded from the particular equation but included in the other equations of the model" (Gujarati, 1988).

The rank condition may be explained as follows. Assuming the following model were set up:

$$y_{1t} - a_{10} - a_{12}\, y_{2t} - b_{11}\, X_{1t} = u_{1t} \qquad [4.7]$$

$$y_{2t} - a_{20} - a_{23}\, y_{3t} - b_{21}\, X_{1t} - b_{22}\, X_{2t} = u_{2t} \qquad [4.8]$$

$$y_{3t} - a_{30} - a_{31}\, y_{1t} - a_{32}\, y_{2t} - b_{32}\, X_{2t} = u_{3t} \qquad [4.9]$$

then, in order to identify the model, the following table should be formed:

equation #	1	y_{1t}	y_{2t}	y_{3t}	X_{1t}	X_{2t}
[4.7]	$-a_{10}$	1	$-a_{12}$	0	$-b_{11}$	0
[4.8]	$-a_{20}$	0	1	$-a_{23}$	$-b_{21}$	$-b_{22}$
[4.9]	$-a_{30}$	$-a_{31}$	$-a_{32}$	1	0	$-b_{32}$

For each equation to be identified, at least one nonzero determinant of order 2 should be formed from the coefficients of the variables excluded from the equation but included in the other equations. For example, to identify equation [4.7], a matrix $\underline{A}$ should be obtained with the following procedure:

(1) strike out the first row from the above table;

(2) strike out the columns in which the coefficients are not zero in the first row;

(3) the remaining coefficients in the table constitute matrix $\underline{A}$.

$$\underline{A} = \begin{bmatrix} -a_{23} & -b_{22} \\ 1 & -b_{32} \end{bmatrix}$$

and its determinant is

$$|\underline{A}| = (a_{23})(b_{32}) + (b_{22}) \neq 0$$

Since the determinant is not zero, the matrix is of full rank [i.e.r($\underline{A}$)=2]. Therefore, equation [4.7] satisfies the rank condition and it is identified.

By using the rank condition method, any model can be detected as an identified or underidentified model. But it is not known whether the model is exactly identified or overidentified. Therefore, Gujarati (1988) suggested to use both order and rank conditions in order to solve the problem. The order and rank conditions of identification are as follows (Gujarati, 1988):

"1) if K-k > m-1 and the rank of the A matrix is M-1, the equation is overidentified;
2) if K-k = m-1 and the rank of the A matrix is M-1, the equation is exactly identified;
3) if K-k > m-1 and the rank of the matrix A is less than M-1, the equation is underidentified;
4) if K-k < m-1,the structural equation is unidentified. the rank of the A matrix in this case is bound to be less than M-1."

Equation [4.7] may be recalled. It is known that this equation is identified since r(A)=2. But it is not clear that the equation is exactly identified or overidentified. Therefore, the order condition method needs to be applied to equation [4.7]. The result obtained from the order condition method is as follows:

K-k (2-1=1) = M-m (3-2=1)

According to the order and rank conditions of identification, this equation is exactly identified.

In this study, the order and rank conditions of identification were used. The results of the identification method indicated the estimated model in this study was overidentified.

4.3.3 Choice of econometric technique

The last part of the data analysis stage involved the choice of econometric technique for the model and then estimating the parameters of the model.

Since the model is a simultaneous-equation model, methods which deal with this kind of model should be selected. These fall into the categories of single-equation methods or that of system methods (see Figure 4.1). Within each category, there are a number of methods which fall into subcategories.

In choosing the appropriate method the following criteria were used:

(1) the method should be easily applied without reducing accuracy or consistency in terms of economic and econometric theory;

(2) the method should be capable of dealing with overidentified equations since the model in this study is an overidentified equation model.

Based on the first criterion, it was decided to choose a single-equation technique for estimation of the model parameters. Single-equation methods are applied individually to each equation in the system of simultaneous equations, without consideration of restrictions in other equations. It is a much easier method to use than any of the system methods. Using the second criterion as a guide it was found that the two stage least squares method (2SLS) was the most suitable method in single-equation methods for this study. The 2SLS method is specially designed to handle overidentified equations although it can be also used to handle equations which are exactly identified. Since it requires less information and can be applied to an individual equation in the system without consideration for other equations, it is the most popular method for estimating a simultaneous-equation model. As Gujarati (1988) said:

"...for solving econometric models involving a large number of equations, 2SLS offers an economic method. It is for this reason that this method has been used extensively in practice."

In the simultaneous-equation model, there is a correlation between the explanatory variables ($\underline{Y}$*) and the disturbance terms ($\underline{u}$). This violates the classical method for estimation of the parameters of a model. By considering overidentified simultaneous-equations, there is not only correlation between $\underline{Y}$* and $\underline{u}$, but there are also more identities than unknown parameters. In this case, the problem is how to estimate unique values for the regression coefficients and how to get rid of the likely correlation between $\underline{Y}$* and $\underline{u}$. The first stage of 2SLS reduces the number of identities to the right number and gets rid of the correlation between $\underline{Y}$* and $\underline{u}$ by the reduced form regression of $\underline{Y}$* on all of the independent variables, i.e: the predetermined variables in the system, not just of that equation. This stage creates new variables ($\hat{Y}$*) which are used in the next stage. The second stage of 2SLS substitutes $\hat{Y}$* into the original equation. Then OLS is applied to the equation. The estimators obtained from stage 2 are consistent. The 2SLS method can be explained as follows.

Model [4.1] is represented as follows:

$$\underline{y} = \underline{Y}^*\underline{b} + \underline{X}\,\underline{c} + \underline{u}$$

(see explanation on page 26 & 27).

Regress each variable in Y* on all independent variables in the model separately. It can denoted as follows:

(1) $\underline{Y}^* = \underline{X}\ \underline{d} + \underline{V}$ $\quad$ ($\underline{V}$ = error term)

The OLS estimator is

$$\hat{\underline{Y}}^* = \underline{X}\ \underline{d}$$

where

$$\underline{d} = (\underline{X}'\underline{X})^{-1}\ \underline{X}'\underline{Y}^*$$

(2) substitute $\hat{\underline{Y}}^*$ for $\underline{Y}^*$ in equation [4.1]

Then the following equation is obtained:

$$\underline{y} = \hat{\underline{Y}}^*\underline{b} + \underline{X}\ \underline{c} + \underline{u^*} \qquad [4.10]$$

(3) now apply OLS to equation [4.10] obtaining

$$[\underline{b},\underline{c}]' = [(\hat{\underline{Y}}^*,\underline{X})'(\hat{\underline{Y}}^*,\underline{X})]^{-1}\ [\hat{\underline{Y}}^*,\underline{X}]'\underline{y} \qquad [4.11]$$

This is the 2SLS solution. This procedure was completed using the SHAZAM package.

4.4 Assumptions of the model

The fourth stage of the methodology is to test the model with the theoretical criteria of economics and econometrics. Economic theory explains why certain variables have a large influence on supply, demand, export and consumption of newsprint production and also indicates the direction of the relationship (positive or negative) between dependent and explanatory variables.

By using the 2SLS method, the following assumptions must be satisfied so that the Best Linear Unbiased Estimators (BLUE) can be obtained:

(1) zero mean value of u_i : $E(u_i) = 0$;

(2) constant variance of u_i : $V(u_i) = \sigma^2$;

(3) zero covariance of u_i: $Cov(u_i,u_j) = 0, i \neq j$;

(4) no perfect multicollinearity among explanatory variables;

(5) the specification of the model is correct;

(6) the number of observations is greater than the number of predetermined variables in the structural system.

Briefly, the 6 assumptions are discussed below. The 2SLS analysis with SHAZAM provided for a residual value analysis. The mean was tested for a zero value using a simple t test at the 5 per cent level of significance.

The assumption of a constant variance was tested by the Goldfeld-Quandt test. This method is applicable if it is assumed that the variance is positively related to one of the explanatory variables in the regression model. For example, assume the following model:

$$\underline{Y}_i = \underline{bX}_i + \underline{u}_i \qquad [4.12]$$

Suppose σ_i^2 is positively related to X_i as

$$\sigma_i^2 = \sigma^2 X_i \qquad [4.13]$$

where σ is a constant.

If equation [4.13] is appropriate, σ_i^2 would be come larger when the values of X_i increase. If this is true, heteroscedasticity is most likely to be present in the model. To carry out this test, the steps suggested by Goldfeld and Quandt are as follows (Gujarati, 1988):

(1) order the n observations according to the values of X from small to large;

(2) eliminate C central observations and divide the remaining (n-C) observations into two groups, of (n-C)/2 observations each. Goldfeld and Quandt suggest that C is 1/4 of the total observations;

(3) run OLS regressions on the two groups and obtain R^2 from each group. R^2 obtained from the small X values is Rss1. R^2 obtained from the large X values is Rss2. Each R^2 has [(n-C)/2]-K degree of freedom (df) (K is the number of parameters to be estimated, including the intercept);

(4) compute the ratio:

$$\lambda = [Rss2/df]/[Rss1/df]$$

which follows the F distribution with both numerator and denominator degree of freedom equal to df.

If the computed λ~F is greater than the F value at the 5 per cent level of significance, the hypothesis of homoscedasticity should be rejected. Otherwise this hypothesis should be accepted. If the hypothesis were rejected, the method of generalized least-squares (GLS) should be applied in order to obtain BLUE. Since this is not the case in this study, OLS could be used.

The 4th and 5th assumptions were already tested in the test of the specification error section (section 4.3.1).

The 6th assumption is satisfied since 30 observations were used to build the model and no equation has more than 7 predetermined variables.

The zero covariance of the disturbance term (u_i) of the original structural equations is tested by the Durbin-Watson test at the 5 per cent level of significance. The test is as follows:

$H_0: \rho = 0$ (the u's have no autocorrelation.)

$H_1: \rho \neq 0$ (the u's are serially dependent.)

Here ρ is the correlation coefficient between the disturbance terms u_i and u_{i-1}.

To carry out this test, a d* value is calculated as:

$$d^* = 2\{1-[(\Sigma e_t e_{t-1})/\Sigma e^2{}_t]\} \qquad [4.14]$$

where:

e_t = estimated error term in period t;

e_{t-1} = estimated error term in period t-1.

The d* value is compared with d_u (the upper value of d* which is obtained from a table of d* values) with (n-L) degrees of freedom (where n is number of the observations; L is the number of explanatory variables). If $d^* > d_u$, accept the null hypothesis of no autocorrelation, otherwise reject the null hypothesis. If H_0 is rejected as was the case in this study, it means that there is an autocorrelation problem in the model. Therefore, the 2SLS estimation is not BLUE. A remedial method should be applied in order to solve the problem.

To solve the autocorrelation problem, the Cochrane-Orcutt method can be used. This method can be understood as follows:

Assume the following model:

$$y_t = b_0 + b_1 X_{1t} + u_t \qquad [4.15]$$

where:

$$u_t = \rho u_{t-1} + v_t \qquad [4.16]$$

(v_t, error term, must satisfy all the assumptions of a random variable)

The relationship for the period t-1 is:

$$y_{t-1} = b_0 + b_1 X_{1t-1} + u_{t-1} \qquad [4.17]$$

Premultiplying this equation by the autocorrelation coefficient (ρ), the following equation is obtained:

$$\rho y_{t-1} = \rho b_0 + \rho b_1 X_{1t-1} + \rho u_{t-1} \qquad [4.18]$$

Subtracting equation [4.18] from equation [4.15], the following generalized difference equation is obtained:

$$y_t - \rho y_{t-1} = b_0(1-\rho) + b_1(X_{1t}-\rho X_{1t-1}) + (u_t-\rho u_{t-1}) \qquad [4.19]$$

Equation [4.19] can be written as follows:

$$y^*_t = b^*_0 + b^*_1 X^*_t + v_t \qquad [4.20]$$

Since $(u_t-\rho u_{t-1})$ equals to v_t, equation [4.19] is BLUE and can be estimated by OLS provided ρ is known. Since this is not the case in this study, ρ should be estimated. Cochrane and Orcutt recommended the following steps to estimate ρ (Gujarati, 1988):

(1) run OLS on equation [4.15] and obtain the residuals;

(2) use the residuals to run the following regression:

$$e_t = \hat{\rho} e_{t-1} + v_t$$

(3) use $\hat{\rho}$ to run equation [4.18].

(4) since it is not known whether $\hat{\rho}$ is the best estimate of ρ, let $b_0{}^* = b_0(1-\hat{\rho})$ and $b_1{}^* = b_1$ (coefficient in equation [4.18]) and run OLS on the following equation:

$$y_t = b_0{}^* + b_1{}^* X_{1t} + u_t$$

From this last equation, the new residuals, e**, can be obtained.

$$e_t^{**} = y_t - b_0^* - b_1^* X_t$$

(5) continue with steps 2, 3 and 4 until the value of two consequent b_1^* values are the same (they converge). In practice 3 or 4 iterations will usually suffice.

The last part of this section deals with the amount of variation explained by the estimated regression model. To test how much variation is explained by the estimated model, the R-square statistic is used. The R-square value is the ratio of explained variation to total variation. In terms of the general model the R-square value is calculated as follows:

$$R^2 = [\underline{A}\underline{X}^{**\prime}\underline{y} - (1/n)(\Sigma Y)^2] / [\underline{y}'\underline{y} - (1/n)(\Sigma Y)^2] \qquad [4.21]$$

where:

$\underline{y} = \underline{Y}^*\underline{b} + \underline{X}\,\underline{c} + \underline{u}$

$\underline{X}^{**} = [\underline{Y}^*, \underline{X}]$

$\underline{A} = [\underline{b}, \underline{c}]$.

If, for example, the R-square value is 0.7, it means that the explanatory variables in a model explain 70 per cent of the total variation in the dependent variable.

4.5 Model validation

The fifth stage of the methodology is to test the forecasting power of the model. The forecasting power of each equation is tested using the following two methods:

(1) test the difference between a predictive and a realized value;

(2) analyse the sources of the forecast error of the model.

The first method is frequently used as a basis for the evaluation of the forecasting power of an econometric model. This can be done by a simple t test. Based on the standard error of the forecast, a t value is computed as follows:

$$t = (Y_a - Y_f)/s_{yf} \qquad [4.22]$$

This t value has Student's t distribution with n-K degrees of freedom.

where:

t = calculated value of the t statistic;

K = number of estimated parameters, including the intercept;

Y_a = actual (observed) value of Y;

Y_f = predicted (forecast) value of Y from the model.

$$s_{yf} = V(u) + V(b_0) + \Sigma[V(b_i)](X_{if} - \overline{X}_i)^2 + 2W$$

where:

$$W = [Cov(b_i b_j)](X_{if} - \overline{X}_i)(X_{jf} - \overline{X}_j)$$

X_{if} = value of ith explanatory variable (X_i) in the forecast period;

X_{jf} = value of jth explanatory variable (X_j) in the forecast period;

$\bar{X}_i$ = mean value of the X_i in the raw data used for estimating the model;

$\bar{X}_j$ = mean value of the X_j in the raw data to be used for estimating the model;

u = disturbance of the estimated model;

b_0 = estimated intercept of the model;

b_i = estimated ith parameter of the model (excluding the intercept).

In this study, data for 1985 and 1986 were used to test the hypothesis H_0: $Y_a = Y_f$ at the 5 per cent level of significance.

To analyze the source of the forecast error, partial inequality coefficients could be calculated. They are:

bias proportion $U_m = (\bar{P}-\bar{A})^2/[\Sigma(P_i-A_i)^2/n]$ [4.23]

variance proportion $U_s = (s_p-s_a)^2/[\Sigma(P_i-A_i)^2/n]$ [4.24]

covariance proportion $U_c = [2(1-r_{pa})s_p s_a]/[\Sigma(P_i-A_i)^2/n]$ [4.25]

where:

s_p = standard deviations of predictors;

s_a = standard deviations of realizations;

r_{pa}= the correlation coefficient of the difference between predicted and realized values;

P_i = predicted (forecast) change in the dependent variable;

$\bar{P}$ = mean of P_i;

A_i = actual (realistic) change in the dependent variable;

$\bar{A}$ = mean of A_i:

n = sample size.

The first component, equation [4.23], shows that the cause of the discrepancy between predictions and realizations is the difference between their means; it is referred to as the bias component of the inequality coefficient. The second component, equation [4.24], shows that another cause of difference between P_i and A_i is the difference between their variances; it is referred to as the variance component of the inequality coefficient. The third component, equation [4.25], shows that still another cause of the discrepancy between P_i and A_i is their imperfect covariance; it is called the covariance component of the inequality coefficient. According to Koutsoyiannis (1977):

"The third source of forecast error is the most dangerous one, in the sense that not much can be done about it. We can never hope that forecasts will be able to produce predictions which would be perfectly correlated with the actual values of the variable. It is natural that $r_{pa} \neq 1$ and hence the 'covariance component' of the prediction error cannot be expected to be zero. The other two sources of error can be reduced in general in the course of time, by the incorporation of additional information in the forecasting process."

4.6 Criteria for choosing the final model

The final stage of the methodology is to evaluate the properties of the econometric model in order to choose the best model. Several criteria, introduced by Koutsoyiannis (1977) and Gujarati (1988), were used in this study. These criteria are as follows:

(1) simplicity: the model should represent the economic relationship with maximum simplicity;

(2) identifiability: this means that for a given set of data the estimated parameters must have unique values;

(3) goodness of fit: the explanation of a model, as measured by R^2, is as high as possible;

(4) theoretical consistency: the model should be compatible with the postulates of economic theory;

(5) predictive power: the model should produce satisfactory predictions of future values of the dependent variables.

5 RESULTS

5.1 Model specification

The market model initially specified the following four equations:

the supply equation:

$$y_1 = c_{10} + b_{12}y_2 + b_{13}y_3 + c_{11}X_{13} + c_{12}X_{50} + c_{13}X_{12} + \\ + c_{14}X_{24} + c_{15}X_{26} + c_{16}X_{36} + u_1 \qquad [5.1]$$

the export equation:

$$y_2 = c_{20} + b_{21}y_1 + b_{23}y_3 + b_{24}y_4 + c_{21}X_{32} + c_{22}X_5 + \\ + c_{23}X_{22} + c_{24}X_{17} + c_{25}X_{18} + c_{26}X_{35} + c_{27}X_{36} + u_2 \qquad [5.2]$$

the consumption equation:

$$y_3 = c_{30} + b_{31}y_1 + b_{34}y_4 + b_{32}y_2 + c_{31}X_5 + c_{32}X_{21} + \\ + c_{33}X_{22} + c_{34}X_{12} + c_{35}X_{17} + c_{36}X_{18} + u_3 \qquad [5.3]$$

the price equation:

$$y_4 = c_{40} + b_{41}y_1 + b_{42}y_2 + c_{41}X_5 + c_{42}X_{21} + c_{43}X_{17} + c_{44}X_{18} + \\ + c_{45}X_{35} + c_{46}X_{12} + c_{47}X_{151} + c_{48}X_{33} + c_{49}X_{36} + u_4 \qquad [5.4]$$

The meaning of $\underline{X}$ and y were explained in Table 4.1.

5.2 Data analysis

The detection of the presence of unnecessary variables was carried out by means of t tests. At the 5 per cent level of significance, unnecessary variables were deleted. Table 5.1 shows the variables in the final model after the elimination of unnecessary variables.

The equations in Table 5.1 were then tested for omitted variables by RESET with the SHAZAM econometrics computer programme. The test results showed that the F values were insignificant at the 5 per cent level in demand, export and supply equations and significant at the 5 per cent level in the price equation. When X_{21} was replaced by $X^2{}_{21}$ in the price equation, the RESET result showed an F value not statistically significant at the 5 per cent level. This meant no necessary variables were omitted from the chosen model. After the two tests, the correct equations of the specified model were obtained and listed in Table 5.2. The complete analyses of each of the equations in Tables 5.1 and 5.2 are given in Appendix II. The results of the tests indicated which explanatory variables have sufficient influence on the dependent variables and should be investigated in this study.

The equations in Table 5.2 were tested for identification by the order and rank condition method. The results of the test showed that all of the four equations in the model were overidentified. This meant there were more equations than unknown coefficients in this study. The methods which deal

Table 5.1: Remaining explanatory variables after detecting the presence of unnecessary variables.

Left hand variable	Right hand variable	Explanation
y_1	y_2	export of newsprint from Canada to the USA
	x_{24}	capacity of Canada's newsprint industry
	x_{26}	operating rate of Canadian newsprint production
y_2	x_5	total shipment of US newsprint
	y_3	total newsprint consumption in the USA
	y_1	total shipment of newsprint in Canada
	x_{17}	index of US advertising expenditures in newspapers
	x_{35}	newsprint consumer stock in the USA
y_3	x_{22}	GNP of the USA
	x_{12}	cost of pulpwood delivered to the mill
	x_{17}	
	x_{18}	circulation of US newspapers
y_4*	x_5	
	x_{21}	US federal reserve discount rate
	y_1	
	x_{17}	
	x_{12}	
	x_{151}	capital expenditure by Canada's newsprint industry
	x_{33}	Bank of Canada prime interest rate

* y_4 is the average newsprint price in the U.S. dollar in the US market.

Table 5.2: Correctly specified model

Equation #	Left hand variable	Right hand variable	Explanation
1	y_1	y_2	export of newsprint from Canada to the USA
		x_{24}	capacity of Canada's newsprint industry
		x_{26}	operating rate of Canadian newsprint production
2	y_2	x_5	total newsprint of US shipment
		y_3	total newsprint consumption in the USA
		y_1	total shipment of newsprint in Canada
		x_{17}	index of US advertising expenditures in newspapers
		x_{35}	newsprint consumer stock in the USA
3	y_3	x_{22}	GNP of the USA
		x_{12}	cost of pulpwood
		x_{17}	
		x_{18}	circulation of US newspapers
4	y_4	x_5	
		y_1	
		x_{17}	
		x_{12}	
		x_{151}	capital expenditure by Canada's paper and board industry
		x_{33}	Bank of Canada prime interest rate
		x^2_{21}	square of the US federal reserve discount rate

with an overidentified model should be used to estimate the model. The results are summarized in Table 5.3 and the complete results presented in Appendix III.

After testing for identification, the equations in Table 5.2 were analyzed by using the 2SLS method from SHAZAM. The results of the analysis are summarized in Table 5.4. The complete results are given in Appendix V. It is shown that equation 4 of Table 5.4 is a good estimation equation for its R-square is high and the Durbin-Watson (D-W) value (d*) is close to 2 (the closer to 2 the better). The Rho value is -0.098 which means the correlation between error terms u_i and u_{i-1} is very low. The result of the D-W test shows that there is no autocorrelation problem in the price equation at the 5 per cent level of significance. But this is not the

Table 5.3: Results of the order and rank conditions of identification

Equation #	No. of pre-determined variables excluded (K-k)	No. of endo-genous vari-ables included less one (m-1)	Rank matrix $\underline{A}$	Identi-fication ?
1	9	1	3	overidentified
2	8	2	3	overidentified
3	7	0	3	overidentified
4	5	1	3	overidentified

Table 5.4: Summary of SHAZAM 2SLS analyses

Eq. #	Explanatory variable	Estimated coefficient	Standard error of coeff.	Standard error of the estimate	R-square for the equation
1	y_2	0.37080	0.1144	103.64	0.9916
	x_{24}	0.74210	0.0588		
	x_{26}	59.31500	7.0330		
	constant	-6086.50000	529.0300		
	Durbin-Watson = 1.32		Rho = 0.32		
2	x_5	-0.57222	0.0660	69.56	0.9883
	y_3	0.20848	0.0470		
	y_1	0.57945	0.0440		
	x_{17}	3.45870	0.5030		
	x_{35}	-0.32777	0.1210		
	constant	995.38000	185.6500		
	Durbin-Watson = 1.07		Rho = 0.44		
3	x_{22}	4.75140	0.4680	144.07	0.9914
	x_{12}	-184.68000	21.0230		
	x_{17}	16.53000	2.4230		
	x_{18}	69.73600	29.4180		
	constant	-3373.10000	2851.7000		
	Durbin-Watson = 1.32		Rho = 0.31		
4	x_5	-0.01621	0.0060	7.12	0.9977
	x^2_{21}	0.44507	0.0790		
	y_1	-0.01750	0.0030		
	x_{17}	1.27490	0.1770		
	x_{12}	6.76460	1.2090		
	x_{151}	-0.04259	0.0100		
	x_{33}	4.11540	0.9720		
	constant	138.32000	14.9420		
	Durbin-Watson = 2.18		Rho = -0.098		

case in the other equations. The D-W values in equations 1, 2 and 3 of Table 5.4 are too far from 2. They are equal to or less than 1.32. They indicate that there is autocorrelation between the error terms at the 5 per cent level of significance. The Rho values are larger than 0.3 which shows the autocorrelation problem is serious in the three equations. Therefore, the 2SLS estimators were not BLUE. A remedial method should be found in order to solve the problem.

5.3 Assumptions of the model

The assumptions of the model were tested by various methods. The test results of the assumptions of 2SLS are summarized in Table 5.5 and the complete test is found in Appendix IV. The results of the tests show that all of the 2SLS assumptions were satisfied for the price equation. Therefore, the 2SLS estimators are BLUE for this equation. This is also shown by Table 5.4. But this is not the case for the other equations. Not all of the assumptions of 2SLS were satisfied for the supply, export and consumption equations. Table 5.5 indicates autocorrelation problems in these three equations. Under these conditions, the 2SLS method breaks down. To solve the autocorrelation problem, the Cochrane-Orcutt method was used in this study. The results of the Cochrane-Orcutt analysis and results of the

Table 5.5: Test for the assumptions of 2SLS

Equation #	$E(u_i)=0$		$V(u_i)=\sigma^2$		$E(u_i,u_j)=0$		perfect multicol-linearity between X's		specifi-cation error	
	Y	N	Y	N	Y	N	Y	N	Y	N*)
1	x		x			x		x		x
2	x		x			x		x		x
3	x		x			x		x		x
4	x		x		x			x		x

*) Y = yes;
N = no.

Durbin-Watson tests are listed in Table 5.6.

The table shows that there is no longer an autocorrelation problem in the three equations after the Cochrane-Orcutt method was applied, since the Rho values appeared sufficiently low. Therefore, the estimators in Table 5.6 are BLUE for the three equations.

The estimated results of the model in Table 5.6 need to be evaluated for economic validity. The criterion used in the evaluation is compatibility with the proposed model structure and the validity of the sign of the coefficient. The evaluated results are listed in Table 5.7. The results

Table 5.6: Summary of the SHAZAM Cochrane-Orcutt analyses

Eq. #	Explanatory variable	Estimated coefficient	Standard error of the coefficient	Standard error of the estimate	R-square for equation
1	y_2	0.33946	0.1317	103.15	0.9926
	x_{24}	0.75824	0.0702		
	x_{26}	62.27100	8.3502		
	constant	-6315.70000	644.1000		
	Durbin-Watson = 1.83		Rho = 0.059		
2	x_5	-0.53957	0.07156	67.60	0.9909
	y_3	0.17286	0.05131		
	y_1	0.61118	0.04758		
	x_{17}	3.40250	0.61651		
	x_{35}	-0.34374	0.11746		
	constant	990.31000	210.39000		
	Durbin-Watson = 1.87		Rho = -0.0017		
3	x_{22}	4.74360	0.49299	135.64	0.9935
	x_{12}	-170.65000	21.86000		
	x_{17}	14.96300	2.46280		
	x_{18}	68.95500	26.80800		
	constant	-3330.10000	2580.60000		
	Durbin-Watson = 1.86		Rho = 0.031		

Table 5.7: Test for economic validity

Eq. #	Predictor variable	Coeff. sign	Is sign of coefficient valid ? Y	N	Does predictor variable fit model structure ? Y	N *)
1	y_2	+	x		x	
	x_{24}	+	x		x	
	x_{26}	+	x		x	
2	x_5	−	x		x	
	y_3	+	x		x	
	y_1	+	x		x	
	x_{17}	+	x		x	
	x_{35}	−	x		x	
3	x_{22}	+	x		x	
	x_{12}	−	x		x	
	x_{17}	+	x		x	
	x_{18}	+	x		x	
4	x_5	−	x		x	
	y_1	−	x		x	
	x_{17}	+	x		x	
	x_{12}	+	x		x	
	x_{151}	−	x		x	
	x_{33}	+	x		x	
	x_{21}	+	x		x	

*) Y = yes;

N = no.

show that all the signs of the explanatory variables were consistent with the sign as specified beforehand in the model. This means the estimated model is consistent with economic theory.

5.4 Model validation

The comparisons of the actual and predicted values for the period of 1955 to 1984 are shown in Figures 5.1 to 5.4. They indicate that the actual and predicted values are very close. Therefore, the estimated model represents the relationship between the dependent and predetermined variables very well for the period.

The data for 1985 and 1986 have been reserved for prediction purposes and have not been used in estimating the model. Table 5.8 indicates the result of testing the significance of the difference between predicted and observed values. The table shows that there are no significant differences between predicted and observed values at the 5 per cent level for the years 1985 and 1986. Table 5.9 shows the results of analyzing the sources of forecasting error about the estimated model. The complete results of Tables 5.8 and 5.9 are found in Appendix VII. Table 5.9 also indicates that the predicted error is largely due to the correlation between the predicted and observed values and accounts for

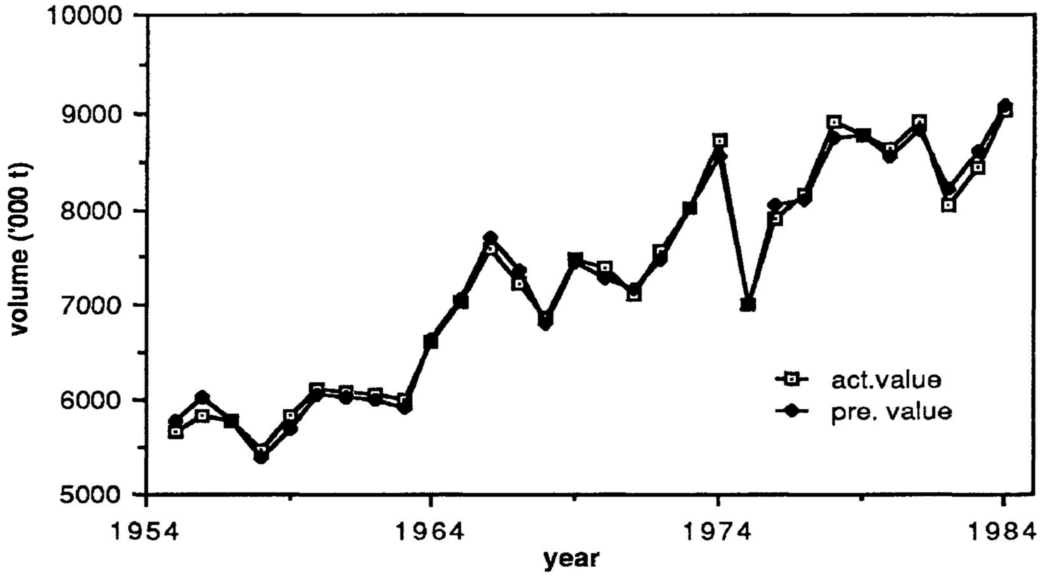

Figure 5.1: Plot of actual and predicted values

(Supply equation)

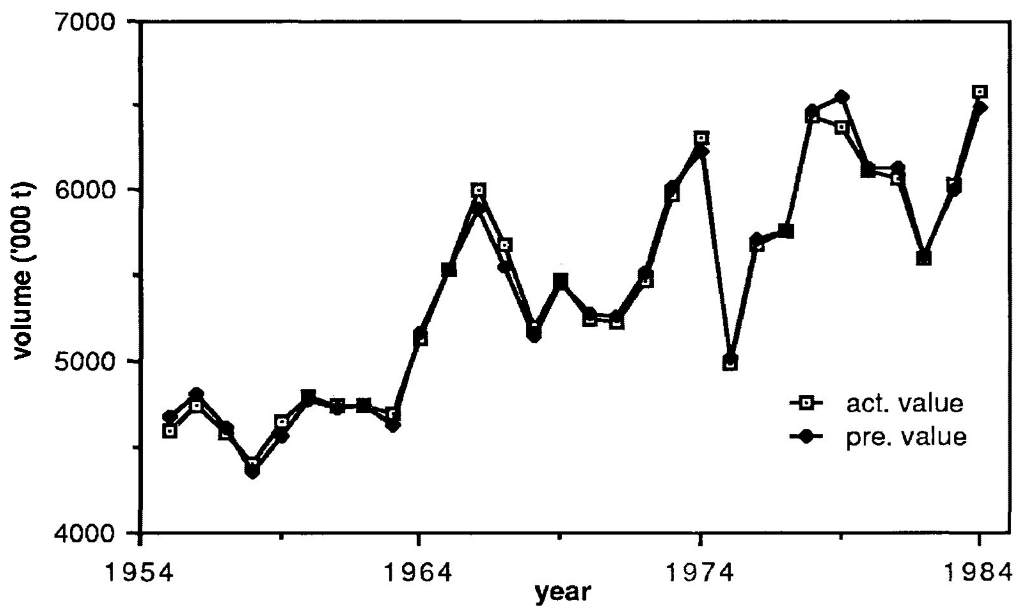

Figure 5.2: Plot of actual and predicted values

(Export equation)

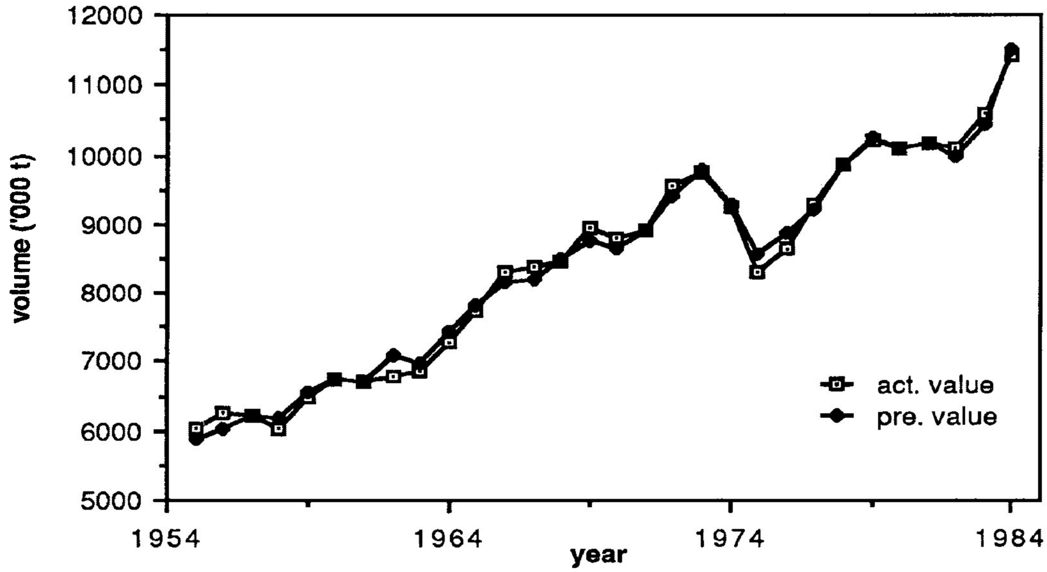

Figure 5.3: Plot of actual and predicted values

(Consumption equation)

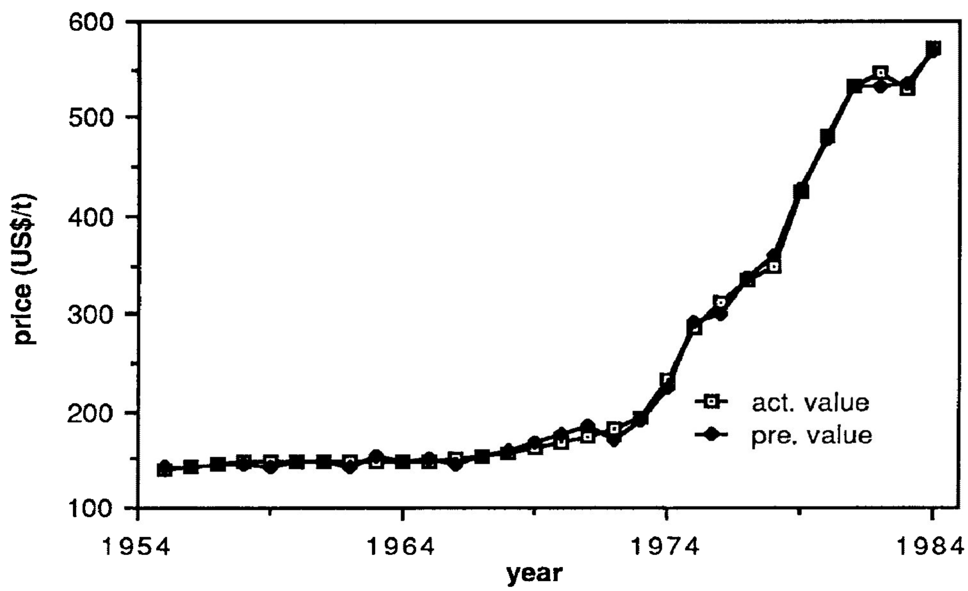

Figure 5.4: Plot of actual and predicted values

(Price equation)

Table 5.8: Test of significance of the difference between prediction and the actual observation at 5 per cent (t-test).

Equation #	Year	Y_a	Y_f	t	df	t(0.05)
1	1985	8996	9094	-0.149	26	2.056
	1986	9302	9319	-0.026	26	2.056
2	1985	6674	6615	0.271	24	2.064
	1986	6734	6827	-0.393	24	2.064
3	1985	11587	12307	-0.278	25	2.060
	1986	11937	13078	-0.440	25	2.060
4	1985	590	635	-1.970	22	2.074
	1986	599	641	-1.789	22	2.074

Y_a: actual (observed) value of Y;
Y_f: predicted (forecast) value of Y from the model.

Table 5.9: The results of analyzing sources of the forecasting error

Equation #	U_m [4.22]	U_s [4.23]	U_c [4.24]
1	0.5263E-03	0.6525E-03	0.9988
2	0.1530E-04	0.3486E-02	0.9965
3	0.4531E-04	0.1625E-01	0.9837
4	0.3058E-29	0.5005E-03	0.9995

more than 98 per cent of the error in each of the four equations. There is nothing that can be done to reduce the covariance proportion (U_c) in order to improve the forecasting of the model, since the correlation coefficient (r_{PA}) cannot equal 1. Therefore, this model is the best estimation model that could be obtained.

5.5 Final model

After all the data manipulation, consisting of the elimination of unnecessary variables, the solution of the identification problem and reduction of autocorrelation, the final model consists of the following equations:

the supply equation:

$$y_1 = \underset{(644.1)}{-6315.7} + \underset{(0.132)}{0.33946}y_2 + \underset{(0.07)}{0.75824}X_{24} + \underset{(8.35)}{62.271}X_{26}$$

$$R^2 = 0.9926 \qquad d = 1.83$$

the export equation:

$$y_2 = \underset{(210.39)}{990.31} - \underset{(0.072)}{0.53957}X_5 + \underset{(0.051)}{0.17286}y_3 + \underset{(0.048)}{0.61118}y_1 +$$

$$+ \underset{(0.617)}{3.4025}X_{17} - \underset{(0.117)}{0.34374}X_{35}$$

$$R^2 = 0.9909 \qquad d = 1.87$$

the consumption equation:

$$y_3 = \underset{(2580.6)}{-3330.1} + \underset{(0.493)}{4.7436}X_{22} - \underset{(21.86)}{170.65}X_{12} + \underset{(2.463)}{14.963}X_{17} +$$

$$+ \underset{(26.808)}{68.955X_{18}} \qquad R^2 = 0.9935 \quad d = 1.86$$

the price equation:

$$y_4 = \underset{(14.942)}{138.32} - \underset{(0.006)}{0.01621X_5} + \underset{(0.079)}{0.44507X^2_{21}} - \underset{(0.003)}{0.0175y_1} +$$

$$+ \underset{(0.177)}{1.2749X_{17}} + \underset{(1.209)}{6.7646X_{12}} - \underset{(0.01)}{0.04259X_{151}} +$$

$$+ \underset{(0.972)}{4.1154X_{33}} \qquad R^2 = 0.9977 \quad d = 2.18$$

The four estimation equations indicate the most important variables which influence Canada's newsprint industry and the US newsprint market. The supply equation shows that the operating rate of the newsprint industry (X_{26}), newsprint exported from Canada to the USA (y_2) and capacity of Canada's newsprint industry (X_{24}) have a large influence on the Canadian newsprint production. The positive coefficients of these explanatory variables support the well known observation of the positive impact of these factors on Canadian newsprint production. The export of newsprint is positively determined by newsprint consumption in the USA (y_3), production of newsprint in Canada (y_1) and the amount of US advertising expenditures in newspapers as indicated by the index (X_{17}) and negatively influenced by shipments of the US newsprint (X_5) and newsprint consumer stock in the USA during the previous year (X_{35}). The consumption equation indicates that the GNP of the USA (X_{22}), the US advertising index (X_{17}) and circulation of US newspapers (X_{18}) are positively correlated with consumption (y_3); the cost of pulpwood (X_{12}) is negatively correlated with consumption (y_3). The newsprint price is positively influenced by the squared US federal reserve discount rate (X^2_{21}), the US advertising index (X_{17}), the cost of pulpwood (X_{12}), the bank rate during the previous year (X_{33}) and negatively determined by shipments of US newsprint (X_5), total shipment of newsprint in Canada (y_1) and the capital expenditures by Canada's paper and board industry two years before (X_{151}). These correlations between dependent and explanatory variables are consistent with economic theory.

6 DISCUSSION

6.1 Model specification

The first, and most important, step in this study was model specification. Sometimes, this is also the weakest and most difficult step in econometrics.

In this step, economic theory serves as a general guide for selecting the variables which were included in this study. Previous econometric studies, such as those by Bulger (1986), Manning (1971), Meyer (1979), McKillop (1969, 1973) and Wallis (1980), provided some detailed information for selecting variables and the mathematical form, such as a linear or nonlinear function. The econometric studies in the newsprint area by Schaefer (1979) and Ghebremichael (1989) provided specific information for selecting variables in this study.

Based on their information and experience, a linear market model was specified for this study, because the linear model is not only the simplest one but also the most popular used mathematical form in econometric research.

6.2 Data sources

In this study, problems related to data collection were as follows:

(1) a lack of consistency in the data; some data series, such as newsprint cost in Canada, are available for most years but not for all;

(2) a lack of available quarterly or monthly data for this study;

(3) the most recent data (1987 and 1988) were not available for some variables;

(4) some data are imperfect, such as newspaper circulation in the USA; this series only contains daily and Sunday paper circulation;

(5) the number of observations was not large enough in order to carry out some tests in this study, e.g. the Durbin-Watson test.

(6) a lack of information about technical change; except the basis weight of newsprint change from 14.5 to 13.6 kg (32 to 30 lbs) in 1975, other technical changes have not been considered by this study.

The lack of consistency in the data was not a serious problem in this study. The capital expenditure of the paper and board industry was substituted for capital expenditure in the newsprint industry in Canada. Newsprint production accounts for 60 per cent of paper and board production. It is reasonable to use these capital expenditures instead of that of the newsprint industry. This introduced only minor inaccuracies in the model.

In this study, repair and construction costs of the newsprint industry should be included according to economic theory. Since the data could not be found for the entire study period, it was excluded from the model. This cost accounts for only a small part of the cost of the newsprint industry and its influence on the newsprint market is negligible. Therefore, it could be excluded without serious violation of the model.

The lack of available quarterly and monthly data is regrettable for these are more sensitive to market changes. If they were available, not only more observations could have been used for the study but also a more sensitive market model of newsprint could have been developed. Quarterly or monthly predictions could be made from a quarterly or monthly model, but now this venue is not available..

The lack of 1987 and 1988 data has not been a big problem for this study. This kind of situation often happens in econometric research. The most recent data is impossible to obtain for each econometric study because it takes time for it to be collected and published. The consequences of this lack of the latest information has been minimal.

The lack of perfect consumption data is a serious problem in other studies. But in the present study, this is not the case, because Sunday and daily newspaper publishers are the major newspaper consumers in the USA. The omitted part is too small to have a serious influence on the US newsprint consumption data. From the major newspapers, the relationship between dependent and explanatory variables could be obtained and a good prediction could be made.

The number of observations was not always large enough to apply the Durbin-Watson test, since the zone of acceptance of the H_0 could not be formed in the test. But this is not a big problem in this case, for ρ is small enough to conclude that no autocorrelation is present. This means that although the autocorrelation coefficient is quite low, the Durbin-Watson value (d value) does not indicate that there is no autocorrelation in the model according to the decision rule of the Durbin-Watson test. The reason is the area of accepting the null hypothesis of no autocorrelation in the model could not be formed. This could not be improved in this study for two reasons. One is that early data have little influence on today's economic performance. Therefore, a data series using the last 30 years seems enough for this study. The other reason is that some data are not available prior to 1955.

6.3 Data analysis

The final test of specification error confirmed that the model was correctly specified. None of the four equations in the model have specification errors according to the t and F tests and RESET. All of the variables which should be contained in the model were included and no superfluous variable was included. All of the equations fitted into the theoretical model structure initially proposed. This confirms the validity of the proposed model.

6.4 Assumptions of the model

Among econometric methods, OLS is the basic and the most important one. Given certain assumptions, OLS has some very attractive statistical properties that have made it one of the most powerful and popular methods of regression analysis.

The assumptions are:

(1) zero mean value of u_i: $E(u_i) = 0$;

(2) no autocorrelation between the u's:

$Cov(u_i, u_j) = 0, i \neq j$;

(3) homoscedasticity or equal variance of u_i:

$V(u_i) = \sigma 2$;

(4) zero covariance between u_i and X_i: $Cov(u_i, X_i) = 0$;

(5) the regression model is correctly specified.

If the model satisfies the above assumptions, OLS will give estimators which are BLUE of the parameters being estimated. This is the famous Gauss-Markov theory.

As mentioned before, simultaneous-equation models cannot satisfy all of the above assumptions. In this model, the fourth assumption is violated. If OLS were applied in this study, the estimation of the model would not only be biased but also inconsistent. Therefore, OLS cannot be applied in this study. Other econometric methods had to be found in order to solve this problem. The 2SLS method can solve the problem and provides for estimators which are BLUE, as mentioned before.

As mentioned in the methodology part (p.42), certain assumptions for using the 2SLS method need to be met. Those assumptions are often violated when using regression techniques on raw data.

The assumption of zero mean of the error term was tested by Student's t test. The test result indicated this assumption is satisfied at the 5 per cent level of significance for all of the equations in this study.

If the assumption of constant variance is violated then the problem known as heteroscedasticity is encountered (Gujarati, 1988). The solution here is to use GLS. But in this study, the assumption of constant variance was satisfied. Therefore, it was not necessary to carry out a GLS analysis.

The other assumptions of the 2SLS method, except zero covariance between disturbance (u_i), were satisfied by this model. The test results were shown in the last section.

When the assumption of zero covariance between disturbances is violated, the 2SLS method breaks down. This means the 2SLS estimate is not BLUE. This happened in this study. The results of the Durbin-Watson test show that there were serious autocorrelation problems in equations 1, 2 and 3. Therefore, the Cochrane-Orcutt method was used to solve the problem in this study. As the results of the method indicated, the autocorrelation problem in equations 1, 2 and 3 was solved at the 5 per cent level of significance. The estimators obtained from this study are BLUE.

6.5 Model validation

The results of comparing the actual and predicted values for the period of 1955-1984 inclusive are very close; their R-square values were greater than 0.99 for each equation in this model. For the data of 1985 and 1986, which were not included in the model, tests of the predicted values with actual values showed that there was no significant difference at the 5 per cent level. In order to analyze the source of predicted error, the error was divided into three parts: U_m, the bias proportion, whose error comes from the mean of predicted and observed values; U_s, the variance proportion for which the error comes from the variance of predicted and observed values and finally U_c, the covariance proportion of the predicted error, which is caused by the covariance of predicted and observed values. The U_m and U_s could be reduced by using additional information in the forecasting process. But U_c could not be reduced during the process because the correlation coefficient of predicted and observed values can never be made to equal 1. Therefore, when U_c accounts for most of the error, there is little that can be done to reduce the predicted error for an estimated model. As Table 5.9 shows, the analysis of the sources of the forecasting error indicates that the major error comes from U_c which could not be reduced. So this model is the best estimated model for this study.

6.6 Final model

The final model had very high R-square values and predictive power. The signs of the coefficients of the model were consistent with economic theory.

A comparison of this market model with previous newsprint market models is not easy. In this study, simultaneous-equation models were developed. In previous studies, most of the models were of the single equation type. For example Johnson (1985) provided a newsprint demand equation for the USA. Ghebremichael (1989) published an econometric study report for the Ontario newsprint industry, in which an equation of the USA demand for Ontario newsprint was given. Roberts and Luck (1985) set up a domestic consumption model for Canadian customers. All of these studies provide a single equation so that it is difficult to compare them with this study.

Although model comparison is difficult, some consistent points emerge. In the first place, the GNP or GDP of the USA is an important determinant of the US demand for newsprint. This fact is clearly expressed in this study and others such as those by Ghebremichael (1989), Johnson (1985) and

Schaefer (1979). Secondly, the circulation of newspapers and advertising expenditures are two important determinants of the consumption of newsprint in this model. This result supports the 1979 study by Schaefer and is consistent with the results of a recent FAO (1986) forecast for the world's demand for paper products. Finally, operating rates have a large influence on the total shipment of newsprint as exposed by this study and that of Schaefer.

It should be pointed out that the technical change variable was excluded from the final model. This seems incorrect because the basis weight was changed from 14.5 kg to 13.6 kg in 1975. This change was expected to have some influence on the shipment of newsprint. But when a test of the influence of the technical change (X_{36}) was carried out, it indicated no significant influence on price, export and supply at the 5 per cent level. Therefore, it was eliminated from the final model. Technical change affects newsprint consumption primarily through the price equation. Therefore, it was not introduced into consumption equation directly

There remains the moot point about the exchange rate. Theoretical considerations suggest that the exchange rate (Can$/US$) has a considerable bearing on Canadian newsprint export to the USA. Ghebremichael (1989) pointed out that if the value of the Canadian dollar increases in relation to the US dollar, Canadian exports will decrease. This is true in the period covered by this study. When the simple regression of exports (y_2) on the exchange rate (US$/Can$) (X_{32}) was carried out, the influence of the exchange rate on the export could be found. The correlation coefficient of X_{32} and export (y_2) is 0.65. Therefore, it was to be expected that the exchange rate would be included in this model. Unfortunately, its t-statistic is quite low in the export equation. In absolute terms, it is only 0.89 in this study, 0.47 in the study by Ghebremichael (1989) and 0.5 in the Schaefer (1979) study. Since the t value was low, the exchange rate was eliminated from the export equation at the 5 per cent level of significance without invalidating the equation. When a lagged exchange rate (3 - 5 years) was used in the export equation, the result was not significantly different from that of the current exchange rate at the 5 per cent level. This result is consistent with some other studies. Bulger (1986) eliminated the exchange rate from his export model. Manning (1975) excluded the exchange rate in his model altogether when he studied the Canadian softwood lumber industry. But some researchers such as Schaefer (1979) and Ghebremichael (1989), insisted that it remain in their model although its t value was small. In this study, the elimination introduced only minor inaccuracies in the model. This does not mean that the exchange rate has no influence on Canadian newsprint export to the USA. For an export oriented industry, the exchange rate is of great importance to individual firms and the entire newsprint industry. From a simple regression of the volume of export on the exchange rate, it was estimated that a 1 cent change in the value of the Canadian dollar in term of the US dollar, would be associated with an opposite change of about 37 thousand tonnes of newsprint exported to the USA. At the current (October 1989) average price of US$598/t, this would amount to a value of $22 million and represents about half of one per cent of the total annual value of newsprint exports to the USA.

6.6 Model limitations

A number of model limitations should be pointed out. In this study they are as follows:

(1) the model cannot be used to predict seasonal or monthly fluctuations since it is an annual model;

(2) sudden or large changes in dependent variables are less accurately predicted; therefore the model is not suitable to use under these situations;

(3) the model cannot be used to predict regional newsprint supply and demand, since it is a national model.

7 RECOMMENDATIONS

7.1 Summary of the study

The final model presents the relationship of the US market and the Canadian newsprint industry as it existed in the period 1955-1986. It shows that the variables remaining in the final model are the most important in the relationship.

The final model also provides an accurate means of predicting supply and demand for Canadian newsprint as well as forecasting consumption and price for newsprint in the USA for the years 1985 and 1986.

7.2 Recommendations

The recommendations for the use of the estimated model are as follows:

(1) The model should be updated to the most recent year (1988 or 1989) for which predictor variables are available;

(2) The model structure should be reviewed periodically

in order to consider any economic change (such as in the bank rate, GNP and price) so that the most accurate prediction can be obtained from the model;

(3) The model should be used with caution during periods of wide economic fluctuations;

(4) Since the model is a national newsprint market model, it is not suitable for predicting provincial newsprint development.

The following recommendations for future studies in this area are made.

(1) Quarterly or monthly data should be used; not only can seasonal variation be predicted but it also increases the number of observations so that more efficient statistical tests, such as the Durbin-Watson test, can be carried out.

(2) Owing to time constraints during this study, data sources such as the circulation of newspapers in the USA were incomplete; for future studies, this problem should be solved in order to obtain a better estimation model.

(3) Factors which influence newspaper circulation such as people's education level, should be considered; it is reported that when the level of education is increased by 1 per cent, the newspaper circulation will be increased by 1 per cent, as well (FAO, 1986).

(4) As mentioned before, expenditures on advertisement are a major source of newspaper income. Therefore, future studies should investigate this fact and analyze its influence on the newsprint consumption of the USA in more detail.

(5) Recycling of used paper should be considered in future studies. Owing to insufficient statistical information of recycled newsprint material, this consideration is excluded from this study.

8 LITERATURE CITED

Bank of Canada, 1955. Statistical Summary , 1955. 336 pp.

Bank of Canada, 1961. Statistical Summary Supplement, 1961. 161 pp.

Bank of Canada, 1963. Statistical Summary Supplement, 1963. 165 pp.

Bank of Canada, 1968. Statistical Summary Supplement, 1968. 177 pp.

Bank of Canada, 1972. Bank of Canada Review, 1972.

Bank of Canada, 1976. Bank of Canada Review, 1976.

Bank of Canada, 1980. Bank of Canada Review, 1980.

Bank of Canada, 1984. Bank of Canada Review, 1984.

Bank of Canada, 1988. Bank of Canada Review, 1988.

Brown, M. and De Cani, J. S. 1963. Technological change and the distribution of income. International Economic Review, 4,289-309.

Bulger,W. D.1986. An econometric prediction model for British Columbia interior softwood lumber exports to the United States. Bachelor thesis. Lakehead University. 121 pp.

Buongiorno, J. and J. K. Gilless. 1987. Forest Management and Economics. Macmillan Publishing Company. New York. 285 pp.

Choudhry, N.K., Y. Kotowitz, J.A. Sawyer and J.W. L. Winder. 1972. The Trace Econometric Model of the Canadian Economy. University of Toronto Press. Toronto and Buffalo. 166 pp.

CPPA,1956. Reference Table, 1956. Canadian Pulp and Paper Association. 32 pp.

CPPA,1957. Reference Table, 1957. Canadian Pulp and Paper Association. 32 pp.

CPPA,1963. Reference Table, 1963. Canadian Pulp and Paper Association. 32 pp.

CPPA,1964. Reference Table, 1964. Canadian Pulp and Paper Association. 32 pp.

CPPA, 1978. Reference Table, 1978. Canadian Pulp and Paper Association. 40 pp.

CPPA, 1981. Reference Table, 1981. Canadian Pulp and Paper Association. 40 pp.

CPPA, 1981. Annual Newsprint Supplement, 1981. Canadian Pulp and Paper Association. 8 pp.

CPPA, 1983. Annual Newsprint Supplement, 1983. Canadian Pulp and Paper Association. 8 pp.

CPPA, 1985a. Reference Table, 1985. Canadian Pulp and Paper Association. 42 pp.

CPPA, 1985b. Newsprint Data, 1985. Canadian Pulp and Paper Association. 27 pp.

CPPA, 1986. Annual Newsprint Supplement, 1986. Canadian Pulp and Paper Association. 8 pp.

CPPA, 1987. Reference Table,1987. Canadian Pulp and Paper Association. 73 pp.

CPPA, 1988. Annual Report for the Year 1988. Canadian Pulp and Paper Association. 24 pp.

CPPA, 1989. Reference Table, 1989. Canadian Pulp and Paper Association. 63 pp.

Davidson, J. E. H., Hendry, D.F., Srba, F. and Yeo, S. 1978. Econometric modelling of aggregate time-series relationship between consumers' expenditure and income in the United Kingdom, Eco. J., Vol.88, 661-692

Davis, J., A.L. Best, J.M. Smith, D.A. Wilson, P.E. Lachance and S.L. Pringle. 1957. The Outlook for the Canadian Forest Industry. Queen's Printer and Controller of Stationery Hull. 261 pp.

Davis, T. E.1952. The consumption function as a tool for prediction. Review of Economics and Statistics, 34,207-277.

FAO, 1986. The world pulp and paper industry outlook: 1995. FAO. Roma. 109 pp

Ghebremichael, A. 1989. An Analysis of the Demand for

Ontario's Newsprint and Market Pulp in the United States. Can. For. Serv., Great Lakes For. Cent., Info. Rep. O-X-397. 21 pp.

Gujarati, D.N. 1988. Basic Econometrics. McGraw Hill Book Company. Toronto. 705 pp.

Hair,D. 1967. Use of Regression Equations for Projecting Trends in Demand for Paper and Board. U.S. For. Serv., Washington, D.C., For. Res. Rep. No.18. 178 pp.

Hoffman, M.S., J. Foley, T. McGuire, D. Young, D.G. Felder, A. Cummings and L. Ratzkin, 1987. The World Almanac and Book of Facts. D. Hendin and P. Henrici, New York. 928 pp.

Hoffman, M.S., J. Foley, T. McGuire, D. Young, D.G. Felder, A. Cummings and L. Ratzkin, 1988. The world Almanac and Book of Facts. D. Hendin and P. Henrici, New York. 928 pp.

Jacques, R. 1988. Impact of Forestry Activity on the Economy of Canada and Its Provinces: An Input-Output Approach. CFS Info. Rep. E-X-39, 34 pp.

Johnson, J. D. 1985. Medium-term Market Outlook for the Ontario Newsprint Industry. Can. For. Serv., Great Lakes For. Cent., Rep. No. 26. 23 pp.

Johnston, J. 1972. Econometric Methods. McGraw Hill Book Company, New York. 437 pp

Koutsoyiannis, A. 1977. Theory of Econometrics. The Macmillan Press Ltd., London. 681 pp.

Manning, G.H. and H.R. Grinnell. 1971. Forest Resources and Utilization in Canada to the Year 2000. Environment Canada, C.F.S., 80pp.

Manning, G.H. 1975. The Canadian softwood lumber industry: a model. Can. J. For. Res., 5(3): 345-351.

McKillop, W.L.M. 1969. An econometric model of the market for redwood lumber. For. Sci., 15, 159-170.

McKillop, W.L.M. 1973. Structural analysis of Japanese-North American trade in forest products. For. Sci., 19: 63-74.

Meyer,P., 1979. Empirische Untersuchungen Ueber den Schweizerischen Holzmarkt. Ph. D thesis, University of Zuerich. 302 pp.

Mills, T.J. and R.S. Manthy. 1974. An Econometric Analysis of Market Factors Determining Supply and Demand for

Softwood Lumber. Michigan State University, Agricultural Exp. Stn. Res. Rep. 238, 59 pp.

Neal, F. and R. Shone. 1976. Economic Model Building. The Macmillan Press Ltd. London and Basingstoke. 172 pp

Roberts, D. and D. Luck. 1985. Canada's Domestic Consumption of Forest Products: 1960-2000. Economics Branch. G.F.S.Info. Rep. E-X-35. 74 pp.

Schaefer, G.P. 1979. The Canadian Newsprint Industry: Econometric Models of Different Market Structures. Bank of Canada, Ottawa, Ont. Tech. Rep. 17. 77 pp.

Statistics Canada, 1956. Bank of Canada, Annual Report of the Governor to the Minister of Finance, FN73 A55. 79 pp.

Statistics Canada, 1957. Bank of Canada, Annual Report of the Governor to the Minister of Finance, FN73 A55. 45 pp.

Statistics Canada, 1958. Bank of Canada, Annual Report of the Governor to the Minister of Finance, FN73 A55. 63 pp.

Statistics Canada, 1959. Bank of Canada, Annual Report of the Governor to the Minister of Finance, FN73 A55. 82 pp.

Statistics Canada, 1960. Bank of Canada, Annual Report of the Governor to the Minister of Finance, FN73 A55. 103 pp.

Statistics Canada, 1964. Logging, 25-201. 12 pp.

Statistics Canada, 1966. Logging, 25-201. 19 pp.

Statistics Canada, 1968. Logging, 25-201. 14 pp.

Statistics Canada, 1969. Logging, 25-201. 14 pp.

Statistics Canada, 1986. Selected Forestry Statistics, FR175 84R02. 180 pp.

Statistics Canada, 1987. Selected Forestry Statistics, FR175 84R02. 188 pp.

Suits, D. B. 1955. An econometric model of the watermelon market. J. of Farm Econ., 37, 237-251.

U.S. Bureau of the Census, 1959. Statistical Abstract of the United States, 1959. Washington, D.C., 1042 pp.

U.S. Bureau of the Census, 1964. Statistical Abstract of the United States, 1964. Washington, D.C., 1041 pp.

U.S. Bureau of the Census, 1969. Statistical Abstract of the United States, 1969. Washington, D.C., 1032 pp.

U.S. Bureau of the Census, 1974. Statistical Abstract of the United States, 1974. Washington, D.C., 1028 pp.

U.S. Bureau of the Census, 1979. Statistical Abstract of the United States, 1979. Washington, D.C., 1057 pp.

U.S. Bureau of the Census, 1984. Statistical Abstract of the United States, 1984. Washington, D.C., 1015 pp.

U.S. Bureau of the Census, 1986. Statistical Abstract of the United States, 1986. Washington, D.C., 985 pp.

Wallis, K.F. 1980. Topics in Applied Econometrics. University of Minnesota Press. Minneapolis, 166 pp.

APPENDIX I : DATA

Data

Year	y_1	y_2	y_3	y_4	x_5	x_{13}	x_{24}
1955	5657	4599	6022	138.89	1406	27.1	5501
1956	5850	4545	6259	143.30	1556	30.6	5667
1957	5773	4586	6228	146.61	1648	31.9	6129
1958	5482	4379	6028	147.71	1598	32.9	6567
1959	5829	4643	6488	147.71	1781	34.9	6823
1960	6125	4789	6737	147.71	1843	36.3	6905
1961	6085	4742	6695	147.71	1893	37.5	7016
1962	6060	4744	6792	147.71	1961	40.6	7116
1963	6007	4699	6847	147.71	2003	43.4	7307
1964	6632	5124	7295	147.71	2062	47.4	7506
1965	7028	5528	7757	147.71	2039	52.2	7639
1966	7607	5997	8304	149.91	2250	58.1	8054
1967	7229	5682	8381	154.87	2441	65.7	8431
1968	6885	5194	8461	156.53	2772	71.4	8212
1969	7484	5470	8940	162.04	3051	79.7	8342
1970	7403	5242	8805	167.55	3136	85.7	8196
1971	7112	5227	8922	174.17	3146	94.5	8519
1972	7570	5469	9551	181.42	3312	105.2	8501
1973	8012	5970	9741	192.90	3182	123.6	8735
1974	8706	6304	9262	231.49	3229	147.5	8899
1975	7010	4980	8325	286.60	3347	165.3	8974
1976	7903	5671	8670	311.40	3381	191.9	9000
1977	8169	5751	9281	333.45	3507	210.2	8983
1978	8913	6443	9864	348.33	3429	232.2	8946
1979	8779	6371	10199	425.27	3689	264.3	9063
1980	8622	6118	10088	480.61	4234	297.6	9130
1981	8912	6058	10164	531.87	4735	339.8	9490
1982	8055	5596	10108	545.46	4525	358.3	9935
1983	8439	6034	10589	528.37	4674	393.7	9970
1984	9019	6586	11431	570.45	5065	429.5	9820
1985	8996	6674	11587	589.74	4927	465.1	9820
1986	9302	6734	11937	599.38	5115	506.5	9885

Source: CPPA, Reference Tables 1981, 1987, 1989.

Data (continued)

Year	x_{26}	x_{32}	x_{35}	x_{50}	x_{151}	x_{21}*	x_{22}*
1955	102.1	0.99	655	121.67	91.9	2.125	658.6
1956	103.6	0.98	610	125.54	69.0	2.875	672.1
1957	94.7	0.96	814	125.52	118.3	3.250	685.0
1958	84.2	0.97	853	126.61	239.6	2.250	672.6
1959	85.0	0.96	823	125.72	234.4	3.500	718.2
1960	88.5	0.97	823	129.55	106.8	3.250	733.4
1961	87.1	1.01	781	132.32	110.2	3.000	747.8
1962	85.3	1.07	750	135.17	141.3	3.000	797.3
1963	82.3	1.08	763	134.71	138.4	3.500	827.0
1964	88.2	1.08	691	133.84	147.8	4.000	868.6
1965	91.7	1.08	724	132.01	181.6	4.500	929.0
1966	94.8	1.08	707	134.74	293.7	4.500	988.1
1967	86.6	1.08	826	138.06	383.8	4.250	1015.3
1968	83.0	1.08	753	147.54	506.4	5.313	1063.4
1969	89.9	1.08	760	148.94	418.5	6.000	1091.1
1970	90.2	1.04	865	149.49	240.2	5.625	1086.0
1971	84.4	1.01	944	152.31	325.7	4.875	1122.2
1972	88.2	0.99	885	155.51	488.3	4.500	1191.7
1973	91.0	1.00	667	167.17	491.0	6.180	1261.9
1974	96.3	0.98	753	213.06	389.0	7.875	1246.0
1975	77.6	1.02	1052	263.52	341.9	7.000	1232.0
1976	89.9	0.99	903	275.84	468.7	5.620	1298.0
1977	90.8	1.06	1206	314.63	486.9	5.620	1370.0
1978	98.1	1.14	1070	337.02	604.8	7.750	1439.0
1979	96.6	1.17	970	380.02	690.2	10.750	1479.0
1980	94.5	1.17	877	431.78	607.3	11.500	1475.0
1981	94.3	1.20	1090	492.14	763.1	13.000	1512.0
1982	81.7	1.23	1477	503.72	1233.0	9.250	1480.0
1983	85.1	1.23	1258	492.44	2051.6	8.500	1535.0
1984	91.8	1.29	1179	534.56	1743.2	8.500	1639.0
1985	91.5	1.37	1329	551.17	903.1	7.750	1658.1
1986	94.0	1.39	1394	559.60	1014.3	6.250	1717.4

Source: CPPA, Reference Tables (1981, 1987, 1989).

* U.S. Bureau of the Census, Statistical Abstract of the United States (1959, 1964, 1969, 1974, 1979, 1984, 1986); Hoffman, *et al* (1987 & 1988) The World Almanac and Book of Facts.

Data (continued)

Year	x_{12}*	x_{17}**	x_{18}**	x_{33}#	x_{36}
1955	9.61	44.4	102.6	1.44	0
1956	10.06	46.4	104.3	1.06	0
1957	9.54	44.7	104.8	2.56	0
1958	9.16	45.8	104.4	3.67	0
1959	9.46	51.0	106.1	2.26	0
1960	8.93	53.0	106.6	4.67	0
1961	9.93	52.0	107.5	3.36	0
1962	9.76	53.0	108.8	2.83	0
1963	10.13	54.0	105.7	3.98	0
1964	10.09	59.0	108.8	3.57	0
1965	10.19	64.0	109.0	3.75	0
1966	11.05	70.0	110.7	3.92	0
1967	11.78	71.0	110.8	4.96	0
1968	11.89	75.0	111.5	4.56	0
1969	11.84	82.0	111.7	6.24	0
1970	12.13	82.0	111.3	7.06	0
1971	12.49	89.0	111.9	6.27	0
1972	12.76	100.0	112.5	3.66	0
1973	14.01	108.0	114.8	3.52	0
1974	16.56	114.0	113.6	5.25	0
1975	20.26	121.0	111.8	7.76	1
1976	22.47	141.0	112.6	7.26	1
1977	24.48	159.0	113.9	8.91	1
1978	25.30	181.0	116.0	7.41	1
1979	26.87	207.0	116.6	8.40	1
1980	28.99	222.0	116.9	11.42	1
1981	31.86	249.0	116.6	12.51	1
1982	33.42	253.0	118.8	17.94	1
1983	36.04	294.0	119.3	14.04	1
1984	36.68	335.0	120.6	9.32	1
1985	35.55	364.0	121.7	11.05	1
1986	35.93	390.0	124.1	9.48	1

Source: * Statistics Canada,25-201 (1964, 1966, 1968, 1969); and Selected Forestry Statistics (1986, 1987).

** CPPA, Reference Tables 1964; Annual Newsprint Supplement (1981, 1983, 1986).

Statistics Canada, Bank of Canada, Annual Report of the Governor to the Minister of Finance (1956-1960); Bank of Canada Statistical Summary (1955); Bank of Canada Statistical Supplement (1961, 1963, 1968). Bank of Canada Review (1972, 1976, 1980, 1984, 1988);

y_1 = total shipments of newsprint in Canada ('000 t)

y_2 = newsprint exported from Canada to the USA ('000 t)

y_3 = total newsprint consumption in the USA ('000 t)

y_4 = average newsprint price in US dollar ($/t) in the US market (delivered at New York).

x_5 = total shipment of US newsprint ('000 t)

x_{12} = cost of pulpwood delivered to the mill (Can$/m^3).

x_{13} = GNP of Canada in current dollars ($10^9).

x_{17} = Index of US advertising expenditures in newspapers (1972=100)

x_{18} = circulation of newspapers (daily + Sunday) in million copies.

x_{21} = US federal reserve discount rate (average, per cent per year).

x_{22} = GNP of the USA in 1977 constant dollar ($10^9).

x_{24} = capacity of Canada's newsprint industry ('000 t).

x_{26} = operating rate of newsprint production in Canada (%)

x_{32} = exchange rate of US dollar in terms of the Canadian currency.

x_{33} = Bank of Canada prime interest rate on three month treasury bills (average and lagged one year).

x_{35} = newsprint consumer stock in the USA (lagged one year).

X36 = technical change variable.

x_{50} = newsprint price in Canada (Can$/t).

x_{151} = capital expenditure of Canada's paper and board industry, lagged two years ($10^6).

APPENDIX II : MODEL SPECIFICATION TEST RESULTS

Detecting the presence of unnecessary variables

Assumption : The equation contains unnecessary variable(s)

Test: t test when one variable was tested

F test when more than one variable needed to be tested

H_0: Equation contains unnecessary variable(s)

H_1: Equation does not contain unnecessary variable(s)

accept H_0 if $t < t_{0.05}$ (df) or $F < F_{0.05}$ (df1,df2), otherwise reject H_0 and accept H_1.

Equation #	Right hand variable	t value	F value	Result accept H_0	reject H_0
1			37901		x
	y_2	3.53			x
	x_{24}	12.83			x
	x_{26}	8.48			x
2			31331		x
	x_5	-9.10			x
	y_3	5.28			x
	y_1	13.53			x
	x_{17}	7.15			x
	x_{35}	-2.81			x
3			21107		x
	x_{22}	10.15			x
	x_{12}	-8.78			x
	x_{17}	6.82			x
	x_{18}	2.37			x

Equation #	Right hand variable	t value	F value	Result accept H_0	reject H_0
4			6229		x
	x_5	-2.90			x
	y_1	-6.01			x
	x_{17}	7.22			x
	x_{12}	5.61			x
	x_{151}	-4.14			x
	x_{33}	4.20			x
	x_{21}	5.62			x

Nb: For equation 1: $t_{0.05} = 2.056$; $F_{0.05} = 2.98$;
For equation 2: $t_{0.05} = 2.064$; $F_{0.05} = 2.62$;
For equation 3: $t_{0.05} = 2.060$; $F_{0.05} = 2.59$;
For equation 4: $t_{0.05} = 2.074$; $F_{0.05} = 2.46$.

Test for omitted variables and incorrect functional form

Assumption : Model is correct

Test : RESET (Regression Specification Error Test)

H_0 : Model is correct

H_1 : Model is misspecified

If $F^* < F_{0.05}$ (W, n-M), accept H_0; otherwise reject H_0, accept H_1.

where :

$$F^* = [(R_1-R_2)/W]/[1-R_1)/(n-M)] \quad [4.3]$$

where :

W = number of new regressors

R_1 = new R-square

R_2 = old R-square

n = observations

M = number of parameters in the new model

Equation #	F* value	df1	df2	$F_{0.05}$	Result accept H_0	Result reject H_0
1	1.34	1	25	4.24	x	
2	2.34	1	23	4.28	x	
3	0.06	1	24	4.26	x	
4	2.37	1	21	4.32	x	

APPENDIX III : IDENTIFICATION RESULTS

Model Identification

Method : Order and Rank conditions of Identification

If $K-k > m-1$ and $R(\underline{A}) = M-1$, the equation is overidentified;

If $K-k = m-1$ and $R(\underline{A}) = M-1$, the equation is exactly identified;

If $K-k > m-1$ and $R(\underline{A}) < M-1$, the equation is unidentified;

If $K-k < m-1$, the equation is unidentified.

where $R(\underline{A})$ is the rank of the matrix $\underline{A}$.

Coefficients in the specification model

Equation #	Variables in the model							
	y_1	y_2	y_3	y_4	1	x_{24}	x_{26}	x_5	x_{17}
1	1	a_3	0	0	a_0	a_1	a_2	0	0
2	b_3	1	b_2	0	b_0	0	0	b_1	b_4
3	0	0	1	0	c_0	0	0	0	c_3
4	d_3	0	0	1	d_0	0	0	d_1	d_4

Coefficients in the specified model (continued)

Equation	Variables in the model						
#	x_{35}	$x^2{}_{22}$	x_{12}	x_{18}	x_{21}	x_{151}	x_{33}
1	0	0	0	0	0	0	0
2	b_5	0	0	0	0	0	0
3	0	c_1	c_2	c_4	0	0	0
4	0	0	d_5	0	d_2	d_6	d_7

Results of the order and rank conditions of identification

Equation #	No.of predetermined variables excluded (K-k)	No. of endogenous variables included less one (m-1)	Rank of matrix A	Identification ?
1	9	1	3	overidentified
2	8	2	3	overidentified
3	7	0	3	overidentified
4	5	1	3	overidentified

APPENDIX IV : TEST RESULTS OF THE MODEL ASSUMPTIONS

Test for zero mean:

Assumption: $E(u_i) = 0$

Test: t test

H_0: $E(u_i) = 0$

H_1: $E(u_i) \neq 0$

Accept H_0 if $t^* < t(0.05)$, otherwise reject H_0 and accept H_1.

where :

$t^* = (u./n)/s_u$

s_u = standard error of the estimate

n = the number of observations (30)

Equation #	u./n	s_u	t*	df	$t_{0.05}$	Result accept H_0	Result reject H_0
1	-1.36E-13	103.64	0.0	26	2.056	x	
2	-1.02E-13	69.56	0.0	24	2.064	x	
3	-2.35E-13	144.07	0.0	25	2.060	x	
4	-1.52E-14	7.12	0.0	22	2.074	x	

Test for zero-covariance (autocorrelation) of the error term.

Assumption: $\rho = 0$ (the u_i is not correlated with u_{i-1})

Test: The Durbin-Watson test

H_0: $\rho = 0$

H_1: $\rho \neq 0$

Accept H_0 if $d^* > d_u(0.05)$. Reject H_0 if $d^* < d_u(0.05)$

where:

$$d^* = 2\{1-[(\Sigma e_t \; e_{t-1})/\Sigma e^2{}_t]\} \qquad [4.13]$$

where:

e_t = estimated error term in period t;

e_{t-1} = estimated error term in period t-1.

Equation #	ρ	d*	$d_u(0.05)$	Result accept H_0	Result reject H_0
1	0.32	1.32	1.65		x
2	0.44	1.07	1.83		x
3	0.31	1.32	1.74		x
4	-0.098	2.18	2.03	x	

Test for constant variance:

Assumption: $V(u^2_i) = \sigma^2$ $i=1,2,\text{------}n$

Test: Goldfeld-Quandt test

H_0: $V(u^2_i) = \sigma^2$ $i=1,2,\text{------}n$

H_1: $V(u^2_i) \neq \sigma^2$ $i=1,2,\text{------}n$

Reject H_0 if $\lambda > F\ (a,b)$, otherwise accept H_0, where

$\lambda = (Rss2/a)/(Rss1/b)$ with $df=(a,b)$

$a = df(Rss2)$

$b = df(Rss1)$

where:

$a = \{(n-C)/2\}-K$

$b = a$

n = # of observations

C = # of omitted central observations

K = number of parameters to be estimated, including the intercept.

Equation #	λ	a	b	$F_{0.05}$	Result accept H_0	reject H_0
1	0.36	7	7	3.79	x	
2	0.44	5	5	5.05	x	
3	0.42	6	6	4.28	x	
4	0.02	3	3	9.28	x	

APPENDIX V : 2SLS ESTIMATES FROM SHAZAM

Two stage least squares - dependent variable = y_1
11 exogenous variables
30 observations

R-square = 0.9925 R-square adjusted = 0.9916
Variance of the estimate = 10742.0
Standard error of the estimate = 103.64
Mean of dependent variable = 7278.6

Variable name	estimated coefficient	Standard error	t-ratio 26 df
y_2	0.3708	0.1144	3.24
x_{24}	0.7421	0.0588	12.63
x_{26}	59.3150	7.0330	8.43
Constant	-6086.5	529.03	-11.50

Observation no.	Observed value	Predicted value	Calculated residual
1	5657.0	5757.0	-100.02
2	5850.0	6023.3	-173.32
3	5773.0	5779.3	-6.32
4	5482.0	5404.8	77.18
5	5829.0	5740.1	88.87
6	6125.0	6062.7	62.28
7	6085.0	6044.6	40.37
8	6060.0	6012.8	47.19
9	6007.0	5959.9	47.07
10	6632.0	6615.1	16.86
11	7028.0	7071.2	-43.23
12	7607.0	7737.0	-129.96
13	7229.0	7413.6	-184.57
14	6885.0	6856.6	28.41
15	7484.0	7464.7	19.34
16	7403.0	7289.6	113.42
17	7112.0	7179.7	-67.70
18	7570.0	7481.5	88.55
19	8012.0	8006.9	5.06
20	8706.0	8566.8	139.15
21	7010.0	7022.4	-12.44
22	7903.0	8027.5	-124.50
23	8169.0	8097.9	71.07
24	8913.0	8760.0	152.97
25	8779.0	8731.2	47.81
26	8622.0	8562.5	59.45
27	8912.0	8795.6	116.40
28	8055.0	8207.2	-152.19
29	8439.0	8597.2	-158.22
30	9019.0	9088.0	-68.97

Durbin-Watson = 1.32   Rho = 0.32
Residual sum = -0.409E-11
Sum of absolute error = 2442.9
R-square between observed and predicted = 0.9925

Two stage least squares - dependent variable = y_2
11 exogenous variables
30 observations

R-square = 0.9904 R-square adjusted = 0.9883
Variance of the estimate = 4838.3
Standard error of the estimate = 69.558
Mean of dependent variable = 5424.7

Variable name	Estimated coefficient	Standard error	t-ratio 24 df
x_5	-0.57222	0.066	-8.61
y_3	0.20848	0.047	4.40
y_1	0.57945	0.044	13.16
x_{17}	3.45870	0.503	6.87
x_{35}	-0.32777	0.121	-2.71
Constant	995.38000	185.650	5.36

Observation no.	Observed value	Predicted value	Calculated residual
1	4599.0	4663.1	-64.145
2	4745.0	4760.2	-15.223
3	4586.0	4592.7	-6.746
4	4379.0	4393.1	-14.070
5	4643.0	4613.1	29.857
6	4789.0	4808.0	-19.012
7	4742.0	4757.8	-15.775
8	4744.0	4738.2	5.780
9	4699.0	4694.1	4.859
10	5124.0	5156.8	-32.829
11	5528.0	5502.2	25.752
12	5997.0	5857.4	139.620
13	5682.0	5509.6	172.440
14	5194.0	5175.3	18.736
15	5470.0	5484.5	-14.485
16	5242.0	5326.3	-84.349
17	5227.0	5174.7	52.282
18	5469.0	5533.6	-64.637
19	5970.0	6002.9	-32.878
20	6304.0	6270.8	33.179
21	4980.0	4951.4	28.587
22	5671.0	5639.3	31.656
23	5751.0	5811.7	-60.704
24	6443.0	6529.7	-86.661
25	6371.0	6495.8	-124.780
26	6118.0	6152.2	-34.173
27	6058.0	6072.9	-14.947
28	5596.0	5571.8	24.162
29	6034.0	6023.0	11.044
30	6586.0	6478.5	107.460

Durbin-Watson = 1.07   Rho = 0.44
Residual sum = -0.307E-11
Sum of absolute errors = 1370.8
R-square between observed and predicted = 0.9904

Two stage least square - dependent variable = y_3
11 exogenous variables
30 observations

R-square = 0.9926 R-square adjusted = 0.9914
Variance of the estimate = 20756.0
Standard error of the estimate = 144.07
Mean of dependent variable = 8407.8

Variable name	Estimated coefficient	Standard error	t-ratio 25 df
x_{22}	4.7514	0.468	10.15
x_{12}	-184.6800	21.023	-8.78
x_{17}	16.5300	2.423	6.82
x_{18}	69.7360	29.418	2.37
Constant	-3373.1000	2851.700	-1.18

Observation no.	Observed value	Predicted value	Calculated residual
1	6022.0	5870.3	151.72
2	6259.0	6002.9	256.07
3	6228.0	6210.0	18.00
4	6028.0	6168.6	-140.58
5	6488.0	6534.3	-46.34
6	6737.0	6772.4	-35.37
7	6695.0	6702.3	-7.34
8	6792.0	7076.1	-284.12
9	6847.0	6949.3	-102.25
10	7295.0	7453.1	-158.13
11	7757.0	7818.2	-61.25
12	8304.0	8158.0	146.04
13	8381.0	8175.9	205.12
14	8461.0	8499.0	-38.05
15	8940.0	8769.6	170.45
16	8805.0	8663.9	141.13
17	8922.0	8926.9	-4.93
18	9551.0	9431.0	120.04
19	9741.0	9826.3	-85.29
20	9262.0	9295.3	-33.31
21	8325.0	8535.7	-210.65
22	8670.0	8827.5	-157.49
23	9281.0	9186.6	94.43
24	9864.0	9873.1	-9.09
25	10199.0	10245.0	-45.81
26	10088.0	10103.0	-15.15
27	10164.0	10174.0	-10.31
28	10198.0	9953.7	154.30
29	10589.0	10444.0	145.25
30	11431.0	11588.0	-157.09

Durbin-Watson = 1.32   Rho = 0.31
Residual sum = 0.7048E-11
Sum of absolute errors = 3205.1
R-square between observed and predicted = 0.9926

Two stage least squares - dependent variable = y_4

11 exogenous variables
30 observations

R-square = 0.9982 R-square adjusted = 0.9977
Variance of the estimate = 50.652
Standard error of the estimate = 7.117
Mean of dependent variable = 251.44

Variable name	Estimated coefficient	Standard error	t-ratio 22 df
x_5	-0.016205	0.006	-2.95
x_{21}	0.445070	0.079	5.61
y_1	-0.017495	0.003	-5.80
x_{17}	1.274900	0.177	7.22
x_{12}	6.764600	1.209	5.60
x_{151}	-0.042586	0.010	-4.12
x_{33}	4.115400	0.972	4.24
Constant	138.320000	14.942	9.25

Observation no.	Observed value	Predicted value	Calculated residual
1	138.89	142.21	-3.3184
2	143.30	143.08	0.2244
3	146.61	145.66	0.9524
4	147.71	144.03	3.6796
5	147.71	141.27	6.4390
6	147.71	148.65	-0.9436
7	147.71	147.80	-0.0913
8	147.71	143.76	3.9538
9	147.71	154.08	-6.3733
10	147.71	147.88	-0.1686
11	147.71	149.57	-1.8575
12	149.91	145.41	4.4982
13	154.87	154.61	0.2578
14	156.53	158.77	-2.2377
15	162.04	166.47	-4.4299
16	167.55	177.50	-9.9488
17	174.17	183.39	-9.2203
18	181.42	169.31	12.1120
19	192.90	189.63	3.2679
20	231.49	223.70	7.7950
21	286.60	291.95	-5.3500
22	311.40	301.02	10.3840
23	333.45	336.88	-3.4319
24	348.33	360.21	-11.8770
25	425.27	427.25	-1.9757
26	480.61	478.01	2.5981
27	531.87	532.87	-0.9952
28	545.46	532.11	13.3470
29	528.37	536.14	-7.7709
30	570.45	569.97	0.4812

Durbin-Watson = 2.18 Rho = -0.098
Residual sum = -0.45475E-12
Sum of absolute errors = 139.98
R-square between observed and predicted = 0.9982

APPENDIX VI: SHAZAM COCHRANE-ORCUTT METHOD RESULTS FOR SOLVING THE AUTOCORRELATION PROBLEM.

Dependent variable = y_1
Estimated Rho = 0.28747

R-square = 0.9926 R-square adjusted = 0.9917
Variance of the estimate = 10641
Standard error of the estimate = 103.15
Mean of dependent variable = 7278.6

Variable name	Estimated coefficient	Standard error	t-ratio 26 df
y_2	0.33946	0.132	2.58
x_{24}	0.75824	0.070	10.80
x_{26}	62.21700	8.350	7.46
Constant	-6315.70000	644.100	-9.81

Durbin-Watson = 1.8297 Rho = 0.05936
Residual sum = 34.404
R-square between observed and predicted = 0.9925

Observation no	Observed value	Predicted value	Calculated residual
1	5657.0	5776.7	-119.68
2	5850.0	6046.3	-196.33
3	5773.0	5784.1	-11.12
4	5482.0	5388.8	93.25
5	5829.0	5711.8	117.21
6	6125.0	6061.0	63.96
7	6085.0	6045.8	39.19
8	6060.0	6007.5	52.52
9	6007.0	5913.0	75.97
10	6632.0	6634.2	-2.25
11	7028.0	7074.6	-46.64
12	7607.0	7716.5	-109.47
13	7229.0	7382.0	-153.01
14	6885.0	6819.2	65.78
15	7484.0	7460.9	23.11
16	7403.0	7297.5	105.54
17	7112.0	7182.0	-70.02
18	7570.0	7490.5	79.55
19	8012.0	8012.3	-0.32
20	8706.0	8557.7	148.27
21	7010.0	7016.9	-6.88
22	7903.0	8054.4	-151.41
23	8169.0	8088.1	80.92
24	8913.0	8747.3	165.70
25	8779.0	8764.4	14.57
26	8622.0	8537.5	84.53
27	8912.0	8817.3	94.74
28	8055.0	8207.4	-152.36
29	8439.0	8603.5	-164.53
30	9019.0	9070.7	-51.66

R-square between observed and predicted = 0.9919
Sum of squares of residuals = 0.30264E+06

Dependent variable = y_2
Estimated Rho = 0.35408
R-square = 0.9909 R-square adjusted = 0.9890
Variance of the estimate = 4596.5
Standard error of the estimate = 67.598
Mean of dependent variable = 5424.7

Variable name	Estimated coefficient	Standard error	t-ratio 24 df
x_5	-0.53957	0.072	-7.54
y_3	0.17286	0.051	3.37
y_1	0.61118	0.048	12.85
x_{17}	3.40250	0.617	5.52
x_{35}	-0.34374	0.117	-2.93
Constant	990.31000	210.400	4.71

Durbin-Watson = 1.8709 Rho = -0.00174
Residual sum = 29.932
R-Square between observed and predicted = 0.9908

Observation no	Observed value	Predicted value	Calculated residual
1	4599.0	4683.5	-84.54
2	4745.0	4806.0	-60.97
3	4586.0	4615.0	-28.99
4	4397.0	4349.9	29.12
5	4643.0	4558.4	84.56
6	4789.0	4770.9	18.09
7	4742.0	4718.9	23.13
8	4744.0	4744.9	-0.89
9	4699.0	4633.5	65.52
10	5124.0	5159.8	-35.84
11	5528.0	5532.2	-4.18
12	5997.0	5882.0	114.84
13	5682.0	5549.1	132.95
14	5194.0	9137.2	56.84
15	5470.0	5452.6	17.39
16	5242.0	5274.1	-32.13
17	5227.0	5253.0	-26.02
18	5469.0	5512.2	-43.17
19	5970.0	6021.6	-51.60
20	6304.0	6232.4	71.64
21	4980.0	5008.6	-28.56
22	5671.0	5710.3	-39.26
23	5751.0	5760.9	-9.91
24	6443.0	6468.8	-25.79
25	6371.0	6553.8	-182.82
26	6118.0	6121.0	-3.04
27	6058.0	6130.3	-72.35
28	5596.0	5619.4	-23.42
29	6034.0	6000.6	33.45
30	6586.0	6487.6	98.36

R-square between observed and predicted = 0.9897
Sum of squares of residuals = 0.12505E+06

Dependent variable = y_3
Estimated Rho = 0.36939
R-square = 09935   R-square adjusted = 0.9924
Variance of the estimate = 18397
Standard error of the estimate = 135.64
Mean of dependent variable = 8407.8

Variable name	Estimated coefficient	Standard error	t-ratio 25 df
X_{22}	4.7436	0.493	9.62
X_{12}	-170.6500	21.860	-7.81
X_{17}	14.9630	2.463	6.08
X_{18}	68.9550	26.808	2.57
Constant	-3330.1000	2580.600	-1.29

Durbin-Watson = 1.8596   Rho = 0.031162
Residual sum = -47.570
R-square between observed and predicted = 0.9934

Observation no	Observed value	Predicted value	Calculated residual
1	6022.0	5893.2	128.78
2	6259.0	6027.6	231.38
3	6228.0	6225.5	2.51
4	6028.0	6181.5	-153.49
5	6488.0	6541.6	-53.63
6	6737.0	6768.6	-31.58
7	6695.0	6713.3	-18.34
8	6792.0	7081.8	-289.76
9	6847.0	6960.7	-113.71
10	7295.0	7453.4	-158.44
11	7757.0	7811.5	-54.50
12	8304.0	8152.1	151.91
13	8381.0	8178.4	202.60
14	8461.0	8495.9	-34.91
15	8940.0	8754.4	185.63
16	8805.0	8653.1	151.89
17	8922.0	8909.5	12.49
18	9551.0	9399.1	151.92
19	9741.0	9797.1	-56.07
20	9262.0	9293.5	-31.52
21	8325.0	8576.3	-251.33
22	8670.0	8866.7	-196.69
23	9281.0	9224.2	56.80
24	9864.0	9885.6	-21.57
25	10199.0	10238.0	-38.80
26	10088.0	10102.0	-14.18
27	10164.0	10171.0	-7.24
28	10108.0	9964.8	143.21
29	10589.0	10427.0	162.46
30	11431.0	11514.0	-82.78

R-square between observed and predicted = 0.9924
Sum of squares of residuals = 0.53417E-06

APPENDIX VII: THE RESULTS OF TESTING THE FORECASTING POWER OF THE ESTIMATED MODEL

Test of significance of the difference between predicted (Y_f) and observation (Y_a) values.

Assumption: $Y_f = Y_a$

Test: t test

H_0: $Y_f = Y_a$
H_1: $Y_f \neq Y_a$

accept H_0 if $|t| < t_{0.05}$, otherwise reject H_0, accept H_1,

where:

$t = (Y_a - Y_f)/s_{Yf}$

Y_a = actual (observed) value of Y

Y_f = predicted (forecast) value of Y from the regression

s_{Yf} = the standard error of the forecast.

Equation #	Year	Y_a**	Y_f**	t	df	$t_{0.05}$	Accept H_0
1	1985	8996	9093.6	-0.149	26	2.056	y
	1986	9302	9318.9	-0.026	26	2.056	y
2	1985	6674	6614.6	0.271	24	2.064	y
	1986	6734	6826.8	-0.393	24	2.064	y
3	1985	11587	12307.0	-0.278	25	2.060	y
	1986	11937	13077.8	-0.440	25	2.060	y
4	1985	589.7	635.4	-1.970	22	2.074	y
	1986	599.4	641.4	-1.789	22	-2.074	y

** units are the same as in Appendix I.

Analysis of the sources of the forecasting error

Measure: Partial inequality coefficients

bias proportion $U_m = (\bar{P}-\bar{A})^2 / [\Sigma(P_i-A_i)^2/n]$

variance proportion $U_s = (s_p-s_a)^2/[\Sigma(P_i-A_i)^2/n]$

covariance proportion $U_c = [2(1-r_{pa})s_p s_a]/ [\Sigma(P_i-A_i)^2/n]$

where:

$\bar{P} = (1/n)(\Sigma P_i)$
$\bar{A} = (1/n)(\Sigma A_i)$
$s_p^2 = (1/n)[\Sigma(P_i-P)^2]$
$s_a^2 = (1/n)[\Sigma(A_i-A)^2]$
$r_{pa} = [\Sigma(P_i-P)(A_i-A)]/(ns_p s_a)$

Equation no.	U_m	U_s	U_c
1	0.5263E-03	0.6525E-03	0.9988
2	0.1530E-04	0.3486E-02	0.9965
3	0.4531E-04	0.1625E-01	0.9837
4	0.3058E-29	0.5005E-03	0.9995