is still the most current available. At the time of writing, the full 1971 Census material was not expected to be published until (at least) mid-year 1974. The hope of the author is that, having demonstrated the feasibility and usefulness of the present approach, the regression equations -- and more importantly, the full econometric model -- can be re-estimated once the 1971 information becomes available. Simple correlation results are outlined in Appendix C, while the (single equation) multiple regression findings are discussed in Appendix E.

The major explanatory variables in the regression analysis are county differences in industrial and occupational structure, as measured by the percentage share of a given industry or occupation of the total 'experienced'²/labour force of a given county. Accordingly, a typical regression might have average earnings per wage earner for the 46 counties -- or more precisely, a vector containing average earnings per wage earner for 46 county observations -- as the dependent variable, with the shares of the county labour force in primary industries, manufacturing and certain service industries providing the right hand variables. The 1961 Census offers labour force information for 26 individual industries and 12 occupational categories. As mentioned later, some of these categories were added together where appropriate.

1

OF/DE

5

21226

NATIONAL LIBRARY
OTTAWA BIBLIOTHÈQUE NATIONALE
OTTAWANAME OF AUTHOR... *DEREK JOHN IRELAND*TITLE OF THESIS... *A Model Of Income**DIFFERENTIALS IN THE*
*ATLANTIC REGION*UNIVERSITY... *CARLETON*DEGREE FOR WHICH THESIS WAS PRESENTED... *MASTERS OF ARTS*YEAR THIS DEGREE GRANTED... *1974*THESIS SUPERVISOR *Keith A.J. Hay*

Permission is hereby granted to THE NATIONAL LIBRARY
OF CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and
neither the thesis nor extensive extracts from it may be
printed or otherwise reproduced without the author's
written permission.

(Signed) *Derek J. Ireland*

PERMANENT ADDRESS

22 Presb. St.
Riverview
*New Brunswick*DATED *October 2* 1974

NL-91 (10-68)

A MODEL OF INCOME DIFFERENTIALS

IN THE ATLANTIC REGION

DEREK J. IRELAND
MASTERS THESIS
CARLETON UNIVERSITY

FALL, 1974

The undersigned recommend to the Faculty of
Graduate Studies acceptance of the thesis "A Model
of Income Differentials in the Atlantic Region"
submitted by Joseph John Ireland, M.A. Honours,
in partial fulfillment of the requirements for
the Degree of Master of Arts.

Thesis Supervisor

Chairman, Department of Economics

Carleton University

July 26, 1973

ABSTRACT

The objective of this thesis is to apply econometric techniques to the analysis of earnings and income differentials in the Atlantic Region. A twenty-one equation econometric model is developed, using cross-sectional data for the forty-six counties in the Region for 1961. Two Stage Least Squares was used to estimate the model, and because of non-linearities, a Gauss-Seidel iterative technique was applied to solve it. The major endogenous variables are average earnings, employment income per capita and the degree of inequality in income, and the model is divided into five blocks: earnings and income; labour force utilization; "changes over time"; demographic; and industry.

The estimation and solution results clearly indicate that a cross-sectional model of this type can provide useful insights into the wage and income determination process in the Atlantic Region. The thesis concludes with a number of policy implications that can be derived from the model.

TABLE OF CONTENTS

	PAGE
Foreword	i
Summary	ii
1. INTRODUCTION	1
2. REVIEW OF THE LITERATURE	14
(a) Factors Associated with Income and Earnings Differentials	14
(b) Dynamic Considerations and the Export Base Theory	36
(c) The Size Distribution of Income	43
(d) Labour Market Theory	48
(e) Summary and Testable Hypotheses	57
3. METHODOLOGY	
4. SUMMARY OF AN ECONOMETRIC MODEL OF INCOME DIFFERENTIALS IN THE ATLANTIC REGION	80
5. EARNINGS AND INCOME BLOCK	91
(a) Theory and Underlying Rationale	91
(b) Estimation Results	94
6. LABOUR FORCE UTILIZATION BLOCK	100
(a) Theory and Underlying Rationale	100
(b) Estimation Results	106

TABLE OF CONTENTS (CONT'D)

	<u>PAGE</u>
7. CHANGES OVER TIME BLOCK	114
(a) Theory and Underlying Rationale	114
(b) Estimation Results	116
8. DEMOGRAPHIC BLOCK	120
(a) Theory and Underlying Rationale	120
(b) Estimation Results	123
9. INDUSTRY BLOCK	126
(a) Theory and Underlying Rationale	126
(b) Estimation Results	130
10. SOLUTION RESULTS	138
(a) Introduction	138
(b) Sensitivity Tests	140
(c) Alternative Scenarios	163
(d) Ontario and Canada Simulations	174
11. FINAL CONCLUSIONS	183
FOOTNOTES	196
BIBLIOGRAPHY	221

TABLE OF CONTENTS (CONT'D)

	PAGE
APPENDIX A: THEORETICAL FOUNDATIONS FOR THE WAGE AND INCOME EQUATIONS	1-18
APPENDIX B: MNEMONICS	1-7
APPENDIX C: SIMPLE CORRELATION ANALYSIS	1-14
APPENDIX D: STANDARDIZATION ANALYSIS: METHODOLOGY AND RESULTS	1-2
APPENDIX E: SINGLE EQUATION ESTIMATION: EARNINGS AND INCOME PER CAPITA	1-48
APPENDIX F: PRINCIPAL COMPONENTS	1-20
APPENDIX G: COMPARISON BETWEEN THE ORDINARY LEAST SQUARES AND TWO STAGE LEAST SQUARES RESULTS	1-23
APPENDIX H: COMPARISON OF PRELIMINARY REPORT RESULTS WITH FINAL RESULTS	1-7
APPENDIX I: DATA SOURCES	1-3

FOREWORD

The present study will hopefully serve three purposes. First, it is to be submitted as a Masters Thesis to the Economics Department of Carleton University. Second, the study represents the author's course project for the Quantitative Analysis Course of the Treasury Board which the author attended from May 1973 to March 1974. Finally, it will be distributed to interested parties in the Department of Regional Economic Expansion and other federal and provincial government departments in the hope that it will contribute to their research on the underlying causes of regional disparities.

The author would like to thank the large number of people who assisted him in this task. A partial listing would include: first, Professor K.A.J. Hay, the author's thesis adviser at Carleton, who attempted -- sometimes in vain -- to keep the economics "honest"; second, Professor David Foot of the Quantitative Analysis Course, who taught the author more econometrics than anyone would have thought possible; and third, the author's two immediate supervisors at DREE as of May, 1973, Messrs. D.J. McDonald and N.G. Mulder, without whose "benign neglect" during the past year this study would not have been possible.

A MODEL OF INCOME DIFFERENTIALS
IN THE ATLANTIC REGION: SUMMARY

In recent years, the problem of regional disparities has become a major policy concern of most developed countries. Nowhere is this more true than in Canada where both the federal and provincial levels of government have been very active in the regional development field since the early 1960's. It is perhaps surprising however that, with such a high level of government activity, so little rigorous analysis has been done on the primary or causal factors underlying regional inequalities in income and economic development. One of the more obvious deficiencies has been the failure to date to apply the more advanced econometric techniques -- multiple regression, principal components, simultaneous equations estimation -- to the study of regional earnings and income differentials. One purpose of the present study therefore is to learn what these techniques can tell us about the root causes of regional income disparities.

The original objective of this thesis was to apply single equation estimation techniques, including both simple and multiple regression, in an attempt to explain earnings and per capita income differentials among counties in the

Atlantic Region. However, because of problems with multicollinearity and insufficient degrees of freedom, the methodological scope of the exercise was expanded considerably to include, first, principal components analysis and, second, the formulation of a full econometric model using cross-sectional data for the forty-six counties in the Atlantic Region for 1961.^{1/} The development, estimation and solution of the model ultimately became the principal objective of this work and thus provides the major focus for the body of the report and for the present summary. The single equation estimation results and principal components analysis are discussed in some detail in the attached appendices.

The final model includes twenty-one equations, of which eighteen are behavioural. The major endogenous variables in the model are average earnings, employment income per capita and the degree of inequality in income (the interquartile range divided by the median in the non-farm income series). Such variables as the unemployment and participation rates, the degree of seasonality of the work force, the levels of urbanization and education, and a number of industry share

^{1/}At the time of writing, the full 1971 Census information was not yet available, particularly the sub-provincial data.

variables^{2/} became "intermediate endogenous" variables, since these can be viewed as endogenous to and thus determined within the economic system. The model is divided into five blocks: Earnings and income; labour force utilization; "changes over time;" demographic; and industry. In the final work, the model was estimated by Two Stage Least Squares and, because it was non-linear, it was solved using a Gauss-Seidel iterative technique. The TROLL package of the National Bureau of Economic Research was utilized in the final estimation and solution of the model.

The present analysis begins with a fairly comprehensive review of the literature concerning the generation of earnings and income differentials and differences in the size distribution of income among regions. This chapter therefore focusses on the main endogenous variables in the model. It was concluded that the literature is relatively weak with respect to the "causal" factors underlying regional income differentials, but it does offer a wealth of testable hypotheses. The literature also provided insights concerning other variables that should be included in the model. The testable hypotheses would include:

^{2/} That is, employment in a given industry as a percentage of total county employment.

- industrial and occupational distributions will likely be the critical determinants but cannot be expected to explain all of the interregional variation in earnings and incomes;

- earnings and incomes can be expected to be negatively related to the percentage of county employment in the primary industries, and positively associated with the importance of secondary manufacturing and most service activities;

- among industries, the exporting activities will perhaps play the primary role in determining earnings and per capita income in a given area, but account must also be taken of mainly residentiary activities;

- urbanization should be positively related to earnings and incomes, reflecting both higher living costs and higher productivity associated with agglomeration economies;

- the rate of labour force utilization will have a strong and direct impact on per capita incomes and, moreover, as a measure of the degree of disequilibrium in labour markets, may also influence average earnings in a given area;

- the degree of seasonality and average hours worked will also affect earnings -- particularly annual earnings -- and per capita income through their impact on the size of the employment base;

- the level of education or skills should have a positive impact on earnings and income -- even after industry composition is taken into account -- resulting from both its "non-neutrality in production" plus the choice of more skill intensive processes by industries in areas where skilled labour is relatively more abundant.

A few hypotheses were also gleaned concerning the elements underlying the size distribution of income:

- the distribution of income will likely be more equal the higher the level of economic development in a given area, as measured by the share of employment in secondary

manufacturing or other more modern industries, or by the occupational distribution, or by per capita income itself;

- the distribution will tend to be more equal the lower the unemployment rate and the higher the labour force participation rate;

These hypotheses, among others, were tested in modelling the three behavioural equations in the earnings and income block, while other antecedents were utilized in modelling the remaining equations in the model. For the five behavioural equations in the labour force utilization block, the literature and theoretical antecedents were almost non-existent, reflecting the cross-sectional nature of the model. Accordingly, a more "heuristic" approach had to be applied here than perhaps in any other section of the model, although a few guidelines were provided by time series models.

For the change over time block, the antecedents varied with the equation; in modelling the net migration rate, an "economic" approach focussing on income and/or employment opportunities was attempted, while for the growth in employment a variant of shift/share analysis was applied. For the demographic block, the underlying rationale was that industries have differing requirements for (i) the amenities, services, etc. offered by an urban location, and (ii) skilled labour, and that these differences generate in turn intercounty variations in the levels of urbanization and education.

Finally, the antecedents for the five behavioural equations in the industry block lie mainly in (i) the concept of input-output access developed by Perloff, (ii) the concept of agglomeration economies, (iii) the work of Chenery who focussed on the role of per capita income and population in generating international differences in the levels of industrialization and services.

The following outlines the endogenous and exogenous variables in the model. (The variables are more fully described in Appendix B.) Then, the "structural form" of the model is presented, focussing in particular on the estimation results for the eighteen behavioural equations: (The t-statistics are indicated under each estimated coefficient, while the three identities are noted by an asterisk.)

ENDOGENOUS VARIABLES (1961 Unless Otherwise Noted)

Earnings and Income Block

- EIC: Employment Income Per Capita (\$)
- ER: Employment Ratio (Employment as Per Cent of Total Population) (%)
- ELFR: Employed Labour Force Ratio (Employment as Per Cent of Total Labour Force) (%)
- W: Average Annual Earnings (\$)
- DII: Degree of Inequality in Income (the Inter-Quartile Range Divided by the Median for the Income Per Non-Farm Family Series)

Labour Force Utilization Block

UR: Unemployment Rate (%)
PR: Labour Force Participation Rate (%)
SE: Seasonality (Per Cent of Wage Earners
Working 26 Weeks or Less Per Year)
HO: Hours (Per Cent of Wage Earners Working Less
than 35 Hours a Week)
PFELF: Per Cent Female in Employed Labour Force

Change Over Time Block

GLFP: Growth in Labour Force Population, 1951-61
Index (No Change = 100) *
NMR: Net Migration Rate, 1951-61
GE: Growth in Employment, 1951-61, Index

Demographic Block

URB: Per Cent Urban
LED: Level of Education (Per Cent of Population 5 and over
not Attending School with at Least Two Years of
Secondary Schooling)

Industry Block

HOBS: Higher Order Business Services (%)
OHVAM: Other High Value Added Manufacturing
(i.e. Excluding Pulp and Paper) (%)
RS: Residentiary Services (%)

EH: Education and Health (%)
LVAM: Low Value Added Manufacturing (%)
IR: Industry Residual (Utilities Plus "Industry
not Specified") (%)

21 Endogenous Variables

EXOGENOUS (OR PREDETERMINED) VARIABLES

AG: Agriculture (%)
FOR: Forestry (%)
FISH: Fishing (%)
WPR: Working Population Ratio (Labour Force Source
Population as Per Cent of Total Population)
PA: Public Administration and Defense (%)
TRANS: Transportation and Storage (%)
CON: Construction (%)
PFR: Per Cent French
MIN: Mining (%)
NRI: Natural Rate of Increase (%), 1951-61
NFD: Newfoundland Dummy Variable
POP: Total Population, 1961
UD: University Dummy Variable
PP: Pulp and Paper (%)

14 Exogenous Variables

EQUATIONS, BY BLOCK

Earnings and Income Block

1.	Ln EIC	=	1.087 (1.58)	+0.827 Ln W (10.24)	+0.751 Ln (ER ÷ 100) (7.37)	.970	
2.*	Ln (ER) (100)	=	Ln (ELFR) (100)	+Ln (PR) (100)	+Ln (WPR) (100)	--	
3.*	ELFR = 100 - UR					--	
4.	$\hat{W}$	=	3,898.75 (19.91)	-20.38 AG (-3.74)	-28.92 FOR (-2.14)	-13.36 FISH (-2.07)	.897
			34.19 LVAM (-3.20)	+12.13 PP (1.89)	-17.99 EH (-1.59)	-15.54 SE (-1.35)	
				-58.41 HO (-1.91)			
5.	$\hat{DII}$	=	1.103 (6.46)	-.0108 ER (-2.11)	+.0158 HO (2.18)		.353
			-.0122 OHVAM (-2.10)	-.0044 FOR (-1.03)			

Labour Force Utilization Block

6.	$\widehat{\text{UR}} = -2.96$ (-0.47)	-0.134 AG (-2.55)	-0.200 LVAM (-2.00)	$+0.147 \text{ RS}$ (0.79)	.709
	-0.151 TRANS (-1.59)	$+0.176 \text{ GLFP}$ (2.18)	-0.106 GE (-1.81)	$+0.158 \text{ SE}$ (1.69)	

$\bar{R}^2$

(7) $\hat{PR} = 33.20$ -0.176 GLFP $+0.216$ GE $+0.171$ PFELF .818
 (6.69) (-3.01) (3.93) (1.10)

$+0.093$ PA $+0.147$ AG $+0.186$ PP $+0.361$ HOBS
 (1.04) (3.31) (2.39) (1.74)

(8) $\hat{SE} = 2.10$ $+0.468$ AG $+0.759$ FOR $+0.516$ FISH .804
 (0.49) (5.90) (3.69) (6.37)

$+0.338$ RS -0.444 EH $+0.452$ CON $+0.472$ UR
 (1.24) (-1.48) (2.20) (1.14)

$+0.053$ PFR
 (2.06)

(9) $\hat{HO} = 13.03$ -0.123 FOR $+0.089$ FISH -0.324 EH .615
 (9.01) (-2.97) (3.12) (-3.07)

-0.121 PA -0.094 TRANS -0.077 MIN
 (-5.64) (-1.40) (-1.63)

(10) $\hat{PFELF} = 5.29$ $+0.656$ HOBS $+0.258$ EH .775
 (2.15) (3.34) (1.26)

$+0.562$ RS -0.225 TRANS $+0.054$ PFR
 (2.23) (-1.90) (3.36)

Change Over Time Block

(11) $\hat{GLFP} = 95.48$ $+0.850$ NMR $+0.829$ NRI .995
 (153.46) (60.61) (34.53)

(12) $\hat{NMR} = -72.84$ $+0.700$ GE -0.112 NRI -0.074 PFR .924
 (-19.79) (17.97) (-1.20) (-2.14)

(13) $\hat{GE} = 170.68$ -0.679 AG -0.988 FOR -1.102 FISH .700
 (3.57) (-1.42) (-1.66) (-1.96)

-1.387 MIN 4.678 RS $+0.880$ PA $+0.347$ URB
 (-1.90) (-1.85) (1.86) (1.69)

Demographic Block

R^2

(14)	$\hat{URB} = -25.50$ (-2.42)	+1.14 PP (2.72)	+1.78 EH (1.52)	+0.79 PA (4.09)	.609
	+1.39 HOBS (1.35)	+4.79 OHVAM (4.02)	+4.83 NFD (4.46)		
(15)	$\hat{LED} = 3.98$ (0.54)	-0.326 FOR (-2.05)	-0.171 FISH (-1.81)	+0.654 HOBS (1.22)	.750
	+0.629 RS (1.08)	+0.358 EH (1.00)	+0.218 PA (2.65)	-0.233 TRANS (-1.13)	

Industry Block

(16)	$\hat{HOBS} =$	-7.27 (-3.67)	+0.273 LED (4.48)	+0.100 EH (0.61)	+0.215 TRANS (2.01)	.711	
		+0.021 (POP $\div$ 1,000) (2.36)		+0.510 RS (2.81)	+0.070 AG (1.74)		
		+0.104 LVAM (1.23)					
(17)	$\hat{OHVAM} =$	-7.55 (-3.49)	+0.303 MIN (3.76)	+0.217 HOBS (1.14)		.553	
		+0.556 RS (2.39)	-1.83 NFD (-2.05)	-1.24 UD (-1.33)			
(18)	$\hat{RS} =$	14.98 (9.04)	+1.013 (EIC $\div$ 100) (3.53)	+0.043 URB (1.74)	-0.084 GE (-4.46)	.492	
(19)	$\hat{EH} =$	3.15 (4.07)	+0.414 HOBS (6.24)	+5.05 UD (7.96)			
(20)	$\hat{LVAM} =$	6.09 (0.82)	+0.188 FISH (1.48)	-0.431 (W $\div$ 100) (-2.13)	+0.729 RS (2.45)	.224	
(21)*	IR = 100-AG	-FOR	-FISH	-MIN	-PP	-CON	-TRANS
		-PA	-HOBS	-OHVAM	-RS	-EH	-LVAM

Closer scrutiny of the equations will indicate some of the key "links" between blocks, defined as a left hand variable in one block which appears as a right hand variable in a second block. Some of these would include: the four labour force utilization measures which are used as explanatory variables in the earnings and income block; GE and GLFP from the change over time block which feed into the labour force utilization section; and the variables from the industry block which appear as right hand variables in all of the remaining blocks in the model. Earnings and income is the most critical block in the model and thus all other blocks, either directly or indirectly, feed into this part of the system. At the same time, there are more linkages, feed backs and within block interdependencies associated with the industry block than with any other, reflecting the critical role played by industry in the specification of the model.

No mathematical problems were encountered in solving the model but the results were not always completely feasible in an economic sense. In particular, negative values were at times secured for the industry residual, making necessary the application of an "iterative procedure" in order to generate a feasible solution set. In general though, the

econometric model provided reasonable solution values.

At first, a series of sensitivity tests were conducted in order to learn the properties of the model; these involved changing only one -- or in the case of industry variables only a few -- of the instrument variables and then comparing the computed endogenous values with the control solution values. After that, "alternative scenarios" for the Atlantic Region were computed and simulations were run using the 1961 averages for Ontario and Canada as the exogenous variables.

Perhaps the most illustrative sets of solutions involve the attempts to analyze how well the model might predict probable developments in the Atlantic Region over the last decade. Outlined below are the control solution values from solving the model, estimates of the 1971 actual values^{3/} for the endogenous variables and the results from two alternative "scenarios" for the Atlantic Region.

^{3/}As noted earlier the full Census material is not yet available. Therefore, 1971 estimates had to be developed. These are based for the most part on National Accounts and Labour Force Survey Data. See pages 168-169 of the text.

ENDOGENOUS VARIABLES	1961 CONTROL SOLUTION	1971 SIMULATION		
		MEAN VALUES	SCENARIO (F)	SCENARIO (H)
EIC	659.61	989.57	775.14	908.03
ER	27.09	30.04	30.13	31.29
ELFR	94.35	95.66	94.69	94.46
W	2,261.6	2,979.0	2,495.7	2,920.8
DII	0.876	--	0.849	0.802
UR	5.65	4.34	5.31	5.54
PR	46.14	46.18	46.79	48.70
SE	31.91	--	17.68	17.51
HO	8.03	--	8.29	7.84
PFELF	21.98	27.58	21.98	26.26
GLFP	109.26	112.98	106.34	107.31
NMR	-5.90	-7.95	-3.49	-2.34
GE	101.30	117.39	103.71	105.35
URB	35.21	42.20	39.21	57.04
LED	20.60	31.04	22.49	30.78
HOBS	10.77	14.77	11.55	14.88
OHVAM	3.10	5.79	3.71	5.98
RS	14.70	15.15	15.84	16.82
EH	8.37	11.00	8.68	10.06
LVAM	8.31	8.44	8.00	7.38
IR	3.76	3.14	10.50	3.18

Scenario (F) simply involved plugging in the 1971 means for the exogenous variables into the 1961 model. This lead to a good "prediction" for the employment ratio, but the predicted values for earnings, and therefore per capita income, were well below the actual 1971 values. Perhaps most importantly,

the solution set is not "feasible" in the sense that the industry residual is much too high.

These results though were far from disappointing for Scenario (F) failed to take account of two important considerations: "overall" productivity gains, i.e. over and above those associated with changes in industry mix, and government policy interventions such as education expenditures and industrial incentives programs. These considerations led to Scenario (H) which involved the 1971 Means for the exogenous variables plus the following "additive" adjustments: (i) 400 was added to the W equation intercept, implying overall productivity gains of about two per cent per year; (ii) 0.5, 1.0 and 0.5 were added to the intercepts in the HOBS, OHVAM and LVAM equations respectively, in order to capture both the direct and indirect effects of incentives programs; (iii) 1.0 was deducted from the intercept in the RS equation to take account of probable productivity improvements in that sector; (iv) 5.0 was added to the intercept in the LED equation to take account of increases in education.

With a few notable exceptions,^{4/} Scenario (H) provides fairly good estimates of the 1971 actual values. Perhaps most important, the solution results for earnings and per capita

^{4/} In particular, the net migration rate and urbanization. These two variables are discussed on page 172 of the text.

income cover two-thirds or more of the distance between the actual 1961 and 1971 values. The model also does fairly well with respect to labour force utilization, with an under-estimate in the "Employed Labour Force Ratio" offsetting an over-estimate in the participation rate. Finally, with the inclusion of the "add" factors, the predictions for the level of education and the five endogenous industries are fairly close to their actual values. From these and other solution results, it was concluded that the model is working fairly well, in that it generally predicts the direction of change and a significant proportion of the required magnitude.

In closing this summary, a listing is provided of the major conclusions that can be drawn from this study. The first two pertain to the econometrics that were applied, while the remainder focus on the economic findings from this analysis.

(1) The estimation results indicate clearly that only the most rigorous econometric techniques will suffice in analyzing income and earnings differentials. Both simple correlation and single equation multiple regression techniques can provide misleading results, indicating that, if at all possible simultaneous equation estimation techniques should be utilized.

(2) Principal components proved to be a very useful analytical device, particularly in pointing out linkages between economic activities, but in the present context, their usefulness in regression analysis was minimal.

(3) The estimation and solution results indicate that:
(i) a cross-sectional model of this type can provide very useful insights into the wage and income determination process

in the Atlantic Region and (ii) a similar model based on 1971 data should prove to be a useful policy tool for the Department of Regional Economic Expansion and other government departments interested in regional development.

(4) Perhaps the most useful property of the model is that its structure and solutions show very clearly the "general equilibrium" nature of the economic relationships underlying the income determination process, and in particular how the alteration in one variable can lead, either directly or indirectly, to important changes in a wide range of other variables.

(5) Probably the model's greatest drawback is that, because it is cross-sectional, its usefulness decreases much more quickly over time, than does a time series model. However some dynamic elements could be added to the present model by: (i) comparing the 1961 and 1971 Atlantic models once the latter is estimated, and/or (ii) comparing the Atlantic model with models estimated for the other regions in Canada (particularly Ontario) and the country as a whole. Such comparisons should provide some useful insights into the future course of the Atlantic economy. It is recommended that the Department of Regional Economic Expansion undertake this work once the requisite 1971 data becomes available.

(6) The estimation and solution results strongly illustrate the all pervasive role played by industrial structure in the generation of earnings and income differentials in the Atlantic Region. Industry mix has a strong impact both directly through the wage equation and indirectly in particular through the manpower utilization equations.

(7) The solutions hint at possible trade-offs between average income and the size distribution of income but a better specification for the degree of income inequality equation -- together with a better measure of income concentration -- would likely be required before the solution results could be viewed as more than illustrative.

(8) Because the data and model are thirteen years out of date, it would probably be a mistake to derive "hard and fast" policy prescriptions from the model discussed here. Still, a number of possible policy implications can be noted:

- strong economic gains can be realized from shifting employment out of the primary industries, but it matters greatly which industries are expanded;

- for example, almost no gains in per capita income accrue from shifting employment into primary manufacturing, while relatively small economic advances are realized from supporting secondary manufacturing of the type already present in the Atlantic;

- fairly substantial income gains are suggested from encouraging certain types of service industries, in particular, transportation, higher order business services and public administration;

- even with no alterations in industrial structure, fairly substantial economic gains can be realized from decreasing underemployment through lowering the degree of seasonality and/or increasing the average number of hours worked.

These and other conclusions are discussed in greater detail on pages 183-195 of the text.

A MODEL OF INCOME DIFFERENTIALS IN THE ATLANTIC REGION

1. INTRODUCTION

In recent years, the problem of regional disparities has become a major policy concern of most developed countries. This is particularly true in Canada where both the federal and provincial levels of government have been very active in the regional development field since the early 1960's. The pace of this activity has accelerated considerably since 1968 when the persistence of disparities in per capita income and economic development between the Atlantic Provinces and Eastern Quebec, on the one hand and the remainder of Canada on the other became a major issue in that year's election campaign. Following that campaign, the Liberal Government established the Department of Regional Economic Expansion. To date, Canada's success in the regional development field must be viewed as mixed at best, but certainly recent financial commitments have been very impressive indeed. DREE's budget for the 1973-74 fiscal year amounts to over 500 million dollars.

It is perhaps surprising however that, with such a high level of government activity, so little rigorous analysis has been done in Canada on the primary or causal factors behind regional inequalities in income and economic development. Probably any of us could provide a 'shopping list' of the

elements associated with below average per capita incomes -- for example, low worker productivity and skills; unfavourable industrial structure and occupational composition, high unemployment and low participation rates, under capitalization, low urbanization and large rural populations, and so on -- but few solid answers have been given to date concerning which factors are primary or causal, and which are secondary, and thus only symptomatic of economic distress. Previous studies, particularly in the Canadian context, have provided some useful insights, but in general have offered government policy makers relatively few guidelines concerning where the emphasis in regional planning should be placed. As a consequence, "seat of the pants" planning has become the norm. The above would suggest that this cannot be totally blamed on the penchant of politicians and government administrators for quick and dirty answers to complex economic problems. Part of the responsibility must be placed at the feet of economic and policy analysts and the academic community generally.

One of the more obvious deficiencies has been the failure to date to apply the more advanced econometric techniques -- multiple regression, principal components, simultaneous equations estimation -- to the study of regional earnings and income differentials. This failing partially reflects the fact that most research on regional disparities and growth was completed between the late 1950's and early 1960's, or before

the full potential of computers for econometric analysis had been developed. One purpose of the present study therefore is to learn what these more sophisticated techniques can tell us about the root causes of income differentials. As a minimum objective, it is hoped that this research will indicate both the strengths and weaknesses of the tools of econometrics in analyzing a major policy concern for Canadian governments.

A further consideration behind the limited use of econometrics has been the lack of comprehensive and consistent sets of data. This problem has largely ruled out the application of these techniques for international comparisons, while for Canada a related limiting factor is that the ten provinces do not provide a sufficient number of observations for cross-sectional econometric study. In the Canadian context then, the only solution is to disaggregate and go below the provincial level. Fortunately, the very extensive county and census division data from the 1961 Census -- both published and unpublished -- provides the opportunity for such analysis, at least on a cross-section basis.

In the case of Canada, the researcher faces another problem, namely, the tremendous diversity between different sections of this country. It was decided therefore to diverge from previous research by focussing on a single (supraprovincial) region within a nation, rather than an entire country or many countries. The hope is that through studying a relatively small

- 4 -

area, 'random' or 'non-economic' elements such as differences in culture, history and topography will play a smaller role in distorting the underlying economic relationships.

Not surprisingly, since the author's concern is with regional disparities and under-development, the Atlantic Region was chosen for this study. The forty-six counties and census divisions of the Atlantic offer a wide range of per capita incomes and economic development. For example, in 1961 the differential in per capita income between the most and least prosperous counties in the Atlantic Provinces was about three times, or well above the differentials of two-fold or less between the high and low income provinces in Canada, and the most and least prosperous states in the United States. This diversity arises however within the context of a fairly common history, topography and set of life-styles. The hope is therefore that income differentials within the Atlantic can be largely explained with reference to economic phenomena.

The original objective of this thesis was to apply single equation estimation techniques, including both simple and multiple regression, in an attempt to explain earnings and per capita income differentials in the Atlantic Region. These differentials were to be analyzed with reference to county (or census division) differences in a wide range of economic variables. However, for reasons to be outlined below, the

1

methodological scope of the exercise was expanded considerably to include, first, principal components analysis and, second, the formulation of a full econometric model. The development, estimation and solution of the model ultimately became the principal objective of the present exercise and thus provides the major focus for the body of this report. At the same time, both the single equation estimation results and the principal components analysis played an important role in the preliminary development of the model and, moreover, are of considerable interest in themselves. Accordingly, the results from both analyses are discussed in some detail in the attached appendices.

The following will briefly discuss how the present study evolved. For the simple and multiple regression analysis, a very large data matrix was originally assembled, with approximately 170 dependent and independent variables. Most of the regression analysis -- as well as the full model -- is based on 1961, but more limited research was also attempted for 1951 and for changes over time between 1951 and 1961. Thus, to some degree, it was possible to test whether the relationships estimated for 1961 also held for an earlier year, and whether the cross-section results are consistent with changes over time. Finally, a few hypotheses were tested for a later year, 1966, using the very limited data provided by the quinquennial census of that year. Admittedly, the 1961 data is out of date, but it

is still the most current available. At the time of writing, the full 1971 Census material was not expected to be published until (at least) mid-year 1974. The hope of the author is that, having demonstrated the feasibility and usefulness of the present approach, the regression equations -- and more importantly, the full econometric model -- can be re-estimated once the 1971 information becomes available. Simple correlation results are outlined in Appendix C, while the (single equation) multiple regression findings are discussed in Appendix D.

The major explanatory variables in the regression analysis are county differences in industrial and occupational structure, as measured by the percentage share of a given industry or occupation of the total 'experienced'² labour force of a given county. Accordingly, a typical regression might have average earnings per wage earner for the 46 counties -- or more precisely, a vector containing average earnings per wage earner for 46 county observations -- as the dependent variable, with the shares of the county labour force in primary industries, manufacturing and certain service industries providing the right hand variables. The 1961 Census offers labour force information for 26 individual industries and 12 occupational categories. As mentioned later, some of these categories were added together where appropriate.

However, all research to date has indicated that only a portion of the variation in earnings and income between regions can be accounted for by industry and occupation composition. As a consequence, other explanatory variables were utilized, including the level of urbanization, unemployment and participation rates, the dependency ratio, the degree of seasonality, average hours worked per week, educational attainment, value added and non-wage added per employee in manufacturing, capital per employee in agriculture (the last three being the only capital or capital proxy measures available), the percentage of females in the employed labour force, per cent Francophone of the total population, and the direction and level of migration over the previous decade. Provincial dummy variables were also employed to see if there are characteristics peculiar to each province that are not picked up by the quantitative variables. Based in part on the multiple regression results, many of these non-industry variables came to play a significant role in the formulation of the econometric model.

Turning to the left hand or dependent variables, our original thinking was that, because of the limitations inherent in each income and earnings series, and because different income and earnings measures possess varying economic implications, a number of independent variables should be looked at. These would

include such productivity measures as average earnings of all wage earners, average male and female earnings, and average income per income recipient in the non-farm sector, together with such 'welfare' measures as average family earnings, income per capita and per family in the non-farm sector, plus employment and personal income per person. The first seven can be taken directly from the 1961 Census, while the last two had to be estimated by the author using Census materials. In the final analysis, most of the regression analysis was conducted using average earnings per wage earner and to a lesser extent employment income per capita as the dependent variables. Also, median earnings per wage earner was utilized for 1961, in order to provide consistent comparisons with the 1951 and 1951-61 regressions. Median earnings is the only earnings or income variable available from the 1951 Census.

Finally, since it can be argued that welfare in a given community depends on not only average income but also the degree of inequality in the size distribution of income, experiments were attempted using a measure of income dispersion as the left hand variable. The measures used here are the inter-quartile range divided by the mean for the earnings and non-farm family income series.^{3/} The objective of this analysis was to ascertain whether or not there might exist 'trade-offs' between the level of income and the distribution of income in a region.

Initially then, single equation regression techniques were applied to analyze county differences in earnings, income and the size distribution of income, using a wide-range of industry and non-industry variables as right hand variables. The first results indicated however that multicollinearity -- and more generally insufficient degrees of freedom -- were causing some difficulties in the regression analysis.^{4/} The decision was made to apply principal components analysis as one possible solution to the multicollinearity problem. Two sets of principal components were derived, one from the matrix of industry variables and a second from the matrix of occupation variables.

It was hoped that the application of principal components would prove to be a useful tool to allow us to more fully utilize the available data and to better explain the underlying economic relationships. It was appreciated however that this technique does have certain limitations. In particular, the economic interpretation to be given to a particular component is sometimes far from clear, reflecting the possible difficulty of associating a component with a variable or set of variables in the matrix of explanatory variables. And even when its economic interpretation is relatively clear-cut, difficulties arise in interpreting the regression coefficient when the component, rather than the

actual variable, is utilized in a regression equation. As Appendix F indicates, both of these problems were encountered in the principal components analysis.

Finally, it was concluded that, as a possible solution to the problem of insufficient degrees of freedom and more importantly to take explicit account of the simultaneous nature of the underlying economic relationships, an attempt would have to be made to formulate and estimate a system of equations -- i.e. a full econometric model -- applying Two Stage Least Squares in estimation. Using 1961 data, such a model was formulated, estimated and successfully solved.

The final model includes 21 equations, of which 18 are behavioural or stochastic. The main dependent variables in the single equation estimation, namely, average earnings, employment income per capita and the degree of inequality in income (i.e. the inter-quartile range divided by the median in the non-farm income series), became the major "endogenous" variables in the model. Further, such variables as the unemployment and participation rates, the degree of seasonality of the work force, the levels of urbanization and education, and a number of industry share variables became "intermediate endogenous" variables, since these can be viewed as endogenous to and thus determined within the economic system. The "genesis" of the model

will be further discussed in Chapter 4 of the text.

The author is not aware of any research where simultaneous equation estimation techniques have been applied to the study of earnings and income differentials on a cross-section basis, particularly at this level of detail.^{5/}

Beyond this introduction, the body of the thesis includes ten further chapters. The second chapter provides a comprehensive and fairly far-ranging review of the literature with respect to earnings and income differentials and differences in the size distribution of income. This section therefore focusses only on the main endogenous variables in the model. Time and space did not allow similar treatment for the remaining endogenous variables; their literary and theoretical antecedents will though be discussed briefly in Chapters 6 to 9 when the "non-income" blocks of the model are outlined.

The review of the literature includes five sub-sections. The first analyzes the many works that have attempted to identify the major factors underlying regional differentials in per-capita income and earnings. Particular attention is given to industrial structure, occupational composition, labour force utilization and urbanization. This analysis therefore is essentially static in nature. In contrast, the second sub-section involves the more dynamic aspects of the regional

growth process and the generation of regional income differentials, emphasizing in particular the export-base-theory associated with North and further developed by Borts and Stein.

Third, there will be a brief discussion of the literature on the size distribution of income. Fourth, given that the study of regional income and earnings differentials is simply a sub-category of labour market economies, a short review will be presented of the current state of that theory, with special emphasis on issues relevant to the present analysis. In the final sub-section, the major implications of the literature will be summarized and a listing provided of the testable hypotheses that the literature offers. The sub-section -- and chapter -- will end with a brief discussion of the basic philosophy and/or approach of the author.

Chapter 3 outlines the methodology applied in this exercise, and in particular the econometric techniques that are utilized. In the fourth chapter, the full econometric model is summarized, noting in particular the variables included in each equation and the linkages between the different blocks of the model. Chapters 5 to 9 discuss each block in turn, outlining the theoretical and literary antecedents for the equation specifications and the actual estimation results. The five blocks are: an income and earnings block, including

the degree of inequality in income; a labour force utilization block; a "growth over time" block; a demographic block; and an industry block.

Chapter 10 outlines in some detail the results derived from solving the model. At the outset, the findings from various "sensitivity tests" are presented; these involve changing only one -- or a few -- variables in order to test the properties of the model. Then, presentations are made of "alternative scenarios" for the Atlantic Region and of "Ontario and Canada simulations". These will be explained later. The final section discusses the conclusions to be derived from the estimation and solution results, and possible areas for further research.

Attached to the body of the report are a series of appendices. The main ones are: a discussion of the theoretical foundations for the wage and income equations in the model (Appendix A); presentations of the single equation estimation and principal components results (Appendices E and F); and a comparison of the original Ordinary Least Squares estimation results with the Two Stage Least Squares results for each behavioural equation in the model (Appendix G). Appendix G also contains most of the major conclusions pertaining to the econometrics applied in this exercise.

2. REVIEW OF THE LITERATURE

(a) Factors Associated with Income and Earnings Differentials

As indicated earlier, this first sub-section will highlight the more "static" factors associated with per capita income and earnings differentials between countries and regions. Particular emphasis will be placed on the work of Chenery, Hanna, Kuznets and Perloff who, in the author's judgement, are the four major contributors in this field. Note will also be made of Canadian contributions, and of studies which emphasize the role of non-industry and non-occupation variables.

Chenery (7), in his 1960 article, attempted to explain differences in the level of industrialization -- and industrial structure generally -- by country by regressing per capita value added on per capita income and total population (i.e. market size), using a log-linear specification. Accordingly, he viewed the level of industrialization and industrial structure as the dependent variable whereas in this analysis it will be used as a right hand argument (and the reverse for per capita income). This point will be taken up later. Chenery utilized cross-section analysis for about 40 countries (varying with the regression) and various industries, both within manufacturing and for broader sectoral groupings. The sample was based on data between 1950 and 1956, depending on the country. While his analysis was cross-sectional, Chenery suggested that his

results possessed implications for changes over time. In this context, he noted the work of Kuznets (to be discussed later) who found the evidence on long-term trends in industrial structure to be in close accord with the evidence provided by the association between international differences in industrial structure and the level of per capita income. Based on this, the present author hopes that his cross-section results will throw out useful insights regarding the regional growth process.

Chenery found a strong association between the level of industrialization (and industrial structure generally) and per capita income, and a somewhat weaker one between industrialization and population. Specifically, the elasticities between per capita income and value added per person were well above one for manufacturing and transportation, slightly above for construction and other services and well below for the primary industries, especially agriculture. In the case of manufacturing, the elasticity between per capita value added and income came to 1.44 compared with 0.47 for agriculture. Further, the elasticity between per capita value added in manufacturing and population size came to 0.20, suggesting at least some relationship between industrialization and market size. All of these coefficients were significant at the usual confidence levels, as were most of the estimated parameters in the remaining regressions. Finally, within manufacturing, the elasticities

between income and value added generally showed the expected tendency, that is higher values for secondary industries than for resource-based manufacturing.

However, as Chenery indicates, some difficulties are attached to the interpretation of the estimated elasticities (or coefficients) since they include both supply and demand effects. Accordingly, Chenery argues that, in looking for explanations for the rise in manufacturing with economic growth, it is necessary to analyze both changes in demand (related for example to Engel's Law) and changes in supply conditions associated with increases in physical and human capital per worker. His conclusion is that the relationship between industrialization and rising income indicates very little concerning the factors causing the rise in income itself. Rather, what his results point out is the pattern of resource allocation that normally accompanies a rise in income. Chenery also notes the complications that arise because of differences in natural resource endowment between countries. Such differences can imply corresponding variations in the human and physical capital required to generate a given level of per capita income. This may be of particular relevance to the Atlantic Region, where it can be argued that differences in the extent and type of natural resource endowments play a major role in producing income differentials within the Region. This is one of the hypotheses that will be tested in the present exercise.

The Chenery article provides a useful example of the application of simple econometric techniques to the explanation of patterns of economic growth. On the other hand, the usefulness of his results are constrained somewhat by the difficulties encountered in interpreting his coefficients. These difficulties should diminish when we move from the international to the inter-regional context (and related to this when per capita income becomes the left hand variable). In inter-regional analysis, differences in demand patterns become much less important; accordingly, the coefficients should largely reflect differences in factor supply conditions, and related to this, industry differences in worker productivity. This point will be returned to later when Kuznets' findings are discussed.^{6/} Further, it is anticipated that, except for highly local-market oriented industries, market size should not play a major role in determining industrial structure, reflecting the relatively free flow of goods and services between regions compared with international flows. In sum, Chenery's study provides some useful insights but his specification differs markedly from the one to be applied here.^{7/}

In an article and later monograph both published in the late fifties, Frank Hanna (24 and 25) analyzed interstate income differentials through the application of standardization techniques.^{8/} Standardization in this case involved

applying national average earnings by industry (or occupation) to the industrial (occupational) structure of each state. The resulting measure is called 'rate constant' earnings, i.e. average earnings in each state standardized for differences between states in earnings within industries.^{9/} Rate constant earnings were then regressed on actual earnings in order to show the extent to which earnings differentials between states are 'explained by' differences in industrial composition. In his monograph, Hanna reported that, in 1949, variations in industrial composition explained 72% of observed differences in state earnings of wage and salary workers, while in the case of occupational composition the proportion was quite a bit higher, namely, 87%. (Similar results were noted in his earlier article.)

Two of his conclusions should be noted here. First, Hanna emphasized that the above proportions reflect only statistical relationships, and should not be interpreted as indicating cause and effect.^{10/} Second, Hanna anticipated that the percentage explained would be higher for occupational composition than for industrial structure, since an industry is not homogeneous with respect to the skills employed in it. However, in the same context he noted the close inter-relationship between industrial and occupational structure, suggesting that industry tends to reflect the current demand for various skills while occupation reflects their current supply.

Hanna's results illustrate clearly the important role played by industrial and occupational structure in explaining state differences in average earnings (which are then transmitted into per capita income differentials). However, the usefulness of his analysis is constrained by the limitations inherent in the standardization technique itself. First, while his results indicate that industry in general is important, they do not indicate which industries are critical in generating earnings differentials between regions. Second, Hanna's industry and occupation results taken together, add up to well over 100% of the state differences in actual earnings, reflecting the impossibility of combining -- and thus interpreting together -- the results of separate standardization computations.^{11/} Both problems suggest strongly that multiple regression and other econometric techniques could be appropriate tools for the study of regional earnings and income differentials.

The work of Kuznets^{12/} and Petloff can be reviewed together, since both used simple pair-wise comparisons to study the association between industry structure and per capita income. More importantly, their findings are generally similar.

Kuznets (34) conducted his initial work in terms of three sectors: the A sector, agriculture, forestry and fishing; the M sector, mining, manufacturing, and construction; and the

S sector, which includes all service activities; but he looked at individual industries as well where data was available. Turning first to his cross-section results based on 1950 data, Kuznets noted a strong negative association between per capita income and the share of employment accounted for by the A sector and a positive relationship for the M sector. The relationship was also positive in the case of the S sector but somewhat weaker than for the M sector.^{13/} This is associated with the smaller dispersion among regions with respect to the S sector shares, which is related in turn to the heterogeneity of the S sector plus the localized character of many sub-sectors within the services grouping.

Within the M sector, the share of employment in manufacturing showed a strong tendency to increase with per capita income, as would be anticipated. Kuznets argued that this positive association reflects not only higher per worker incomes in manufacturing but also the tendency for large manufacturing complexes to give rise to high income service activities. The positive association between income and total manufacturing was a consequence of the strong positive correlation for secondary manufacturing. In sharp contrast, Kuznets discovered a negative correlation between per capita income and the proportion of employment in mainly raw materials oriented industries, reflecting the tendency for such activities to locate in areas where

agriculture and/or mining is a major part of the economic base.^{14/} For the remaining industries in the M sector, income was found to be positively correlated with the share in construction and negatively associated with the share in mining. Finally, most components of the service sector showed a positive correlation with per capita income, although the strength of the association varied from industry to industry.

In analyzing trends over time between 1920 and 1950, Kuznets discovered substantial differences among states in the growth in per capita income. More importantly, he found that the more rapidly growing states showed the sharpest declines in the A share of employment and the greatest increase in the M share. His time trend results therefore were in general accord with his cross-section findings, (although the agreement was less close in the case of the S sector where the change over time findings were mixed.) In short, Kuznets found that the greater the change in industrial structure, the faster the rate of growth in per capita income.^{15/}

Kuznets also briefly analyzed intersectoral differences in participation income (or earnings) per worker. He noted that earnings per worker were generally below the state-wide average in the case of the A sector and well above for the M and S sectors. More important, he discovered that the inequality in income per worker between the agricultural and

non-agricultural sectors was significantly narrower in the high income relative to the less prosperous states. Bringing this together with his previous analysis, the implication is that inequality in income per worker between sectors is positively associated with the share of employment accounted for by the A sector, or alternatively, that the differences between sectors tend to narrow the more "favourable" the industrial structure.

Perloff (48 and 49), in his 1956 article and more particularly in the 1960 text that he co-authored, extended the analysis of Hanna and Kuznets, and clarified many of the issues that they raised. In his analysis of state income and earnings differentials in the United States in 1950, Perloff applied the standardization techniques of Hanna, and computed simple correlation coefficients to test the association between per capita income and industrial structure. His results were generally in accord with the two previous authors and thus will not be detailed here. Still, a few of his specific findings should be noted.

(1) In dividing the tertiary sector between business-oriented and consumer-oriented activities, he found a positive correlation between per capita income and the share of employment in business-oriented services, but no association between income and consumer-oriented activities.

(2) Similar to Kuznets, Perloff noted a positive association between income and secondary manufacturing but found that the correlation between income and primary manufacturing was negative. Both of the industry aggregations outlined in points one and two will be utilized in the present study.

(3) He computed a correlation coefficient of $-.63$ between state per capita income and the percentage of employment in agriculture. While statistically significant, this coefficient still allows for considerable variation in income between states that could be considered agricultural. Perloff's conclusion was that concentration in manufacturing and services was no guarantee of a high level of income.^{16/}

Together with their specific findings, note should also be made of some of the more general conclusions of Kuznets and Perloff:

(1) Kuznets concluded that, for cross-country comparisons, per capita income should be the independent variable (thus making differences in income elasticities the key determinant), but that for regions in a single country, industrial structure should be viewed as the right hand variable. Accordingly, in the inter-regional context, per capita income differentials were to be viewed as largely a function of differences in productivity between sectors. Following his arguments, per capita income and/or earnings per wage earner will generally be used as the dependent or left hand variables in the present analysis. However, as will be discussed later, income might be an appropriate right hand argument in an equation explaining the share of employment in a county in highly local-market oriented service industries, reflecting the high income elasticity of demand for services.

(2) Kuznets noted the wider contrast between industrial structure and per capita income in the intra-national than international comparisons, or stated differently, the coexistence of wide differences in industrial structure with relatively narrow per capita income differentials in the intra-national context. Such wide variations are a function of a well developed transportation system which allows wide differences in the patterns of production and consumption within a given region. The suggestion is then that, because distances are even smaller, an even wider contrast between industrial structure and per capita income should be found in sub-regional analysis.

(3) Kuznets also indicated in his international comparisons that the shift from agriculture towards higher product per worker sectors was a contributing factor in the growth in income

per capita in the countries analyzed. His main conclusion though was that differences between countries in the industrial structure of the labour force is of less importance than intra-sectoral differences in the levels of output per worker in determining inequalities between countries in output -- and thus earnings^{17/} -- per worker. However, it can be shown that, given the negative association between (all sector) output and earnings per worker, and both the share of employment in the A sector and inter-sectoral earnings differentials within countries (as discussed in previous paragraphs), intra-sectoral differences between countries will become narrower as industrial structures become more comparable. In other words, the argument can be made that intra-sectoral differences in earnings per worker are a function in part of differences between countries in industrial structure.^{18/}

(4) Perloff argued -- perhaps somewhat tautologically -- that the association of per capita income with the importance of a given sector in a state's industrial structure arises from the fact that the sector is characterized -- on the average -- by a given level of 'income producing power' relative to other sectors. From his other analysis, it can be surmised that Perloff views 'income producing power' as being mainly a function of the human capital and physical capital to labour ratios, and perhaps the scale economies associated with a given sector.

(5) Similar to the standardization results of Hanna, Perloff reported a strong tendency for states with above-average reported earnings to also possess above-average "rate-constant" earnings. More important to the present exercise, there was also a strong tendency for states with above-average rate constant earnings -- i.e. a favourable industrial structure -- to have reported earnings which exceeded its rate constant earnings, pointing out the importance of intra-industry earnings differentials between states. Based on this, Perloff concluded that earnings differentials are a function of both industrial structure and higher earnings within industries in the more prosperous states. This again illustrates the inter-relationship between industrial structure and intra-industry earnings differentials. Perloff argued that these intra-sectoral differentials could be explained in part by the factor proportions, (or Heckscher-Ohlin) theorem. This theorem would suggest, for example, that in areas where capital is relatively more abundant, entrepreneurs in a given industry would choose a more capital intensive technique, resulting in turn in relatively higher labour productivity and higher earnings.

Implicit here is the assumption that transport costs are not zero and/or that factors are not perfectly mobile, and accordingly factor price equalization between regions is not complete.

(6) In studying rates of income change between 1920 and 1955, Perloff found some decline in the relative variation between states in per capita incomes, (although the state rankings were essentially unchanged.) However, in looking at the farm and non-farm sectors separately, he reported little change in the relative variation among states. The conclusion is that, at this highly aggregated level, convergence in industrial structure -- rather than decreases in intra-industry differentials -- was the crucial factor in bringing about more equal incomes among states in the United States.

7) Perloff suggested that, while income per person employed in a given industry is a function of capital and skill intensity, the size of the capital component is the crucial variable for most industries. This conclusion differs considerably from those of other researchers, and from the longer-term situation hypothesized by neo-classical theory, under which skill levels would be the key determinant of earnings.^{19/}

(8) Finally, Perloff cautioned that the economic sector approach to the study of income differentials required a certain degree of disaggregation and refinement, and that the relationship between income and industrial structure was not best analyzed through the overaggregated classes employed to date. One of Perloff's suggestions was to recombine industries into groupings with essentially common supply and/or demand conditions. This suggestion will be followed -- to the extent allowed by the data -- when grouping industries for the present analysis.

In many respects, the analyses of Kuznets and Perloff provide the starting point for the present study for their work offers a wealth of hypotheses to be tested. Still, their results may be very misleading since they are based only on simple pair-wise correlations or comparisons. This argues

strongly for the application of multiple regression and other econometric techniques in order to focus on the 'partial' relationships and thereby to separate the key industries from those of secondary importance.

Beyond industrial and occupational structure, other considerations have been noted in efforts to analyze income and earnings differentials between regions. Four articles are reviewed in this context. Polzin (51) attempted to explain differences in hourly wages of manufacturing production workers between states, once the effects of industry mix were removed. To accomplish the latter, he weighted the hourly wage of a given industry in each state by the relative national importance of that industry. The resulting measure is called 'industry constant' wages and is the opposite of 'rate constant' earnings that were discussed earlier. 26/ Industry constant wages by state were then used as a dependent variable in an estimated regression that included as explanatory variables the capital/labour ratio, female employment and non-white employment in manufacturing by state, plus the degree of unionization and the concentration of manufacturing employment in urban areas in each state.

Polzin stated that only the capital/labour ratio unambiguously represented a behavioural hypothesis. Traditional theory predicts that wages will be positively related to the capital/labour ratio; underlying this are the two assumptions of constant returns to scale and a 'unique' positive association between wages and the marginal product of labour which assumes in turn profit maximizing behaviour on the part of entrepreneurs. In contrast, the remaining variables are largely structural in nature in that they mainly reflect compositional differences between states. The division of explanatory variables between 'behavioural' and 'structural' provides a useful two-part breakdown and will be applied in the present analysis.

Data for 1958 were used, with regressions being estimated for all states and for the South, North and Midwest states separately. In the all-United States regression, four of the five explanatory variables were statistically significant at the 90% confidence level or better, while the fifth, non-white employment, at least added to the adjusted R^2 . Further, all coefficients possessed their anticipated signs, with the five variables together explaining about 70% of the variation in industry constant hourly wages. In the regional equations, the urbanization variable was significant in all three, while the other four were significant in two each. Among the

explanatory variables, Polzin concludes that the urbanization measure is probably the strongest, while he finds the positive coefficient on the capital/labour ratio particularly noteworthy for it suggests that some of the variation in industry constant wages is associated with differing factor proportions in production across states. The latter situation is consistent with traditional theory. Polzin's analysis is particularly relevant to the present exercise for it points out some of the variables that could be included in a regression once industry mix has been taken into account.

Turning briefly to three other articles, Mansfield (40) found some association between city size and median income in the United States, although he did note that, because of regional effects, two size-classes had to differ substantially before the inter-class differences were large relative to intra-class variations in income. Still, his results suggest that not only urbanization per se but also the composition of the urban population by city-size (for example, metropolitan area, smaller urban centre, resource town) might play a role in generating income differentials between regions.

Logan (37) tested the association between the economic health of a region and its linear distance from the nearest large urban centre. He tested this hypothesis using 1960 data

for 241 counties in Minnesota, Wisconsin and Michigan. Logan found a significant association between the level of urbanization and median income, but little evidence that rural poverty is necessarily a function of distance from a large city. His conclusion was that transformation of physical distance into some form of functional distance might provide more meaningful results, since physically distant areas may be functionally close because of their integration into the wider economic system.

Walsh (64) cast some light on the role of skilled labour in production. He emphasized its 'non-neutrality' in production, related to the ability of skilled labour to better allocate factors between competing uses and to select the 'right' quantity of purchased inputs. Thus, a positive coefficient on a skilled labour variable in a wage determination equation would likely reflect both the payment for higher skills plus these 'second-order' effects which would be expected to lead to increased payments to other factors of production.

The Canadian contributions in this field are relatively meager, being largely limited to the two Economic Council monographs of Chornick and Denton. These will be reviewed in some detail, while brief mention will be made of one other work.

The major finding of Chernick (8) concerned the persistence of disparities in per capita income among the ten provinces in Canada. He reported that, similar to the United States experience highlighted above, there was little change in the rank ordering of provinces between 1927 and 1963. Moreover, using the coefficient of variation as a measure of relative dispersion, income disparities among the ten provinces have been relatively constant over the period, and at the very best only a marginal tendency towards convergence is apparent.^{21/} (This is in contrast to the United States experience where the tendency towards convergence has been fairly marked.) Perhaps most significant is Chernick's finding that the degree of convergence is reduced by more than a half when disparities are measured in terms of earned income per capita rather than personal income per capita. This reflects the growing equality in the inter-regional distribution of per capita government transfers.

Attempts were also made to explain the existence and persistence of regional disparities in Canada. First, it was noted that in 1963 the index of dispersion for earned income per worker was 40% below that for earned income per capita, illustrating the importance of inter-regional differences in the availability and utilization of manpower.

Second, Chernick divided the economy into its agricultural and non-agricultural components in order to evaluate the impact of inter-regional differences in the structure of

employment and income per worker. He concluded that income disparities in Canada would not be altered significantly if the structure of employment in each region were identical with the national average, but if sectoral productivity levels in each region remained unchanged. He assumes therefore, that the two factors are independent. In contrast, it was argued earlier that the two elements are closely inter-related and that one approach -- and perhaps the most reasonable approach -- is to assume that intra-sectoral productivity differentials are at least partially explained by variations in employment structure. This would greatly heighten the significance of employment structure in explaining income differentials between regions. Moreover, as Perloff suggests, it is probably dangerous to analyze the association between income and industrial structure at such a high level of aggregation.

Third, Chernick studied the relationship between income and urbanization and concluded that provincial differences in the extent of urbanization appear to be an important factor in explaining inter-regional income differentials. He failed though to ask the further question concerning the determination of the level of urbanization. One important determinant could be the structure of employment, whose significance Chernick had earlier minimized. The rate of labour force utilization -- a further factor that Chernick finds important -- could also be

viewed as a function of industrial structure. This again points out the danger of analyzing factors separately and alternatively illustrates the necessity for applying multiple regression and other econometric techniques to the study of income differentials.

Together with his inter-regional comparisons, Chernick analyzed the degree of dispersion within the ten provinces and five regions of Canada, using 1961 Census data on non-farm incomes. He discovered a fairly strong tendency for the degree of intra-regional dispersion to be higher in the lower income provinces and regions.^{22/} Thus, to the extent that a region's welfare depends both on average income and its distribution, the position of the Atlantic Region is even more inferior than average income alone would suggest. Chernick's finding further points out the need to study income differentials at the sub-regional and subprovincial level. This is the objective of the present study.

Chernick's work provided the background for the Denton (17) study that followed. Like Chernick, Denton emphasized the wide dispersion in per capita earned incomes among the five regions and ten provinces in Canada. Accordingly, in the 1961-64 period, earned income per person in the Atlantic amounted to only two-thirds of the Canadian average and more

than 40% below the levels in Ontario and British Columbia. (While there has been a slight relative improvement over the last decade, the gap in earned income per person between the Atlantic and the country as a whole remains essentially unchanged today.)

Denton analyzed these differentials in terms of manpower utilization and productivity, that is, average earnings per worker. At the outset, he noted that the smaller size of the employment base accounts for nearly one-half of the earned income differential between the Atlantic Region and total Canada. The Atlantic performance is unfavourable in all four variables that make up the size of the employment base, namely, the age distribution of the population, the labour force participation rate, the unemployment rate and the seasonal variation in employment.

But while the size of the employment base plays an important role, also noteworthy is the fact that, between 1961 and 1964, the average employed person in the Atlantic earned about 20% less than his Canadian counterpart. Denton attempted to account for this and other regional differentials in earned income per person employed by applying standardization techniques to the following variables: industrial and occupational structure, hours worked per week, age distribution, educational

levels, and degree of urbanization.^{23/} He reported that none of these variables, operating on their own, accounted for more than a small proportion of inter-regional earnings differentials in Canada. Further interpretation of these findings proved troublesome, reflecting the usual problem that results from different standardization calculations cannot be added together.

The interaction between income and urbanization was further analyzed through the use of multiple regression. Denton regressed average earnings of male wage-earners on the proportions of the population living in various urban-rural categories, using 1961 Census data for the 238 counties and census divisions in Canada. The coefficients on the categories all possessed their anticipated signs and were significant at the usual confidence levels, but taken together, the urban-rural categories only accounted for about 60% of the variation in average earnings. Following from this, the coefficient of multiple correlation improved considerably to 0.78 when four regional dummy variables (excluding Ontario to prevent perfect multicollinearity) were included in the regression. As would be expected, since the reference point is Ontario, the dummy variables for the Atlantic and Quebec were significantly negative; in contrast, the coefficient for British Columbia was significant and positive while, for the Prairies, the coefficient was not significantly different from zero. Denton

concluded this section by pointing out that, while a substantial association exists between earnings and urbanization, differences in urban-rural distribution may be more a symptom of lower income levels rather than a cause.^{24/}

In summary, the work of Denton and Chernick provides some useful insights into the Canadian experience. Generally speaking, however, the two papers possess fairly severe weaknesses on both theoretical and methodological grounds. The theory or rationale underlying their statistical tests is often far from clear. This likely lies at the root of the questionable nature of a number of their conclusions. The usefulness of their work is further constrained by their failure to apply econometric techniques to the explanation of regional income differentials and, related to this, their failure to take full account of the inter-relationships between key economic variables.

In a 1967 article, Munro (44) reviewed the Economic Council monographs and further developed some of their findings. He argued that, in measuring regional welfare, income per family is preferable to income per capita, for the former takes explicit account of possible economies of scale in consumption, and partially allows for differences among regions in the age-structure of the population. Munro reported that, when measured in per family terms rather than per capita terms, the degree of

inter-regional disparity in Canada is lower and the rate of convergence over the last forty years has been somewhat higher. Following Munro's analysis, the present study will use both income per capita and per family in the regression analysis. Finally, using Gini coefficients, he discovered that the size distribution of income was more equal in the more prosperous regions of the country. Together with Holmes (29), Munro followed this up with the development of a model to explain regional differences in Canada in both per capita income and the size distribution of income. This model will be discussed in a further sub-section.^{25/}

(b) Dynamic Considerations and the Export Base Theory

This sub-section will focus on the work of North and Borts (together with Stein). North developed the export base theory of regional growth and income determination, a theory which has generated considerable controversy in the literature. North's theory will be briefly outlined together with Stabler's synthesis of the export base theory and the stages of growth hypothesis that North attempted to supplant. It is perhaps misleading to associate the broader research of Borts (and Stein) with the export base theory, but certainly many of his major findings -- particularly in the context of

the present analysis-- illustrate the importance of the export base in determining the economic health of a region.

North (41) argued in his 1955 article that the 'stages of growth' (or location) theory of regional growth holds little relevance for the development of regions in North America and other 'New World' economies.. The stages of growth theory focussed on the evolution of regions from a subsistence economy through to industrialization and the final tertiary stage. In contrast, similar to Innis' staple theory for Canada, North concluded that the export base (which could include both commodity producing and service industries) is the critical factor in determining the level of absolute and per capita income in a given region, with residentiary industry playing a purely passive role. Accordingly, the rate of growth of a region is largely determined by the success of its exports, either through relative improvements in the position of existing exports or through the development of new exports. Further, the level and growth of per capita income is critically tied to the character of its exporting industries, particularly their capital and skill intensity and their scale economies.^{26/}

Finally, there is the suggestion that the export base plays an important role in determining the level and type of urbanization in a region.

Under the export base hypothesis, industrialization -- which is generally thought to be the most difficult stage to achieve -- is no longer a necessary prerequisite for continued growth in total and per capita income. He further argues that a strong export base oriented towards primary industries will, through forward linkages and growing incomes, give rise to primary manufacturing and residentiary manufacturing industries. Finally, with maturity comes greater diversity in a region's export base, as foot-loose industries are attracted to the region and as previous residentiary activities become exporters.

At this point, North's export base theory seems to merge with the stages of growth hypothesis. This point was taken up by Stabler (60) in his 1968 article when he offered a synthesis of the two theories. Stabler reasoned that the export staples are the initial engine of growth but that, once initial activity is firmly established, further growth depends for the most part on other factors, some endogenous and some exogenous to the region. These would include population growth, changes in tastes, new discoveries and/or depletions, changes in technology and linkage effects. Also, Stabler agreed with North that evolution to a manufacturing base was neither necessary nor desirable for all regions. Stabler emphasized, however, that the development of a manufacturing base held several advantages related to the forward and backward

linkages associated with manufacturing industry, and the high income elasticity of demand for manufactured goods and personal services.

Borts (4 and 5) studied the interaction between resource earnings and regional growth patterns by state in the United States in his 1960 article and a 1964 text which he co-authored with Stein. He discussed and tested two different hypotheses concerning regional growth: (1) that it is largely a function of resource movements generated by earnings differentials, which are a consequence in turn of initial differences in regional endowments of capital and labour;^{27/} (2) that differentials in resource movement and regional growth patterns are mainly a function of differences in production functions^{28/} or in the demand for a region's exports. He tested these hypotheses by analyzing the changes in capital, labour and wages among the 48 states in three time periods: 1919 to 1929, 1929 to 1948, and 1948 to 1953.

Borts found that the data was consistent with the first hypothesis in only one of the three periods, 1929 to 1948. Accordingly, only in this period did wage differentials appear to be the primary determinant of the marginal efficiency of investment,^{29/} leading to wage convergence among states. Borts notes however that "this period also witnessed an influence

favourable to wage increases which is not likely to be repeated. This was the increase in the marginal product of labour arising from the relative increase in prices in the labour-intensive sectors of growing states".^{29 a/}

For the remaining two periods, Borts found that product demand was the critical factor in explaining relative movements of capital and the increase in wages in different states. As a consequence, there was no wage convergence among states in the 1919 to 1929 and 1948 to 1953 periods. He concludes that his findings provide strong support for a model of regional growth based on the demand for a region's exports. His final conclusion was that wage convergence among states was a function of three elements: (1) continued migration from low to high wage states; (2) elimination of high population reproduction rates in low wages areas; (3) the direction of capital formation to low wage areas. Note was again made of the fact that in two periods these elements interacted to prevent convergence.

Borts findings hold important implications for traditional economic theory, for in two out of three periods, capital movements were not in accord with the predictions of that theory.^{30/} The above points out clearly that, once more dynamic elements are introduced such as differential changes in export demands and/or prices, or differing lags among regions

in the adoption of new technology, the predictive power of 'comparative statics' theory becomes questionable. These considerations might have particular relevance for the Atlantic Region, in explaining both the growth performance of the total Region, and growth patterns and earnings differentials within the Region.

Some of the other conclusions should also be noted briefly, taken mainly from his 1964 text:

(1) Convergence in per capita income among states between 1880 and 1950 can be explained to a large extent by improvements in the allocation of resources within states (i.e. shifts in industrial structure). The conclusion is very pertinent to the Atlantic Region where employment in low productivity industries (particularly primary activities) continues to be very high relative to the remainder of the country.

(2) As states become more comparable in terms of industrial structure, intersectoral wage differentials tend to narrow. This is in accord with previous analyses.³¹

(3) Consistent with neo-classical theory, a positive association was found between the growth in earnings by state and the increase in the capital/labour ratio.

(4) Borts reported a positive correlation between the growth in per capita income and the growth in manufacturing, noting that historically manufacturing has been the most important growing export sector.

It has often been argued that the significance of the export base theory becomes greater the smaller the area being analyzed. This -- plus the past development of the Atlantic economy -- would suggest that the export base theory may play

an important role in explaining growth patterns and income differentials within the Atlantic Region. For the present analysis, this would suggest that in specifying an earnings equation based on industry, the exporting industries should be the first explanatory variables to be included.³²

Still, some of the implications that North drew from the export base theory are of questionable validity, particularly for the Atlantic Region. The role of the residentiary industries is surely not purely passive, for an efficient residentiary sector can be a critical factor in attracting foot-loose industries, or in the development of forward linkages out of the primary sector. Moreover, while the present author would agree that industrialization is not necessary in all cases, the existing export base in many parts of the Atlantic is so weak that industrialization provides the only viable alternative to economic stagnation and depopulation. Related to this, the existing base which is generally oriented towards primary activities, is too weak to allow the development of forward linkages, while low incomes and depopulation prevent the development of residentiary manufacturing activities. It may still be true that, for some sections of the Atlantic, depopulation is the only feasible option but certainly other possibilities should not be neglected.

(c) The Size Distribution of Income

Compared with the analysis of income and earnings differentials, relatively little has been written concerning the economic variables underlying regional differences in the size distribution of income. The following will review the literature that is available and some of the inferences that can be drawn from it. The sub-section will close with a fairly detailed look at the model (mentioned earlier) that was developed to analyse the impact of different variables associated with regional differences in income levels and income distribution in Canada.

Most studies have found a positive association between the level of economic development and the degree of equality in the size distribution of income. Kuznets (35) in his international comparisons reported that the income shares of the upper income groups were distinctly larger in the under-developed than developed countries. 33/

A partial explanation for this relationship lies in the wider intersectoral differentials in product and income per worker in the lesser developed countries and the less prosperous United States regions.

Williamson (66) focussed on the degree of income dispersion among regions in a given country -- or counties within a region -- rather than the size distribution of income, but Canadian experience (and perhaps intuition) would suggest that the two are probably closely related.^{34/} The objective of his study was to analyze how regional inequality varies over the development process. His hypothesis was that at the outset the degree of income inequality among regions would likely increase but that convergence would occur after a certain level of development was achieved.^{35/} He found that the available data was consistent with his hypothesis, that is, that the association between per capita income and regional inequality was in the form of an inverted 'U', with regional disparities reaching a peak in the middle income group of countries. He also noted that, for the United States, county inequality was higher in the lower income than more prosperous states.^{36/}

The general consensus then is that the same economic factors which bring about improvements in average per capita income also work towards a more equal distribution of income. But the exact mechanisms linking average incomes and its size distribution are less clear; related to this, the literature offers few guidelines concerning whether or not these economic factors operate on the two variables with equal force. If not, then trade-offs could arise.

Chiswick (10), in a 1971 article, suggests that inequality in earnings (and thus income) tends to decline with economic development because of smaller inequalities in wealth and schooling, and because economic development brings about more efficient capital markets (including presumably markets in human capital) and thus less inequality in rates of return.

Alternatively, this problem could be approached through the industrial and occupational dimensions. This approach was taken by Reder (52) in a 1955 study of occupational wage differentials (or skill margins) among United States regions. His hypothesis was that differences in skill margins -- and thus for our purposes differences in the distribution of income -- are a function of two factors. Both involve the ease of substituting less skilled for more skilled workers. First, differences in educational levels among regions are a consideration.³⁷ Second, and perhaps more important, Reder argues that skill margins are negatively associated with the number of semi-skilled workers and thus the importance of manufacturing in a region. His reasoning is that labour market competition between skilled and non-skilled workers will be more effective when it is indirect, that is, when it operates through the link provided by a large pool of semi-skilled workers. Accordingly, under Reder's hypothesis, the size distribution of income will tend

to be more equal the greater the employment share of manufacturing and other industries with a large semi-skilled component (such as construction, mining and transportation).

A further variable (mentioned by Williamson (60)) is the labour force participation rate, particularly for women. The expectation is that the participation rate will be positively associated with the degree of equality in the size distribution of income, particularly when measured in terms of family income. A higher participation rate suggests greater opportunities for low income families to supplement their income through female employment. Participation rates should in turn be positively correlated with the employment shares of (mainly service) industries with a high female component. Therefore, the size distribution of income can again be viewed as a function of a region's industry mix.

The above suggests then that the factors underlying average income and income distribution are highly inter-related. Still, the possibility exists that the industry mix that leads to the highest possible average income will differ considerably from that which offers the most equal distribution of income. Some attempt will be made in the present analysis to investigate such possible trade-offs. In this regard, a particularly noteworthy contribution is the recent formulation of a small model by Holmes and Munro (29) to analyze the role of different

variables in generating regional differences in income levels and income distribution in Canada. They constructed a four equation simultaneous model using 1961 data for the 238 Canadian counties and census divisions. The endogenous variables were non-farm mean family income, income concentration (measured by the Gini coefficient),^{38/} the employment ratio (i.e. employment as a proportion of the total population), and investment income as a proportion of total income. The exogenous variables were median years of schooling, the percentage of the labour force in agriculture, the percent unemployed, and median age, together with provincial dummies. The system was estimated using Two Stage Least Squares.^{39/}

The two critical variables in determining average income and the size distribution of income were found to be education and the employment ratio (or the rate of labour force participation), with both having a favourable impact on the two main endogenous variables. Increases in the employment ratio result in a more equal distribution by increasing the number of employed persons in low income families (as discussed earlier), or by raising incomes above the low levels provided by government transfer payments. At the same time, increased levels of education promote a more equal distribution of income by raising the earning capacity of low income groups; the two authors note however that the importance of education

may be over-stated by the exclusion of occupational and other variables.

Finally, their analysis suggests that the goals of higher and more equal incomes may conflict, reflecting the interaction between these two variables and investment income as a percentage of the total. They note that this conflict can be avoided only through a more equal distribution of income bearing assets. In the present analysis, this potential conflict does not become a factor because of the income measure used (i.e. employment income per capita rather than non-farm family income or some other 'overall' measure), but other trade-offs could arise through the use of more disaggregated data. In sum, Holmes and Munro have developed a very interesting little model which, although underspecified, offers a useful point of departure for the joint study of income levels and income distribution.^{40/}

(d) Labour Market Theory

The study of regional differentials in income and earnings is simply a sub-category of the broader field of labour market economics. Labour market economics has generally focussed on earnings differentials between and within industries and occupations, but of course all of these considerations

influence the spatial dimension as well. No attempt will be made here to review the vast literature in the labour market field. Rather, brief mention will be made of a few concerns raised in a recent review by Corina -- together with their implications for the present exercise -- in order to provide some feel for the current 'state of the art'. After that, a review will be offered of one article whose subject matter and methodology is especially relevant to the present analysis. This article is solidly embedded in the 'traditional' school of labour market economics.

One controversy noted by Corina (11) concerned the role of marginal productivity in wage determination. The dispute mainly centres around the relevance of equilibrium analysis and maximizing behaviour to wage determination. The antagonists argue that marginal productivity is largely irrelevant and that the institutional setting is the critical factor in the determination of wages.

The implications of this controversy for the present exercise are obvious. To the extent that marginal productivity theory no longer holds, it will not be possible to predict the sign or size of an industry coefficient in a wage determination equation with reference to known characteristics of the industry's production function. Instead, reference would have

to be made to the degree of monopoly or monopsony power in the relevant factor and product markets.

Moreover, to the extent that the institutional setting within an industry varies greatly from county to county, the predictive power of industrial structure in this type of analysis would be greatly diminished. The critical determinant of wages would likely be the 'average' institutional setting in a given county, together with some 'disequilibrium' conditions such as the degree of slack in the labour market. In this study, the initial assumption will be that there is at least some link between industry earnings and the industry's production function and therefore that wages are determined at least in part by marginal productivity or some variant such as average value productivity.^{41/} The possibility of other determinants however will not be ignored although the direct test of the impact of institutional factors is impeded by data constraints.

A second controversy noted by Corina involves the relevance of the competitive hypothesis^{42/} over the longer term relative to the short run. Many authors have argued that, while institutional factors are dominant over the short run, the competitive hypothesis is in general a good predictor of longer term trends in relative earnings.^{43/} Corina concluded

that "since satisfactory evidence of long run competition cannot be derived directly, indirect relationships between relative wages and selected economic variables have been usually invoked to provide a second-order battery of tests, but even here (on current evidence) there appear to be no firmly established or consistent relationships".^{44/}

Compared with the previous dispute, the implications of this controversy for the present analysis are far from obvious. The present study is cross-section in nature, providing a 'snap-shot' of the Atlantic economy at a single point in time. As a consequence, the expectation is that the results will be a mixture of shorter and longer term effects. One implication is that, to the extent that Reder et al are correct and the competitive hypothesis does hold over the longer term, at least some of these findings should be consistent with that hypothesis. Unfortunately, we cannot say a priori which results will be for the most part a reflection of longer term effects.

It seems clear from the Corina review that labour market economics is in considerable disarray at the present time. He further notes however that this condition is in turn largely a function of problems within the economics discipline generally, for example the growing doubt concerning the

relevance of marginalist assumptions to economic decision-making. Perhaps his most important conclusion is that a labour market model cannot hope to approach completeness unless it includes both institutional and economic (or traditional) elements. This point was kept in mind in conducting the present research, although the direct test of either set of elements, which would require their complete separation, was not possible in the present context.

The article reviewed below provides a useful look at the 'traditional' approach in labour economics. Scully (58) used multiple regression analysis to analyze inter-state wage differentials of production workers in different United States manufacturing industries for the year 1958. His specific objective was to focus on the classic case of the large and persistent earnings differentials between northern and southern workers. The inspiration for Scully's study was provided by previous work which found that, even after regional differences in industry mix were accounted for, a sizeable portion of the overall North-South differential still remained to be explained.

Scully put forward the hypothesis that the failure to achieve North-South wage uniformity could result from three major sources: (1) barriers to the free flow of resources among regions, resulting in differing factor proportions;

(2) non-homogeneity of the work force, in particular, interstate variations in the amount of human educational capital embodied in workers; (3) institutional factors. Accordingly, he relates within industry wage differentials among states to five economic factors: (1) variations in the capital-labour ratio; (2) variations in human educational capital; (3) interstate differences in union activity; (4) variations in the per cent of non-white production workers; (5) within industry variations among states in the percentage of employment accounted for by female production workers. The expected sign of the regression coefficient was positive for the first three explanatory variables and negative for the last two. A North-South dummy variable was also employed.

In testing the above, regression equations were estimated for 17 separate manufacturing industries by regressing average hourly earnings on the explanatory variables in each industry. The American states were utilized as observations, with the number of states varying from industry to industry. At the outset, the equations were estimated using all independent variables, but afterwards variables were excluded when insignificant, or were combined -- in particular, human and physical capital, and non-white and female labour -- when 'harmful' multicollinearity became a problem. The coefficient of multiple

correlation was significant in 14 of 17 industry equations, suggesting that in general his basic hypothesis could go a long way towards explaining inter-state differentials in industry earnings. Moreover, all of the coefficients possessed their anticipated signs. Among the explanatory variables, the strongest was the human capital to labour ratio, while the unionization variable was next in line. In contrast, the capital/labour ratio was generally not too important, while the results for the non-white and female variables were mixed and clouded by the multicollinearity problem.

Scully also pooled his observations in order to run separate regressions for the low, medium and high wage industries. (Dummy variables were included for individual industries in light of the different intercept values in the separate industry equations.) In this case, his results were more clear-cut -- reflecting in part the much larger sample size -- but in general the findings did not differ greatly from the results for the individual industry equations. The human capital variable remained the most significant factor in explaining inter-state wage differentials, but the other independent variables were also significant.

Finally, Scully partitioned all of his data into northern and southern groupings and then ran two new regressions in order to directly analyze the North-South wage differential.

Once the equations were estimated, he was able to identify the sources of the North-South differential by inserting the mean values of the independent variables into the regression equations. He discovered that, outside of the effects of the industry dummies which were unfavourable to the South,^{45/} the North-South differential was largely explained by the human capital variable and the non-white variable, while unionization played a very subordinate role. In sharp contrast, the net effect of the capital/labour ratio (i.e. other things being equal) was to create a wage differential in favour of the South, suggesting that factor proportions have adjusted in the right direction (and perhaps even overadjusted).

His major conclusion was that it may prove very difficult to obtain complete regional wage uniformity. For example, he notes with respect to human capital: "...problems arise due to this factor's being embodied in the labour-input to the production process. Thus, for human capital to flow from the North to the South, labour must flow, but the combined average reward to the two factors does not encourage this."^{46/} Moreover, Scully cites evidence that the movement of non-whites from the South to the North is on the decline, implying that this factor cannot be counted on now to play a major role in eliminating the wage differential. Scully concludes that, at the present time, the North-South wage differential is largely

a non-economic phenomenon and that political and social tools must be looked to for its elimination.

The focus of the Scully study -- with its major emphasis on intra-industry differentials -- differs considerably from that of the present exercise. Even so, both in terms of content and methodology, this is perhaps the most important contribution to date in relation to the present exercise, and thus has been given greater weight in this review than any other piece of analysis. Four factors underlie this:

(1) Scully's work illustrates better than any other the usefulness of cross-section analysis for the study of wage determination and differentials.

(2) His study indicates the type of independent variables that ideally should be included in a regression equation once the explanatory power of industry composition has been exhausted. (Unfortunately, we have no unionization variable and the proxy variables for the others may be weak.)

(3) Much of Scully's methodology will be adopted for this study, both in terms of overall strategy and particular techniques. For example, the Farrar-Glauber equations were used to test for multicollinearity, and variables were reformulated (e.g. through adding similar industries together) in order to minimize the multicollinearity problem.

(4) The general conclusions that emerge from his work are pertinent to the present exercise. In a limited sense, his results seem to support the proponents of the competitive hypothesis, for the human capital/labour ratio was consistently the strongest independent variable.⁴⁷ However, a broader interpretation of Scully's work would likely confirm Corina's conclusion that both competitive and institutional forces should be taken into account in the study of wage determination.

As the bibliography indicates, other studies in the labour market field are being utilized by the author. Lack of space prevents their review here, but sources will be given when the findings and conclusions of these studies are employed explicitly in the present analysis.

(e) Summary and Testable Hypotheses

The above review of the literature has touched on a broad range of topics, rather than being limited to the generally 'static' question of the association between per capita income and other socio-economic variables. In particular, attention has often been focussed on the more dynamic aspects of the regional growth process and regional income determination. Admittedly, it is questionable to what extent these more dynamic forces will be picked up in a cross-section model, (although many inferences have already been made concerning their possible impacts.) The author would argue however that the cross-section results can be properly interpreted only in the context of a dynamic regional development model. It is hoped that the analysis to follow will justify the length and broad nature of the present section.

No attempt will be made here to summarize the large and diverse literature reviewed above. Instead, a brief review will be presented of the major factors that prevent income and earnings equalization between regions. Then, a listing will be offered of the various testable hypotheses that can be derived from the above literature. In this manner, a large proportion of the issues raised earlier should be touched on in this concluding sub-section.

The existence and persistence of income and earnings differentials have always been something of an embarrassment to traditional economists, for they compromise many of the fundamental tenets of competitive theory, namely, that markets equilibrate, factors of production move in response to market signals, and economic decisions are based on maximizing behaviour. Thus, a great deal of the literature has been taken up with attempting to justify the existence of such differentials in the context of a 'competitive' economy.^{48/} Yet, this embarrassment and the resulting justifications seem odd in light of the stringency of the conditions that must be met in order to bring about factor price equalization among regions. Factor prices -- that is, wages and the rental cost of capital -- will not equilibrate spatially as long as at least one of the following circumstances are present:^{49/}

- (1) transportation and information costs between regions;
- (2) economies and diseconomies of scale in production, and related to this, the existence of agglomeration economies;
- (3) pecuniary considerations such as differences in living costs among regions;
- (4) differences in non-pecuniary advantages associated with the relative attractiveness of communities and industries;
- (5) fixed factors of production, such as natural resource endowment or location (e.g. a port site or an appropriate location for a military base);
- (6) institutional differences between regions, for example, in the extent of monopoly power in product and factor markets;
- (7) regional differences in the time lags with respect to the adoption of new technology, which is related in turn to the broader question of variations among regions in the quality of entrepreneurship;
- (8) changing demands for the export goods of different regions.

But even if factor price equalization is approximated, this does not imply equality among regions in per capita incomes. This could reflect any one of the following considerations:

- (a) unequal distribution of property income -- returns may be equalized, but regional differences exist in the per capita volume of assets owned;

- (b) differences in labour force participation rates and age structure, reflecting perhaps only regional variations in life-styles;^{50/}
- (c) regional differences in the stock of human capital -- again, there may be price equalization for a given skill group but variations among regions in the per capita stock of human capital. A further consideration is different lags among regions in the acquisition of new skills.

Most or all of these factors prevail to some degree within the Atlantic Region. Of particular relevance are considerations (5) and (c) taken together. Natural resources -- viewed in the broadest sense to include location -- play a very important role in the Atlantic economy. But the quantity and type of natural resource endowment varies tremendously among different parts of the Region.^{51/} Being a fixed factor, the existence of natural resources alone is sufficient to prevent factor price equalization. More important, however, production functions -- and particularly skill requirements -- vary greatly between different resource industries. As a consequence, the 'stock' of human capital would likely be relatively small in areas where fishing -- with low skill requirements -- is the dominant activity, but would be much higher in areas such as Halifax and Saint John where port-related activities are an important part of the economic base. As noted above, regional variations in the per capita stock of human capital -- mainly

associated in the Atlantic with county differences in natural resource endowments -- can lead to substantial inter-regional differentials in per capita income, even when the prices for given skill groups are equalized.

A further consideration is differences among Atlantic counties in institutional setting, especially with respect to the unionization of the work force. In some parts of the Atlantic, the work force is highly unionized, particularly those areas where pulp and paper, mining and smelting and refining are important industries. In other areas, for example, where fishing is dominant, unionization was almost non-existent in 1961. Since much of this unionization is industry specific, some of these effects will likely be reflected in the regression coefficients on individual industries, although it will be impossible to separate these effects from other forces.

To close out this section, a partial listing is provided of the testable hypotheses offered by the literature concerning the generation of earnings and per capita income differentials among regions. (The relevant authors are noted in parentheses, where appropriate.)

(1) Industrial and occupational distributions will likely be the critical determinants but cannot be expected to explain all of the inter-regional variation in earnings and incomes. (Hanna and Polzin).

(2) The literature (mainly Kuznets and Perloff) indicates that earnings and incomes will be negatively related to the percentage of employment in the primary industries and primary manufacturing, and positively associated with the importance of secondary manufacturing and most service activities.

(3) Among industries, the exporting activities will perhaps play the primary role in determining earnings and per capita income in a given area, and thus should probably receive first priority in equation specification; but account must also be taken of mainly residentiary activities (North, Borts and Stabler).

(4) Urbanization should be positively related to earnings and incomes, reflecting both higher living costs and higher productivity associated with agglomeration economies (Denton, Richardson, Polzin and Mansfield).

(5) The rate of labour force utilization -- as measured by the unemployment and participation rate, and the age structure of the population -- will have a strong and direct impact on per capita incomes. Moreover, as measures of the degree of disequilibrium in labour markets, these variables may also influence average earnings in a given area (Denton, Chernick and Reder^{52/}).

(6) The degree of seasonality in the work force together with average hours worked will also affect earnings -- particularly annual earnings -- and per capita income through their impact on the size of the employment base (Denton).

(7) Age structure may have a further influence on earnings and income, reflecting the possible negative association between skill levels and the average age of the work force (Reder^{52/}).

(8) The level of education or skills should have a positive impact on earnings and income, even once industry composition is taken into account, resulting from both its 'non-neutrality in production', plus the choice of more or less skill intensive production processes within industries depending on the relative abundance of skilled labour in a given area (Walsh and Scully).

(9) Besides the level of education, relative skills could also be picked up by occupational variables, such as the share of employment in the professional and technical groups, or in the white-collar groups generally (Hanna).

(10) Similar to education, capital per worker -- given the availability of the requisite data -- could also pick up important within industry differences among regions related again to relative factor endowments (Scully and Polzin).

(11) Ethnicity could play a role in generating earnings and income differentials between regions, resulting perhaps from discrimination or differences in life-styles among cultural groups (Report of the Royal Commission on Bilingualism and Biculturalism).

(12) The direction and amount of migration over the preceding number of years could be an important explanatory variable, but the anticipated sign of the coefficient is unclear for it depends on whether migration is equilibrating or disequilibrating (Richardson and Vanderkamp⁵³).

(13) A further dynamic factor could be differential changes in demand among the export industries of differing parts of the Atlantic. Other things being equal, average earnings would be expected to be lower in a county whose export industries had experienced unfavourable changes in demand over the preceding period. This may further help to explain the sign and size of different industry coefficients in a wage determination equation (North and Borts).

A few hypotheses can also be gleaned concerning the elements underlying the size distribution of income:

(a) The "consensus" of the literature is that the distribution of income will likely be more equal the higher the level of economic development in a given area, as measured by the share of employment in secondary manufacturing or other more modern industries, or by the occupational distribution, or more directly by per capita income itself (Kuznets, Williamson, and Reder).

(b) The distribution will tend to be more equal the lower the unemployment rate and the higher the labour force participation rate. It should also be related to the age structure of the population (Williamson, Munro and Holmes).

As the above hopefully suggests, the literature may be relatively weak with respect to the 'causal' factors underlying regional income differentials, but, in terms of testable hypotheses, it has provided a firm foundation for the present analysis.

Before closing this section, this might be an appropriate place to outline explicitly the "philosophy" or basic approach of the author. One theme flowing through the review of the literature was that the industrial dimension provides a useful starting point for the study of regional differences in per capita income and the size distribution of income. The author is not arguing that industrial structure is the only causative factor, for it is clearly inter-related with a large number of other variables, such as the rate of labour force utilization, the level of urbanization, the skills and attitudes of the work force, and productivity differentials between and within industries. Clearly then, we are working with a 'general equilibrium' type of model, with all of the usual attendant problems of separating cause from effect. The author's basic position is, however, that in the majority

of cases it is more reasonable to view industrial structure as a 'cause' than a symptom. Or at the very least, the industrial dimension often represents the easiest or most appropriate vehicle for breaking into the 'circle of causation'.

This approach is closely linked with the perspective of the regional planner. When the planner arrays all of the various elements associated with economic distress, industrial structure probably stands out as the variable most easily altered through policy action. (Accordingly, industrial incentives and other 'job creation' programs have been a major part of the regional development efforts of this and other countries.) Of course, other elements also have to be taken into account, such as the availability of skills and the quality of entrepreneurship. But the regional planner would probably view the encouragement of shifts in industrial structure -- implying alterations in the demands for labour -- as an appropriate starting point for policy action. He would then expect -- or at least hope -- that the encouragement of a more 'favourable' industrial structure would lead to other favourable changes, such as improvements in the rate of labour force utilization and reductions in intersectoral productivity differentials (with attendant improvements on the size distribution of income).

To this extent, the perspective of the regional planner will be adopted for this exercise, with analysis focussing first on the industrial dimension and then going on to other variables. It should be emphasized moreover that this approach is consistent with traditional 'marginal productivity' theory which focusses on differences in the amounts and types of the demands for and supplies of labour. A review of the traditional theory and its implications for the present study is offered in Appendix A.

The basic approach of the author should become clearer when the econometric model is summarized in Chapter 4 and the five blocks are discussed in Chapters 5 to 9. Many of the issues raised in the review of the literature will be noted again when the underlying rationale and the estimation results for each block are discussed.^{54/} In that context, emphasis will be placed on how the various hypotheses offered by the literature were empirically tested.

3. METHODOLOGY

The following section will briefly outline the methodology applied in this exercise, with particular emphasis on the econometric techniques. The background for this section is provided in part by Appendix A which discusses the traditional theory underlying the present work -- in particular the equations in the income and earnings block -- and how that theory could be tested directly if the requisite data were available.

As indicated in Appendix A, severe data constraints prevent the direct testing of the theoretical framework developed there. The only available information is (all-industry) average earnings by county, and employment (and thus employment shares) by county, for 26 industry groups. This information does however allow us to approximate the third procedure discussed in the appendix. This can be done by regressing actual average earnings by county on (a part of) ⁵⁵ the employment shares matrix in equation 2 (Appendix A). The estimated coefficients, together with the constant term, would take the place of "competitive wages derived from the industry production functions. And similar to Appendix A, once a proper specification is developed using industry variables alone, further variables can be added to the regression -- urbanization, the rate of labour force utilization, the level of

skills, etc. -- in an effort to explain the remaining variation in earnings. Therefore, initial thinking was that a typical specification would take the following form:

$$(9) \quad W = a + \beta_1 \text{ PRIM} + \beta_2 \text{ LVAM} + \beta_3 \text{ HVAM} + \beta_4 \text{ PA} + \beta_5 \text{ URB} + \beta_6 \text{ ER} + \beta_7 \text{ LED}$$

where

W	=	Average earnings (\$)
PRIM	=	Primary industries
LVAM	=	Low value added manufacturing
HVAM	=	High value added manufacturing
PA	=	Public administration and defence
URB	=	Urban population as % of total population
ER	=	Employment ratio (i.e. employment as % of total population)
LED	=	Proportion of the population five years and over not attending school with at least two years of secondary education.

and where all of the industry variables are industry shares, that is, employment in a given industry as a percentage of total county employment. The various specifications for a wage determination equation are outlined in Appendix E.

The upshot is therefore, that the fairly rigorous theoretical framework discussed in Appendix A could not be tested directly with presently available data. The framework did though provide a useful background for the empirical analysis to follow, particularly for the critical income and earnings block. Also it did offer certain inferences which proved useful in

further work. First, the theory gives further evidence concerning the expected signs for the coefficients on individual industries. Together with previous empirical work, the theoretical framework suggests significantly positive coefficients for those industries that are highly capital and/or skill intensive. Second, and perhaps more tentatively, the above framework suggests that, to the extent that a properly specified industry equation totally exhausts the variation in earnings, some kind of 'equilibrium' wage has likely emerged in each of the different industries in the Atlantic Region. This is not to argue that factors are being paid in accordance with marginal productivity theory but simply that intra-industry differentials between counties are relatively small, and thus the inclusion of urbanization, unemployment rates, level of education and so on adds little to the explanation of the variation in the dependent variable. It should be emphasized that this second inference is offered very tentatively and may not bear up under closer scrutiny.

The above was suggestive of one possible multiple regression strategy which was in fact applied. First, an effort was made to specify as complete a wage determination equation as possible using industrial structure alone. Then, a series of regression equations were estimated adding only one further explanatory variable to each equation. This served to point

out which industry variables are simply acting for another explanatory variable and thus the extent of specification error from depending on industrial structure alone. After this, a 'complete' wage determination equation was developed, using both industrial and other explanatory variables. Once a wage determination equation was estimated, similar regressions were run using different dependent variables and at the outset the same matrix of explanatory variables; then appropriate adjustments were made to the latter. As indicated earlier, the estimation results from applying this methodology are fully outlined in Appendix E. With respect to both the development of the underlying rationale and empirical testing, more effort was expended on the wage determination equation than any other equation in the model. This reflects the critical role played by the wage equation, for it both provided the starting point for the development of the model and to some extent established the model's dimensions.

The single equation regression analysis -- using income and earnings as the dependent variables -- offered some useful insights into the wage and income determination process in the Atlantic Region. However, as discussed in the Introduction, the problem of multicollinearity -- and more generally that of insufficient degrees of freedom -- made it difficult to estimate the 'true' underlying relationships and prevented

25

OF/DE

us from fully utilizing the large data matrix that had been assembled. The technique of principal components seemed to offer one possible solution to these two problems. This technique essentially involves the extraction of a series of 'orthogonal' -- i.e. totally non-collinear -- vectors from the matrix of explanatory variables. To quote from Johnston (32): "The nature of principal components may be approached in a number of ways. One is to ask how many dimensions or how much independence there really is in the set of K variables. More explicitly, we consider the transformation of the X's to a new set of variables which will be pairwise uncorrelated and of which the first will have the maximum possible variance, the second the maximum possible variance among those uncorrelated with the first, and so forth."^{56/} The most important principal components can then be used as the independent variables in a regression equation. The derivation of the principal components is further discussed at the beginning of Appendix F. As indicated there, when the explanatory variables are for example, industry shares, the "weights" used to compute a given principal component indicate the importance of a certain industry in the composition of that component. The make-up -- and thus the possible economic interpretation -- of a given principal component can be further discerned by calculating pair-wise correlations between each of

/...

the principal components and each explanatory variable. Both the "weights" and the correlation coefficients are provided when the industry and occupation components results are presented in Appendix F.

In summary, principal components have the 'nice' property of being 'orthogonal', that is, totally non-collinear. Moreover, they are a useful analytical device for pointing out those industries, occupations and so on that tend to 'cluster' together, thus indicating which counties are similar in economic structure and why they are similar. Finally, they were used here to provide further evidence of the industries and other variables that should be included in the regression equations and the econometric model.

On the other hand, the economic interpretation to be given to a principal component is sometimes far from clear, even after correlation analysis. Moreover, since a principal component is a 'constructed' or 'artificial' variable, its usefulness when applied in a regional planning context is limited. If at all possible, actual values and variables should be utilized in attempting to estimate the possible impact of a given policy change. Still, multicollinearity and the more general problem of insufficient degrees of freedom prevented us from including all of the relevant variables in a

single regression equation. Furthermore, and this is probably the critical consideration, single equation estimation using Ordinary Least Square fails to take account of the underlying interdependencies or simultaneity among the key economic variables, leading to the serious econometric problem of OLS 'bias' in the estimated coefficients. This brings us to what became the major objective of the present exercise, the development of a simultaneous equation model to explain income and earnings differentials in the Atlantic Region.

In developing an econometric model of this type, three related problems must be considered at the outset, the problem of identification, the type of estimation technique to be applied, and the matter of solution. First, an econometric model can be properly estimated and solved only if every equation in the model is either 'just identified' or 'over identified', or, in more layman's terms, if each equation is sufficiently different from every other equation in the model. A necessary condition for identification (either 'just' or 'over') is that the number of predetermined variables^{57/} that are excluded from an equation must be at least as great as the number of endogenous variables included less one. Or, as an alternative formulation of the same condition, the number of endogenous and exogenous variables excluded from a given equation must be at least as great as the number of equations in the model less one. This is the

so-called order condition for identifiability. There is also the rank condition for identification that must also be met but, in most cases, analysis of the order condition is all that is attempted.^{58/}

The question of the type of technique to be applied arises because (1) Ordinary Least Squares is no longer the appropriate tool, and (2) because of the relationship between this question and the matter of identification. OLS cannot be applied because the error term and the endogenous variable are correlated when the latter appears as a right-hand variable in an equation. This violates one of the fundamental assumptions underlying the OLS technique, that the error term and the right hand variables be uncorrelated. When OLS is applied to a simultaneous system, the estimated coefficients are not only biased in small samples but also inconsistent (or biased) in the case of large samples.^{59/}

Accordingly, simultaneous equation estimation techniques should be applied. The type of technique depends on whether the equations in the system are 'just' or 'over' identified. If they are just identified, that is, if the number of exogenous variables excluded from an equation, just equals the number of endogenous variables less one, Indirect Least Squares can be used.^{60/} However, in the final specification outlined in the following

sections, every equation in the model is over identified; therefore, another estimation technique had to be applied. The one utilized here was Two Stage Least Squares or 2SLS (which is the technique that is generally used for simultaneous systems).^{61/} The coefficients acquired from 2SLS are biased for small samples but they are consistent -- that is, they tend to converge towards their 'true' values -- for larger samples.^{62/}

Third, once the model has been properly specified and estimated, we must turn to the final question of solving the system of simultaneous equations. This is necessary in order to use the model, either for forecasting or, more likely in the case of a cross-section model, for simulation purposes involving the development of 'alternative scenarios'. Solution basically concerns solving for the endogenous variables in the model, given values for the predetermined or exogenous variables over the simulation period. This is a relatively easy task with a ~~totally linear~~ model.^{63/}

However, in its final form, the present model is non-linear. This results from the inclusion of employment income per capita as a dependent variable, which is a multiplicative function of earnings (i.e., employment income per person employed) and the employment ratio (employment as a proportion of the

total population). Moreover, the employment ratio is a multiplicative function of three labour force utilization variables.^{64/} With a non-linear system an iterative technique must be applied to solve the model, the most popular being the Gauss-Siedel method which is available in library computer programs. The software package used here will be noted later in this section.

Three problems can arise in solving a non-linear model. First, as with linear models, the solution may not make economic sense. This problem can be partially solved during the estimation stage through ensuring that the sign and size of the estimated coefficients are reasonable (although this does not ensure that the problem will not arise). Second, the solution may not converge, or may converge very slowly. Finally, there may be multiple solutions; in essence, the system reduces to a general polynomial in the endogenous variable(s) and accordingly there will be as many solutions as the order of the polynomial.^{65/} As outlined in Chapter 10, only the first problem arose in solving the final specification of the model, that is, solution values were always secured, but these values were not always feasible in an economic sense.

Next, a few remarks should be made concerning regression strategy:

(1) While Two Stage Least Squares was used to estimate the complete model, OLS was applied to each equation initially to gain some insights into the underlying economic relationships. This results from the fact that for some of these equations, particularly outside of the income and earnings block, the literature provides relatively few guidelines concerning which right-hand variables should be included in a regression, and more generally, which exogenous variables should be included in the complete model.^{66/} As implied above, the exogenous variables must be known before 2SLS can be applied. In Appendix G, the results from OLS are compared with those from 2SLS in order to indicate the extent of bias from not applying simultaneous-equation estimation techniques.

(2) As a general rule of thumb, a variable was retained in an equation as long as three conditions are met: (i) the sign is 'correct', that is, it is the expected sign based on economic theory or known relationships in the Atlantic Region; (ii) the size of the coefficients is 'reasonable'; (iii) the t-value -- i.e. coefficient divided by its standard error -- is greater than one, which indicates that the coefficient is adding to the adjusted multiple correlation coefficient (that is, adjusted for degrees of freedom). With 40 degrees of freedom, the t-value at the 95% confidence level is 1.684 for a one-tail test -- with a one-tail test being applied when the sign of the coefficient is known -- and 2.021 for a two-tail test.^{67/} These conventional guidelines will be relaxed here, reflecting in part the probable multicollinearity in the system.^{68/} In general, however, the first two conditions must be met before the third became operative. Moreover, once or twice, when the author felt strongly enough about the sign and size of a coefficient, the variable was retained in the equation even when the t-statistic was slightly below one. This was particularly true for intermediate specifications of an equation, for example, for an OLS specification for an equation in the model.

(3) In attempting to specify equations in the model other than the wage and income equations, two initial estimations were made, one based on the author's 'best first guess' of the possible specification, and the second based on the industries included in the final wage determination equation that was estimated previously. Comparing these two regressions provided a useful starting point for the specification of the non-wage equations in the model.

Further, some of the problems related to multiple regression analysis should be reviewed here, together with their implications for the present exercise and their possible solution:

(1) The problem of multicollinearity has already been noted in some detail.^{69/} Possible solutions to multicollinearity include principal components -- discussed in Appendix F -- and applying transformations to the data. In the present exercise, the latter solution was attempted by adding together the industry shares of two industries that are highly collinear. This is similar to the application of 'restricted' Least Squares since it involves restricting the coefficients on the two industries to having the same values.^{70/}

(2) A second problem in multiple regression is that of auto -- or serial -- correlation which relates to the association between the error terms.^{71/} When auto-correlation is present in the system and OLS is applied, the standard errors of the coefficients tend to be under estimated. Since this is a cross-section model, the problems with auto-correlation should be minimal, although geographic and other similarities between contiguous counties could be a factor.^{72/}

(3) A further problem is that of heteroscedasticity which violates the OLS assumption of constant variance for the error term.^{73/} The consequence of heteroscedasticity is to increase the size of the standard errors of the coefficients; this might offer further justification for accepting coefficients with t-values less than the usual critical levels as long as their sign and size seem reasonable.

(4) There may be simultaneity in the system. The solution to this involves the development of a full model and the application of some type of simultaneous-equation estimation technique. This, of course, has become the focal point for the present exercise. The remainder of the text will be taken up with analyzing the results from specifying, estimating and solving a simultaneous system to explain income and earnings differentials in the Atlantic Region.

Finally, note should be made of the computer software packages utilized in the present exercise. For the initial OLS specifications, the Massager program was used, while the APL programs of I.P. Sharp were used to complete the OLS estimations and to estimate the initial specification of the model. However, the APL programs do not include a complete Two Stage Least Squares estimation package^{74/} or an algorithm for solving non-linear systems. Therefore, the TROLL package -- standing for Time-Shared Reactive On-Line Laboratory -- of the National Bureau of Economic Research was utilized in the final estimation and solution of the econometric model. TROLL allows an interactive approach in estimating a model and provides both OLS and 2SLS estimation results. Then, once a correct specification is acquired, the package allows the user to go directly to solution, using a Gauss-Seidel iterative technique.

4. SUMMARY OF AN ECONOMETRIC MODEL OF INCOME DIFFERENTIALS
IN THE ATLANTIC REGION

Before summarizing the model, it might be useful to discuss how the model evolved. As indicated previously, the original objective of this exercise was more limited in scope, involving in the main the application of single-equation estimation techniques to the analysis of earnings and income differentials in the Atlantic Region. However, over time, two things became apparent. First, because of insufficient degrees of freedom, it was impossible to include in a single equation all of the very large number of variables that, according to the literature, are associated with per capita income and/or earnings differentials. Second, since we are focussing on both earnings and income which is, in turn, a function of earnings and labour force utilization, the beginnings of a model were already present. This feeling grew with the realization that unemployment and participation rates could be a function of some of the same variables that determine earnings. Moreover, there was the possibility that earnings might be an appropriate right hand variable in an unemployment or participation rate equation, bringing up the further possibility of simultaneity. In short, the possible existence of interdependencies or two-way causation convinced the author that the total economic system had to be considered. This lead eventually to the development of a full econometric model and the application of simultaneous equation estimation techniques.

Further expansion of the model evolved from the fact that, once a variable was added to the right hand side of any equation, a decision had to be made concerning whether or not that variable was largely determined within the economic system of a county. If so, that variable must be viewed as "endogenous", requiring a separate equation for its determination. If not, the variable is "exogenous", implying that it is largely determined by forces outside of a county's economic system. Following from this, the share of employment in the primary industries, mining and transportation are viewed as exogenous since these are based on a county's natural resource endowment (broadly defined to include location). Further, public administration is exogenous, since a military base is a function of location (e.g. the availability of suitable land for a base), and the site of a provincial capital is mainly related to history or "chance". The two other industry variables that are exogenous are construction and pulp and paper, both of which proved impossible to model.^{75/} The remaining exogenous variables are either demographic variables - per cent French, total population, etc. - or dummy variables.

On the other hand, the author's thinking was that such variables as the unemployment and participation rates, seasonality, hours, urbanization and the level of education, would be determined at least in part by the industrial structure and other economic characteristics of a county, and that similar economic forces would play a major role in determining the importance of such industries as higher order business and

residential services, and low and other high value added manufacturing. Accordingly, these became endogenous variables, that is, variables determined within the econometric model. The following presents a full listing of the endogenous and exogenous variables included in the model. Also outlined is the mean of each variable, both to indicate relative magnitudes, and to provide the interested reader with the necessary data to calculate the elasticities associated with different coefficients.

ENDOGENOUS VARIABLES (1961 unless otherwise noted)

	<u>MEAN</u>
EIC: Employment Income Per Capita (\$)	667.26
ER: Employment Ratio (employment as per cent of total population)	27.27
ELFR: Employed Labour Force Ratio (employment as per cent of total labour force)	94.35
W: Average Annual Earnings (\$)	2,256.94
DII: Degree of Inequality in Income (the inter-quartile range divided by the median for the income per non-farm family series)	0.87
UR: Unemployment Rate (%)	5.65
PR: Labour Force Participation Rate(%)	46.18
SE: Seasonality (%) (per cent of wage earners working twenty-six weeks or less per year)	21.93
HO: Hours (%) (per cent of wage earners working less than thirty-five hours a week)	8.01
PFEFL: Per Cent Female in Employed Labour Force	22.18
GLFP: Growth in Labour Force Population, 1951-61, Index (i.e. no change = 100)	109.03
NMR: Net Migration Rate, 1951-61	-6.18
GE: Growth in Employment, 1951-61, Index	100.91

ENDOGENOUS VARIABLES (1961 unless otherwise noted) - (cont'd)

	MEAN
URB: % Urban	36.08
LED: Level of Education (%) (per cent of population five and over not attending school with at least two years of secondary schooling)	20.82
HOBS: Higher Order Business Services (%) (communications, wholesale, finance, insurance and real estate, hotels and restaurants, other business services)	10.92
OVHAM: High Value Added Manufacturing excluding Pulp and Paper (%) (i.e. includes primary metals, metal fabricating and other manufacturing)	3.21
RS: Residential Services (%) (retail trade and other personal services)	14.85
EH: Education and Health (%)	8.43
LVAM: Low Value Added Manufacturing (%) (food products, wood products and textiles)	8.44
IR: Industry Residual (utilities plus "industry not specified")	3.14

21 Endogenous Variables

EXOGENOUS (OR PREDETERMINED) VARIABLES

AG: Agriculture (%)	8.41
FOR: Forestry (%)	5.59
FISH: Fishing (%)	6.69
WPR: Working Population Ratio (labour force source population as per cent of total population)	62.23
PA: Public Administration and Defense (%)	10.23
TRANS: Transportation and Storage (%)	8.25
CON: Construction (%)	7.13

EXOGENOUS (OR PREDETERMINED) VARIABLES - (cont'd)

PFR:	Per Cent French (Francophones as per-cent of total population)	19.59
MIN:	Mining (%)	2.34
NRI:	Natural Rate of Increase, (%); 1951-61	22.67
NFD:	Newfoundland Dummy	0.22
POP:	Total Population, 1961	41,248.37
UD:	University Dummy (for university towns)	0.15
PP:	Pulp and Paper (%)	2.37

14 Exogenous Variables

The following outlines the right hand variables - divided into endogenous and exogenous - included in each equation. The 21 equations are separated into five blocks with the three identities indicated by an asterisk.

EQ. No.	LEFT HAND VARIABLES	RIGHT HAND VARIABLES	
		ENDOGENOUS	EXOGENOUS
<u>INCOME AND EARNINGS BLOCK</u>			
1.	EIC	W, ER	
2.*	ER	ELFR, PR	WPR
3.*	ELFR	UR	
4.	W	LVAM, EH SE, HO	AG, FOR, FISH, PP
5.	DII	OHVAM, ER, HO	FOR

EQ.
NO.LEFT HAND
VARIABLESRIGHT HAND
VARIABLESENDOGENOUSEXOGENOUSLABOUR FORCE UTILIZATION BLOCK

6.	UR	LVAM, RS, GLFP, GE SE	AG, PA, TRANS
7.	PR	HOBS, GLFP, GE, PFELF	PA, AG PP
8.	SE	RS, EH, UR	AG, FOR, FISH, CON, PFR
9.	HO	EH	FOR, FISH, PA, TRANS, MIN
10.	PFELF	HOBS, EH, RS	PFR, TRANS

CHANGES OVER TIME BLOCK

11.	GLFP	NMR	NRI
12.	NMR	GE	NRI, PFR
13.	GE	RS, URB	AG, FOR, FISH, MIN, PA

DEMOGRAPHIC BLOCK

14.	URB	EH, HOBS, OHVAM	PP, PA, NFD
15.	LED	HOBS, RS, EH	FOR, FISH PA, TRANS

EQ. NO.	LEFT HAND VARIABLES	RIGHT HAND VARIABLES	
		<u>ENDOGENOUS</u>	<u>EXOGENOUS</u>
	<u>INDUSTRY BLOCK</u>		
16.	HOBS	LED, LVAM, EH, RS	TRANS, POP, AG
17.	OHVAM	HOBS, RS	MIN, UD, NFD
18.	RS	EIC, URB, GE	
19.	EH	HOBS	UD
20.	LVAM	RS, W	FISH
21.*	IR	HOBS, OHVAM RS, EH, LVAM	AG, FOR, FISH MIN, PP, CON, TRANS, PA

Some of the model's characteristics can be further illustrated by noting the linkages within and between each block. Block. As the above table suggests, there are numerous linkages and to a lesser extent feedbacks between the blocks. Also, considerable interdependency is apparent within some of the blocks, particularly the labour force utilization and industry sections. The "links" between blocks, defined as a left hand variable in one block which appears as a right hand variable in a second block, are outlined in the following table.

LINKS BETWEEN BLOCKS

<u>LINKS</u>		
<u>FROM</u>	<u>TO</u>	<u>LINK VARIABLES</u>
Earnings and Income	Industry	EIC, W
Labour Force Utilization	Earnings and Income	PR, UR, SE, HO
Changes Over Time	Labour Force Utilization	GLFP, GE
Demographic	Changes over Time	URB
Demographic	Industry	URB, LED
Industry	Income and Earnings	LVAM, EH, OHVAM
Industry	Labour Force Utilization	LVAM, RS, HOBS, EH
Industry	Changes over Time	RS
Industry	Demographic	HOBS, EH, RS, OHVAM

Earnings and Income is perhaps the most critical block in the model and thus all other blocks, either directly or indirectly, feed into this part of the system. On the other hand, there are only two separate links from the earnings and income block to other sections of the model; (that is, there are only two equations in other parts of the model - both in the industry block - where left hand variables from this block appear as explanatory variables). As will be indicated later, other attempts were made to use earnings or per capita income

as right hand variables, particularly in the labour force utilization and change over time sections, but with no success.

Four of the five dependent variables in the labour force utilization block feed directly into the earnings and income part of the model. In turn, two of the four remaining blocks are linked directly to the labour force utilization section. There is also some interdependency within the labour force utilization block, in particular between the unemployment rate and the degree of seasonality. The factors underlying this will be mentioned below.

The change over time sector links directly to only one block, labour force utilization, while variables from two blocks, demographic and industry feed into the change over time section of the model. The demographic block provides explanatory variables for two of the four remaining sections of the model, but only the industry section feeds into the demographic block.

Reflecting the critical role played by industry in the specification of the model, there are more linkages, feed-backs and within-block interdependencies associated with the industry block than with any other. Left hand variables from the industry block are used as explanatory variables in all of the remaining parts of the system while, in turn, the earnings and income and demographic blocks provide right hand variables for the industry section. No

direct interdependencies exist within the block but a few are present indirectly through variables (such as LED and URB) from other sections of the model.

Finally, it should be noted that, despite its apparent complexity, there are surprisingly few direct feed-backs (or interdependencies) between blocks in the system. Out of six possibilities such feed-backs arise in only two cases: between demographic and industry; and between the industry, and earnings and income blocks.

The links and feed-backs can be further illustrated with the use of a flow chart:

The above illustrates clearly the two direct feedbacks in the system, together with the "indirect" feedbacks that exist, for example, between the earnings and income and demographic sections through the industry block. More generally, the chart shows the pivotal roles played by the earnings and income and industry blocks within the structure of the model.

Chapters five through nine will discuss the underlying rationale and estimation results for the five blocks. Throughout these chapters, references will be made to the differences between the Ordinary Least Squares and Two Stage Least Squares specifications. Comparisons between the two sets of estimation results are more fully set out in Appendix G. In the following five chapters, only the estimation results for the final 2 SLS specification are presented. In discussing the results for each equation, emphasis will be placed on the various hypotheses that were tested empirically and, based on these tests, either accepted or rejected.

5. EARNINGS AND INCOME BLOCK

(a) Theory and Underlying Rationale

The earnings and income block includes five equations, three of which are behavioural. Employment income per capita, average earnings and the degree of inequality in income are determined through behavioural equations, while identities are used to define the employment ratio and the employed labour force ratio.

The theoretical background for the wage and per capita income equations is set out in Appendix A, while the foundations in the literature for all three behavioural equations were outlined in the literature section. Accordingly, this sub-section will offer only a brief review of the antecedents for these equations.

Turning first to the wage equation, the work of Hanna (25), Kuznets (33 and 34) and Perloff (48 and 49) emphasized the critical role played by industrial structure in the determination of wages and per capita income. Accordingly, following the methodology outlined in the last chapter, the decision was made to start with industrial structure in specifying a wage determination equation. Further, the work of Kuznets and Perloff suggested that, in the estimation results, negative signs could be anticipated for the primary industries, mining and resource-based and/or low value added manufacturing^{76/} with positive signs for secondary (i.e. high value added manufacturing) and most service industries. Beyond the industry variables, the results of various authors, including Reder (53)

Polzin (51) and Scully (58) suggest that other variables could add to the explanatory power of a wage equation; these variables include various labour force utilization measures plus the levels of education, capitalization,⁷⁷ and urbanization. Experiments were attempted with all of these variables (both industry and non-industry) plus a few occupation variables, using Ordinary Least Squares. The results of these OLS specifications are fully discussed in Appendix E. Only the final OLS and 2 SLS specifications are discussed below.

Turning to the equation determining the degree of income inequality, the work of Kuznets, Reder and Munro and Holmes suggests that the degree of income concentration is associated with the level of economic development or industrialization and, more particularly, will tend to be lower the higher the share of employment in secondary manufacturing, construction, mining and other industries with a large semi-skilled component. Moreover, the work of Williamson, plus Holmes and Munro, indicates that the degree of inequality should be higher the lower the participation and/or employment rate, while Chiswick suggests a negative association between income concentration and the level of education. All of these hypotheses were tested before a final specification was found.

The remaining behavioural equation is that for employment income per capita. This should be viewed as an "adjustment" equation for its purpose is to adjust for the inadequacy of earnings per wage earner as a proxy for earned income per employee. Ideally, this should be an identity, since employment income per capita is identically equal to employment income per employee multiplied by the employment ratio. Thus, in its log-linear form, the constant term should be equal to zero, and the two coefficients equal to one. As indicated below, not all of these conditions were met.

An alternative approach would be to use employment income per employee rather than average earnings as the primary endogenous variable in this model. The author, however, was persuaded from doing this by the fact that the former would be an "estimated" variable - as is employment income per capita - whereas earnings can be taken directly from the Census. Given the critical role played by the wage equation in the formulation of the model, it was thought best to use an actual variable here and then use an adjustment equation to determine employment income per capita. Certainly though, arguments could be made for the alternative approach.

Equations (2) and (3) are identities defining ER and ELFR. A separate equation for ER was required for the following reason: when an alternative specification for EIC was attempted using separately the logs of ELFR, PR and WPR (all divided by 100) rather than the log of the employment ratio, the wrong sign was attached to the employed labour

force ratio. That is, the sign was negative rather than positive, probably for the reasons outlined in Appendix C when the simple correlation results are discussed.^{78/} Finally, equation 3 defines ELFR as a function of the unemployment rate as the latter was used throughout the model (being the more "traditional" measure).

(b) Estimation Results

$$1. \ln \hat{EIC} = 1.087 + 0.827 \ln W + 0.751 (\ln (ER \div 100))$$

$$(0.690) \quad (0.081) \quad (0.102)$$

$$(1.58) \quad (10.24) \quad (7.37)$$

$$R^2 = .976 \quad \bar{R}^2 = .970 \quad F\text{-STAT} (2,43) = 151.74 \quad SEE = 0.104$$

$$2. \ln \left(\frac{ER}{100} \right) = \ln \left(\frac{ELFR}{100} \right) + \ln \left(\frac{PR}{100} \right) + \ln \left(\frac{WPR}{100} \right)$$

$$3. ELFR = 100 - UR$$

$$4. W = 3898.75 - 20.38 AG - 28.92 FOR - 13.36 FISH$$

$$(195.78) \quad (5.42) \quad (13.54) \quad (6.45)$$

$$(19.91) \quad (-3.74) \quad (-2.14) \quad (-2.07)$$

$$-34.19 LVAM + 12.13 PP - 17.99 EH - 15.54 SE$$

$$(10.68) \quad (6.41) \quad (11.32) \quad (11.48)$$

$$(-3.20) \quad (1.89) \quad (-1.59) \quad (-1.35)$$

$$-58.41 HO$$

$$(30.52)$$

$$(-1.91)$$

$$R^2 = .915 \quad \bar{R}^2 = .897 \quad F\text{-STAT} (8,37) = 50.01$$

$$SEE = 166.10 \quad CV = 7.73\%$$

$$5. DII = 1.103 - .0108 ER + .0158 HO$$

$$(0.171) \quad (.0051) \quad (.0072)$$

$$(6.46) \quad (-2.11) \quad (2.18)$$

$$-.0122 OHVAM - .0044 FOR$$

$$(.0058) \quad (.0042)$$

$$(-2.10) \quad (-1.03)$$

$$R^2 = .410 \quad \bar{R}^2 = .353 \quad F\text{-STAT} (4,41) = 7.13$$

$$SEE = 0.0902 \quad CV = 10.33\%$$

As would be expected since it is a replacement for an identity, the explanatory power of the employment income per capita equation is fairly high. However, only one of the three conditions outlined above are met. The constant term is not significantly different from zero using either a one- or two-tail test, but both coefficients on the explanatory variables are significantly different from one. The suggestion is that the estimated equation rather than the identity will provide the more accurate results in the model.^{79/}

Equation (4) is the end product of the large number of specifications that were attempted for the wage determination equation. The results for the various OLS specifications are fully analyzed in Appendix E. The specification represented by equation four is for the most part quite respectable. The explanatory power of the equation is very high, while all but two of the ³ coefficients are significantly different from zero at the 95 per cent confidence level using a one-tail test. Finally, all of the signs on the co-efficients can be supported from the theory and /or literature. Particularly noteworthy here are the negative coefficients on the three primary industries plus primary manufacturing. Still, the specification is not without its disappointments or limitations.

First, there is a lack of balance among negative and positive coefficients, since seven of the eight coefficients are negative. In the earlier OLS specifications, public administration and higher order business services were entering the equation with positive and significant coefficients^{80/} but PA ceased to be significant when HO and SE were added to the equation while HOBS - together with residential services - fell out in moving from OLS to 2 SLS estimation.^{81/} Accordingly, beyond the positive coefficient on pulp and paper, the only positive influences on earnings operate indirectly through the seasonality and hours variables.

The lack of positive industry coefficients may reflect one of three considerations: (1) it could be based on specification error;^{82/} (2) the possibility exists that the industries not included in equation four all pay slightly above average wages, but none are sufficiently above the average to be picked up in regression analysis; (3) for the missing industries, particularly in the tertiary sector, within industry earnings differentials between counties (related perhaps to intra-industry production function differences or the lack of within industry homogeneity among counties) may be too large to allow a significant coefficient to be estimated; the implication is that an industry may be a low-wage activity in one part of the Atlantic and a high wage activity in another. The final consideration suggests that, while industrial structure has in general

proved to be a powerful tool in explaining earnings differences between counties, this approach may have its limitations with respect to particular industrial activities, especially in the non-goods producing sectors.

As a second "disappointment", one industry notable by its absence is other high value added manufacturing (mainly secondary manufacturing). This is totally inconsistent with the literature, where almost all authors pointed out the positive association between secondary manufacturing and per capita income and/or earnings. It is apparent that as of 1961 secondary manufacturing was not a high-wage industry in the context of the Atlantic Region. The reasons for this can probably be discerned by comparing the composition of secondary manufacturing in the Atlantic with that in the more industrialized parts of Canada. As of 1961, the Atlantic Region possessed little or no activity in consumer durables, chemicals or the other more "modern" manufacturing industries, particularly in comparison with Ontario. Most secondary manufacturing activity in the Region is in primary metals, shipbuilding and local market oriented industries. Atlantic representation in these more modern industries has been on the increase in recent years, suggesting the possibility of different results when a 1971 model is estimated.

Third, among all of the non-industry variables that were attempted, only seasonality and hours added to the explanatory power of the equation. What is disappointing here is that seasonality and hours are for the most part only "compositional" or "structural" variables. Accordingly, none of the behavioural hypotheses regarding the levels of education, labour force utilization, urbanization and so on held up under empirical testing. At the same time, this specification points out the crucial role played by industry in the wage determination process in the Atlantic Region, which of course is supportive of the author's main hypothesis outlined earlier. Moreover, these results further suggest the wisdom of developing a full econometric model for, without a full model and the solution results derived from it, we would not be able to trace out the effects of changes in urbanization, education, etc. on income and earnings.

Because it never appears as a right hand variable, the degree of inequality in income is not an integral part of the model. This is perhaps just as well because it turned out to be very difficult to model, reflecting the lack of variability - and a few "aberrant" observations - in the series. After much experimentation, an OLS specification was found which seemed to satisfy most of the hypotheses outlined above (Table G-5). However, the specification fared badly

when estimated using Two Stage Least Squares, with four of the nine t-statistics falling below one and the adjusted R^2 decreasing from .46 to .36. During further experiments using 2SLS, only four variables, ER, HO, OHVAM and FOR, held up under repeated testing. All four coefficients are consistent with the hypotheses outlined in the literature; particularly noteworthy here is the negative and relatively large coefficient on OHVAM. Therefore, it was decided to go with equation (5), despite its low explanatory power. Greater success may have been achieved if a different measure of income concentration were used, in particular the GINI coefficient. It is recommended that the latter be utilized if a 1971 model is estimated.

6. LABOUR FORCE UTILIZATION BLOCK

(a) Theory and Underlying Rationale

This block contains five behavioural equations determining the unemployment rate, the participation rate, the degree of seasonality, hours worked, and the percentage of females in the employed labour force. Because this is a cross-section model, the literature and theoretical antecedents for these equations are almost non-existent. As a consequence, a more "heuristic" approach had to be used here than perhaps in any other section of the model. Still, a few insights are provided by time-series models, such as Candide (McCracken (38)), RD X 2 (Helliwell, etc. (27)), and the Wharton model in the United States (Evans (20)). These will be noted where appropriate. The underlying rationale for each equation will be discussed in turn.

In most time-series models,^{83/} the unemployment rate is derived as a residual once labour demand is developed implicitly from a production function for the total economy, and labour supply is derived from the participation rate and demographic factors. This would suggest that two variables that could be included in an unemployment rate equation in a cross-section model is the growth in employment (indicating labour demand) and either the growth in the labour force (i.e. actual labour supply) or the growth in the labour force source

population (potential labour supply). It was decided to use the latter because (1) in an overall welfare sense and perhaps in a behavioural sense as well, potential labour supply is more relevant than actual labour supply, and (2) the growth in the labour force source population is less highly collinear with the growth in employment than is the growth in the labour force. GE and GLFP together then are indicative of shorter term pressures on the labour market, with a negative sign anticipated for the growth in employment and a positive sign for the growth in the working population.

At the same time, it was concluded that industrial structure would be the best medium for picking up longer term or structural influences on the unemployment rate. Positive coefficients were anticipated for forestry, fishing, pulp and paper, mining and construction, where employment tends to be more volatile, and negative parameters were expected for the tertiary activities, where stability in employment is generally the norm. Our expectations concerning these signs are perhaps strengthened by the fact that 1961 was a recession year.

One problem might be encountered in interpreting the estimated coefficients. Since this is cross-section analysis both short and long term forces can be reflected in a single coefficient; for example, a strongly positive coefficient

on a given industry in the UR equation may be indicative of both the longer term employment characteristics of that industry, and a recent downturn in employment opportunities brought about by a recent deterioration in the Atlantic's comparative advantage in that industry. It is hoped though that this problem is minimized through (1) using the shorter term labour market variables in conjunction with the industry variables, and (2) making the growth in the labour force source population and especially the growth in employment endogenous to the model. The expectations are therefore that the coefficients on the industry variables are mainly reflective of longer term forces.

Three further right hand variables were investigated. The first was average earnings; the expected sign here was positive, under the hypothesis that, for a given industrial structure, the unemployment rate would rise with the wage rate as employers tend to economize on labour. - (The higher wage rate might be a reflection in part of higher levels of unionization.) Also, the unemployment rate should be inter-related with the degree of seasonality since both are measures of the amount of slack in the labour market. Moreover, in the present context, this inter-relationship should be all the stronger because in the Census the unemployment rate is measured at a single point in time, June, when unemployment in the Atlantic is well below normal. Therefore, it is not the usual annual

average. Because of this, the role played by the unemployment rate in conventional time series models is shared between the unemployment rate and the degree of seasonality in the present model. Finally, it was thought that the net migration rate might enter with a positive sign, indicating that, other things being equal, the greater the level of in-migration, the greater the pressure on the labour market.

Turning to the participation rate, it was thought that many of the same variables appearing in the unemployment rate equation would probably also appear in the participation rate equation. The initial expectation was though that non-industry variables would likely play a more important role here than in the previous equation. For example, the participation rate would perhaps be negatively related to the unemployment rate, plus the seasonality and hours variables, reflecting discouraged worker withdrawal from the labour force.^{84/} Also, it should be positively associated with the percentage of females in the employed labour force, using the latter as a crude measure of job opportunities for the "secondary" labour force,^{85/} while the growth in employment and the labour force source population would again pick up shorter term influences. Finally, a specification was attempted using earnings as a right hand variable, under the hypothesis that, for a given economic structure, participation rates would rise with earnings because of the "substitution effect" (i.e. because leisure becomes more expensive).^{86/}

In contrast to the participation rate, it was thought that the degree of seasonality would largely be a function of the industrial structure of a given county. The critical factor here of course is the weather and its differing impacts on different industrial activities. To the extent that a county's industrial structure is highly skewed toward industries which are greatly affected by winter weather, county employment will in general tend to be highly seasonal in nature. And as the wage equation indicates this in turn will have an important downward effect on average annual earnings.

Based on the above, positive coefficients were anticipated for the three primary industries plus construction while negative signs were expected on the tertiary industries where weather is less of a consideration (at least directly). The one exception in the tertiary sector might be residential services, in light of the large amount of part-time employment in that industry. The only non-industry variable considered at the outset was the unemployment rate, reflecting its inter-relationship with seasonality.

Based on time-series models, the expectation was that the hours variable - defined as the percentage of wage earners working less than thirty-five hours a week^{87/} - would be mainly a function of non-industry variables and, in particular, the three other labour force utilization measures, the unemployment rate, the participation rate and the degree of seasonality.

More specifically, the expectation was that the number of wage earners working below average hours would be a function of the general softness of the labour market and accordingly would be positively related with the unemployment rate and the degree of seasonality in the work force, and negatively related with the participation rate.

Two other variables that were investigated were per cent female in the employed labour force - in hopes of picking up the effects of part-time employment - and average earnings, under the hypothesis that the amount of overtime tends to be higher the lower the wage rate (since overtime is relatively "cheaper" to the employer). The expected sign on wages then was positive.⁸⁸ In contrast to previous equations, it was thought that industrial structure would play a relatively minor role in generating inter-county differences in hours worked, since the length of the work week tends to some extent to be standardized across industries.

Turning to the final equation in this block, the expectation was that the percentage of female workers in the employed labour force would be largely based on the employment characteristics of the industry mix of a county. Positive signs were anticipated for such industries as services and perhaps manufacturing where female employment opportunities tend to be relatively high, while the primary industries and perhaps construction and transportation would likely have negative signs, reflecting the male orientation of these activities.

(b) Estimation Results

6. UR = -2.96 -0.134 AG -0.200 LVAM
 (6.36) (0.053) (0.100)
 (-0.47) (-2.55) (-2.00)

+0.147 RS -0.137 PA -0.151 TRANS
 (0.186) (0.080) (0.095)
 (0.79) (-1.71) (-1.59)

+0.176 GLFP -0.106 GE +0.158 SE
 (0.081) (0.059) (0.094)
 (2.18) (-1.81) (1.69)

$R^2 = .761$ $\bar{R}^2 = .709$ F-STAT (8,37) = 14.69
 SEE = 1.754 CV = 31.03%

7. PR = 33.20 -0.176 GLFP +0.216 GE +0.171 PFELF
 (4.96) (0.058) (0.054) (0.155)
 (6.69) (-3.01) (3.93) (1.10)

+0.093 PA +0.147 AG +0.186 PP +0.361 HOBS
 (0.090) (0.044) (0.078) (0.207)
 (1.04) (3.31) (2.39) (1.74)

$R^2 = .846$ $\bar{R}^2 = .818$ F-STAT (7,38) = 29.90
 SEE = 2.058 CV = 4.46%

8. SE = 2.10 +0.466 AG +0.759 FOR
 (4.30) (0.079) (0.206)
 (0.49) (5.90) (3.69)

+0.516 FISH +0.338 RS -0.444 EH
 (0.081) (0.272) (0.300)
 (6.37) (1.24) (-1.48)

+0.452 CON +0.472 UR +0.053 PFR
 (0.205) (0.413) (0.026)
 (2.20) (1.14) (2.06)

$R^2 = .838$ $\bar{R}^2 = .804$ F-STAT (8,37) = 24.00
 SEE = 3.597 CV = 16.40%

9. HO = 13.03 -0.123 FOR +0.089 FISH -0.324 EH
 (1.45) (0.041) (0.029) (0.105)
 (9.01) (-2.97) (3.12) (-3.07)

-0.121 PA -0.094 TRANS -0.077 MIN
 (0.021) (0.067) (0.047)
 (-5.64) (-1.40) (-1.63)

$$R^2 = .666 \quad \bar{R}^2 = .615 \quad F\text{-STAT} (6,39) = 12.96$$

$$SEE = 1.320 \quad CV = 16.47\%$$

$$10. \quad \widehat{PFELF} = \begin{matrix} 5.29 & +0.656 \text{ HOBS} & +0.258 \text{ EH} \\ (2.46) & (0.196) & (0.205) \\ (2.15) & (3.34) & (1.26) \end{matrix}$$

$$\begin{matrix} +0.562 \text{ RS} & -0.225 \text{ TRANS} & +0.054 \text{ PFR} \\ (0.251) & (0.118) & (0.016) \\ (2.23) & (-1.90) & (3.36) \end{matrix}$$

$$R^2 = .775 \quad \bar{R}^2 = .747 \quad F\text{-STAT} (5,40) = 27.63$$

$$SEE = 2.424 \quad CV = 10.93\%$$

Most of the hypotheses tested for the unemployment rate equation proved to be consistent with the data, although not all of the industries noted earlier entered the preliminary OLS specifications with their anticipated signs. Moreover, there were a few disappointments. In early experiments, both average earnings and the rate of migration entered the regression with the wrong sign, suggesting in both cases that causation was in the opposite direction (that is, from unemployment to wages and migration). For the former, our earnings measure is probably not the appropriate variable for this equation since, being an annual figure, it directly includes the effects of manpower utilization. A better measure would be the average hourly wage rate by county but no such variable is presently available.

Further, two of the industry coefficients probably require some explanation. The coefficient on agriculture is negative, reflecting the fact that underemployment rather than

unemployment is the major problem in that sector,^{89/} while residentiary services entered the regression with a positive sign, probably a consequence of the large amount of part-time work in that activity.^{90/} Note should also be made of the fact that the two remaining primary industries are not included in the equation. They do however enter indirectly through the seasonality variable.

One encouraging result was that GE and GLFP are both very strong variables despite a high degree of collinearity between them.^{91/} Finally, as Table G-6 (Appendix G) indicates this specification held up very well in moving from OLS to 2 SLS estimation. Only one variable, HOBS, had to be removed from the equation. In general then, equation (6) seems to be a reasonable specification. Most important, the regression is explaining 71 per cent of the variation in the dependent variable, a fairly creditable result for an unemployment rate equation in a cross-section model.

In the case of the participation rate, only three of the seven non-industry variables noted above - namely GLFP, GE and PFELF - stood up under repeated testing. The hours variable held up fairly well for a while but became negligible in magnitude once further industry variables were included in the regression. Earnings entered with a negative rather than a positive - sign with a t-value well below one.^{92/}

Finally as Table G-7 illustrates, the UR and SE variables - reflecting the "discouraged worker withdrawal" hypothesis - were kept in at the OLS stage (despite marginal t-values), but their coefficients become negligible when the equation was estimated using Two Stage Least Squares. In the final analysis then, industrial structure came to play a much larger role than was originally anticipated, with four industry variables being retained in the final equation. Despite these disappointments, the explanatory power of the final participation rate equation is fairly high, accounting for 84 per-cent of the inter-county variation in the dependent variable.

Turning to the seasonality equation, Equation (8) indicates that most of our original expectations were realized. Beyond the original hypotheses, an analysis of the residuals from a preliminary OLS equation suggested that the equation might be under-estimating the degree of seasonality in areas that were predominantly Francophone. As anticipated, PFR entered the equation with a significantly positive coefficient. The coefficient in the final equation is fairly small, but still it indicates that, other things being equal, the degree of seasonality will be five percentage points higher in a totally Francophone county relative to one that is entirely Anglophone. This variable may be picking up the influence of important differences in life-style between the two cultures. There

was however some worry that PFR may simply be acting for another variable not included in the regression. As perhaps the most likely candidate, PFR was replaced with the level of urbanization in the equation, but the coefficient on the latter was negligible in size and far from significant.

In contrast to a number of equations, the OLS specification for seasonality held up fairly well under simultaneous equation estimation, related in part to the fact that most of the right hand variables are exogenous (Table G-8). In the final analysis, only one variable, PA, had to be dropped in moving from OLS to 2SLS estimation.

In sharp contrast with the previous three equations, the hours variables turned out to be one of the more difficult left hand variables to model. Most important, our hypotheses regarding the three labour force utilization variables did not hold up under testing. When an early OLS specification was estimated using only the three manpower utilization measures, the unemployment rate entered with the wrong sign and the three variables together explained less than a fifth of the inter-county variation in hours. Further experiments were attempted with these variables at a later stage but with no success. Difficulties were also encountered with other non-industry variables. In particular, the coefficient on PFELF turned out to be negligible, while the coefficient on earnings

although significant at the 90 per cent confidence level - had a negative sign, suggesting that causation is in the opposite direction, i.e. from hours to earnings as in the earnings equation. Again, the hourly wage rate would likely be a more appropriate right hand variable for the hours equation.

It became necessary therefore to fall back on industrial structure using an iterative procedure to search out the best specification. Some success was achieved, but still equation (9) has its limitations. First, there is a lack of balance between positive and negative coefficients with only one variable entering with a positive sign. One other industry that was anticipated to be positively related with the HO variable was residential services (again reflecting part-time work) but, while the sign on this industry was always positive, the size of the coefficient varied greatly among specifications in early OLS work and its t-statistic was generally below one. Second, the explanatory power of the regression is relatively weak, as nearly 40 per cent of the inter-county variation in hours is left unexplained. The one possibly favourable aspect of the equation was that the specification altered very little in going from OLS to 2SLS estimation, reflecting the fact that five of the six right hand variables are exogenous (Table G-9).

In sharp contrast, Per cent Female in the Employed Labour Force proved to be relatively easy to model, as the estimation results were generally consistent with our original hypotheses. The one major divergence lies in the fact that none of the primary industries appear in the final equation.^{93/} This leads in turn to the major drawback to the specification, the fact that only one of the five coefficients is negative. Otherwise, the specification seems to be fairly robust, with the five right hand variables explaining three-quarters of the variation in the dependent variable.

Note should be made of the one non-industry variable in the equation. The analysis of the residuals from an initial regression suggested that the equation might be under-estimating PFELE in counties that are predominantly Francophone. As expected, PFR entered the equation with a positive and significant coefficient.^{94/} The positive sign on PFR is perhaps contrary to expectations since the conventional wisdom is that there are closer ties to "hearth and home" among the French community. Because of this, there was some concern that PFR was simply acting for another variable. The most likely possibility was employment income per capita, since the predominantly French counties tend to be among the lower income areas in the Atlantic. However, when EIC was added to the equation, the coefficient

on PFR declined only marginally, whereas the coefficient on EIC - while possessing its expected sign, i.e. negative - had a t-value well below one.

Before closing this section, a few general comments should be made. Because of the lack of a theoretical base, the approach used to develop this block of the model was more iterative and/or "heuristic" than one would have liked. The final specification for each equation possesses certain limitations and in a number of cases differ quite significantly from the original hypotheses offered at the beginning of the chapter. Perhaps most importantly, industrial structure came to play a more significant role in equation specification than was originally anticipated. Still, it should be emphasized that each of the five equations is the end product of a large amount of experimentation. Moreover, each specification can probably be defended if not in terms of traditional economic theory, then at least in terms of known characteristics of the Atlantic economy. Finally, as Chapter 10 will indicate, all five of the labour force utilization equations performed quite well when it came to solving the model.

7. CHANGES OVER TIME BLOCK

a. Theory and Underlying Rationale

This section focuses on the three 1951-61 growth variables, GLFP, NMR and GE, that are included in the model. There is considerable variation in the underlying rationale for the three behavioural equations in this block.

The equation for growth in the working population (i.e. the population fifteen and over) is close to being an identity. It is a function of two variables, the net migration rate, that is the net additions to and subtractions from the labour force population through migration, and the birth and death rates (the net additions from "natural" causes). Here, the latter is summarized by the rate of natural increase between 1951 and 1961; in the case of births, this acts as a proxy variable for the birth rate between 1936 and 1946, which is the period relevant to net additions to the working population from 1951 on. Implicit here is the assumption that, while the rates have likely changed, the rank ordering of counties with respect to the rate of natural increase has probably altered relatively little over this period. The "rate" is calculated by dividing the number of births minus the number of deaths into the 1951 population.

Similarly, the net migration rate is defined as the number of "net" migrants between 1951 and 1961 divided by the 1951 population. Therefore, a positive figure indicates net in-migration, a negative figure net out-migration;

(following from this, the average for the 46 Atlantic counties is negative). The expected signs on both variables are positive.

The literature suggests that the net migration rate can be related to a wide assortment of variables. Included here would be the unemployment rate (including in the present context the degree of seasonality), the participation rate, hours, average earnings, recent employment trends and perhaps the rate of natural increase (in order to include forces related to population supply).^{95/} All of these variables were tested in attempts at modelling the net migration rate.

Turning to the final variable in this block, the expectation was that the growth in employment would be largely a function of industry mix. Therefore, the rationale applied here is similar at least in spirit to shift/share analysis and particularly to the proportionality effect in shift analysis.^{96/} Accordingly, the hypothesis is that a county will experience above average growth in employment if its

industrial structure is skewed towards those industries experiencing above average growth in the Atlantic as a whole (i.e. the so-called proportionality effect). As a corollary, it is assumed that within industry shifts in comparative advantage among counties - the so-called differential effect - play a relatively minor role in generating inter-county differences in the growth in employment. The 1961 industry mix

is used, although ideally the 1951-61 averages would likely be more appropriate.^{97/} Negative signs were anticipated for the primary industries plus primary manufacturing, while positive coefficients were anticipated for most of the tertiary industries plus perhaps higher value added manufacturing.

(b) Estimation Results

$$11. \text{ GLFP} = 95.48 + 0.850 \text{ NMR} + 0.829 \text{ NRI} \\ (0.622) (0.014) (0.024) \\ (153.46) (60.61) (34.53)$$

$$R^2 = .995 \quad \bar{R}^2 = .995 \quad F\text{-STAT} (2,43) = 4402.13 \\ SEE = 1.507 \quad CV = 1.47\%$$

$$12. \text{ NMR} = -72.84 + 0.700 \text{ GE} - 0.112 \text{ NRI} - 0.074 \text{ PFR} \\ (3.68) (0.039) (0.094) (0.034) \\ (-19.79) (17.97) (-1.20) (-2.14)$$

$$R^2 = 0.929 \quad \bar{R}^2 = 0.924 \quad F\text{-STAT} (3,42) = 184.38 \quad SEE = 5.479$$

$$13. \text{ GE} = 170.68 - 0.679 \text{ AG} - 0.988 \text{ FOR} - 1.102 \text{ FISH} \\ (47.83) (0.479) (0.593) (0.562) \\ (3.57) (-1.42) (-1.66) (-1.96)$$

$$-1.387 \text{ MIN} + 4.678 \text{ RS} + 0.880 \text{ PA} + 0.347 \text{ URB} \\ (0.732) (2.526) (0.473) (0.205) \\ (1.90) (-1.85) (1.86) (1.69)$$

$$R^2 = .747 \quad \bar{R}^2 = .700 \quad F\text{-STAT} (7,38) = 16.04 \\ SEE = 14.655 \quad CV = 14.52\%$$

As befits a "near identity", the explanatory power of the GLFP equation is very high, with the adjusted R^2 approaching unity and the t-statistics on both explanatory variables very high. It seems then that the natural rate of increase between 1951 and 1961 has acted as a fairly reliable proxy for the rate between 1936 and 1946. The estimation

results barely changed at all in moving from OLS to 2 SLS estimation (Appendix G).

Greater difficulties were encountered in modelling the net migration rate. In initial experiments, none of the 1961 (i.e. "level") variables - including per capita income, earnings and the unemployment rate - entered the regression with both its anticipated sign and a t- value over one. This likely reflects the fact that the more appropriate right hand variables for the present equation would involve some type of 1951-61 average. Rather than introducing a totally new set of variables to the model, the decision was made to focus on the two "growth" variables, the growth of employment (as a summary measure of employment opportunities) and the rate of natural increase (to indicate population supply pressures).

However, the residuals from an intermediate specification pointed out that the equation was under-estimating the rate of net out-migration from the largely Francophone counties. As equation 12 illustrates, PFR entered with a negative sign and a t- statistic greater than two. Similar to previous equations, the negative sign on PFR - suggesting that other things being equal Francophones have a greater propensity to emigrate than Anglophones - is perhaps contrary to conventional wisdom; however, this result can perhaps be explained to some extent by the proximity of the majority of the Francophone counties to the Quebec border.

Despite our lack of success with the more conventional variables (unemployment rate, per capita income, etc.), the NMR equation above represents a fairly robust and probably readily defensible specification. The critical explanatory variable is the growth in employment, indicating that migrants are attracted to areas where the net additions to employment are largest. Perhaps most significantly, the three variables together are explaining over 92 per cent of the inter-county variation in net migration, a highly creditable result for a cross-section equation.

In accordance with the hypothesis offered above, experiments were attempted with almost all of the available industry variables in modelling the growth in employment equation. One difficulty encountered in early OLS experiments was that the more urban-oriented activities, such as higher order business services and higher value added manufacturing, generally had positive coefficients but t-statistics below one. The suggestion was that the level of urbanization might be a more appropriate variable, to the extent that the "high-growth" nature of these industries might be a function in turn of their urban bias. As equation 13 indicates, the coefficient on URB was both positive and significant (using a one-tail test). In general then, equation 13 seems to be a fairly satisfactory specification, particularly in light of the fact that the equation takes little account - except perhaps

through the ~~UR~~ variable - of intercounty shifts in comparative advantage. The equation explains 70 per cent of the intercounty variation in the growth in employment, while the signs on the variables are in accord with initial expectations.

Special note should be made of the negative coefficient on residentiary services. It seems reasonable to assume that, in counties whose export base industries showed employment declines between 1951 and 1961, residentiary services perhaps provided the only available outlet for surplus labour. The suggestion is then that, other things being equal, the share in residentiary services in 1961 would be higher the lower the growth in employment over the preceding decade. In this sense, the share in residentiary services and the growth in employment are inter-related, with causation going in both directions. This point will be returned to in the industry block.

In summary, some measure of success was achieved in modelling the three "change over time" variables, although not all of the explanatory variables suggested by the literature ended up in the final specifications (particularly in the case of the net migration rate equation). And similar to the equations in the labour force utilization block, these equations performed well in solving the total model.

8. DEMOGRAPHIC BLOCK

(a) Theory and Underlying Rationale

The two behavioural equations in this block determining the level of urbanization and the level of the education have a similar underlying rationale. Both are predicated on the assumption that industries have differing requirements for (1) the amenities, services, and material and human inputs offered by an urban location, and (2) skilled labour, and that these differences generate in turn inter-country variations in the levels of urbanization and education. Accordingly, industry mix provides the right hand variables for the two equations in this section of the model.

The demographic block should probably be viewed in conjunction with the industry block for there is considerable interdependency between the two. This reflects the "two-way" causation between industrial structure on the one hand and the levels of urbanization and education on the other. From one perspective, the importance of a given industry in a county's economic base is a function of the county's access^{98/} to basic inputs, for example, resource inputs in the case of the primary industries and primary manufacturing, or its access to both inputs and markets, e.g. skilled labour together with urban services, amenities and markets in the case of higher

order business services and secondary manufacturing. This is the perspective that will be used in modelling the equations in the industry block.

However, from a different perspective, the levels of urbanization and education can be viewed as a function of industry mix in the manner outlined above. We are though, dealing with more than a question of perspective. The argument for two-way causation between the two blocks is further reinforced by the fact that the growth in urbanization and skilled labour on the one hand and the attraction of certain economic activities to a given location on the other can probably best be viewed as two parts of a single cumulative process, with the former feeding on the latter and visa versa. It is this cumulative process that underlies in part the concept of agglomeration economies associated with large urban centres.^{99/} Some of these issues will be returned to at the beginning of the next chapter.

The expectation was therefore that the level of urbanization would be mainly a function of the importance of urban oriented industrial activities in a given county. Thus, positive coefficients were anticipated for higher order business services and most other service activities, plus other high value added manufacturing, while the primary industries would

/...

likely have negative signs. Also, positive signs might be anticipated for pulp and paper, mining and lower value added manufacturing, since urbanization here is based on the broad Census definition of the percentage of the population in centres of 1,000 or more. $\frac{100}{\text{ }}$

Similarly, the level of education in a given county should depend on the skill requirements (i.e. the "demand" for skills) of the industries that make up its economic base. In particular, the level of education should be higher in those counties whose industry mix is skewed towards skill-intensive activities -- for example, higher order business services, education and health and perhaps high value added manufacturing -- and lower in counties concentrating in the primary industries and other less-skilled activities. One further hypothesis was tested here. As discussed in the literature section, the level of education might be positively related to the rate of net migration over the preceding decade, under the assumption that emigrants are largely composed of the most skilled elements of the population. Accordingly, one experiment was attempted using the net migration rate as a right hand variable.

(b) Estimation Results

14. URB = -25.50 +1.14 PP +1.78 FH
 (10.54) (0.42) (1.17)
 (-2.42) (-2.72) (1.52)

+0.79 PA +1.39 HOBS +4.79 OHVAM +24.83 NFD
 (0.19) (1.03) (1.18) (5.56)
 (4.09) (1.35) (4.02) (4.46)

$R^2 = .661$ $\bar{R}^2 = .609$ F-STAT (6,39) = 12.67

SEE = 13.924 CV = 38.59%

15. LED = 3.98 -0.326 FOR -0.171 FISH
 (7.33) (0.159) (0.095)
 (0.54) (-2.05) (-1.81)

+0.654 HOBS +0.629 RS +0.358 FH
 (0.535) (0.584) (0.359)
 (1.22) (1.08) (1.00)

+0.218 PA -0.233 TRANS
 (0.082) (0.206)
 (2.65) (-1.13)

$R^2 = .789$ $\bar{R}^2 = .750$ F-STAT (7,38) = 20.32

SEE = 3.650 CV = 17.53%

Equations 14 and 15 satisfy most of our initial expectations. Turning first to the URB equation, this seems to be a fairly reasonable specification, despite a lack of negative coefficients and a fairly low adjusted R^2 .

Turning to the first problem, the primary industries in the initial results generally had negative signs but the t-values were below one (in absolute terms). Note should also be made of the single non-industry variable. The analysis of the residuals from an intermediate specification indicated that the equation was consistently under-estimating the level of urbanization in Newfoundland, perhaps reflecting the importance in that province of fishing outports -- at least as of 1961 -- which in the Census are likely defined as "urban". As was anticipated, the Newfoundland dummy variable entered with a coefficient that was both positive and significant. Finally, as Table G-14 indicates, the specification probably improved in moving from OLS to 2SLS estimation, as two somewhat dubious variables ceased to be significant, namely, MIN and LVAM, both with positive coefficients in the OLS equation. 101/

Even fewer difficulties were encountered in developing the level of education equation. As anticipated, positive coefficients were generated for four service industries, while negative signs were indicated for forestry, fishing and transportation. However, our hypothesis regarding net migration did not prove to be consistent with the data. The migration rate entered the regression with a negative sign -- rather than positive -- although the t-value was less than one. Accordingly,

once industrial composition is taken into account, the data does not support the hypothesis that out-migration tends to depress the level of education in the source county.^{102/}

Finally, the specification and results altered very little when the equation was estimated using Two Stage Least Squares rather than Ordinary Least Squares.

9. INDUSTRY BLOCK

(a) Theory and Underlying Rationale

The antecedents for the industry block lie mainly in (1) the concept of input - output access developed by Perloff et al (49), (2) the concept of agglomeration economies popularized by Richardson (55 and 56) and (3) the work of Chenery (7) who, as discussed earlier, used per capita income and population as right hand variables in cross-sectional equations explaining the levels of industrialization and services. The work of these authors is highly inter-related. Perloff has noted that, first, industries differ greatly in terms of their requirements for skilled labour, resource and intermediate inputs, proximity to markets and so on, and second, a given industry will be attracted to a location that possesses the right combination of access characteristics -- in terms of inputs and markets -- required by that industry. ^{103/}

Richardson, among others, has argued that certain types of activities -- particularly the more "modern" activities -- are attracted to locations where agglomeration economies have already been developed. Such economies are generally based on large market size -- as reflected in high levels of urbanization and population concentration together with high per capita incomes -- and the presence of complementary

industrial activities, e.g. secondary manufacturing and business services. These concepts provide a number of testable hypotheses for the five behavioural equations in this block. The sixth equation defines an industry residual in terms of the thirteen other industries -- both endogenous and exogenous -- included in the model.

Turning first to higher order business services, the proportion of employment in HOBS was anticipated to be positively related to the level of education (reflecting its requirements for skilled labour) and perhaps per capita income, reflecting its high income elasticity (and related to this, the possibility that some segments of HOBS may be local-market oriented). Further, the expectation was that HOBS would be a positive function of urbanization and/or population size, the result of its urban orientation and the necessity for access to large markets. Finally, such complementary activities as other high value added manufacturing, education and health and transportation were expected to have positive coefficients, the consequence of forward linkages from these activities to higher order business services.

Initial expectations were ~~that~~ the share of employment in other high value added manufacturing would be a positive function of the level of education and urbanization and/or

population size, plus such industries as HOBS, RS, EH and TRANS (reflecting the tendency for secondary manufacturing firms to be attracted to areas where such services are already available). Also the principal components results suggested that OHVAM would be positively associated with mining, reflecting the forward linkage out of that primary activity.

The initial presumption was that the share of employment accounted for by residential services would be a positive function of urbanization (and/or population size) and per capita income (the result of the high income elasticity for services), and, based on a previous equation, would be negatively related to the growth in employment over the previous decade. Turning to education and health, it was assumed that the share of employment in this activity would be positively associated with the level of education, per capita income, the level of urbanization and employment in higher order business services (the industry that is perhaps most complementary with education and health).

Finally, our a priori expectations were that lower value added manufacturing would be positively related to the three primary industries (the consequence of its requirements for resource inputs), plus the level of urbanization and/or residential services (since primary manufacturing requires

access to at least some complementary service activities), and would be negatively associated with the wage level. The latter reflects the labour intensity of this activity. However, an alternative approach was tested for lower value added manufacturing. This involved splitting up LVAM into its three components -- food products, wood products and textiles and clothing -- and modelling each separately. An identity would then be included adding the three together. The hypothesis was that different underlying factors were associated with each component. The expectation was that (1) food products would be a function of agriculture, fishing, earnings and perhaps residential services, (2) wood products would be largely determined by its counterpart primary industry, and (3) textiles and clothing would be mainly related to earnings and perhaps the level of urbanization.

The industry residual includes utilities and the residual in the Census data, i.e. the "industry not specified" category. Utilities does not appear in any equation as a right hand variable and would likely be impossible to model. Following the arguments of Tobin and Brainard (6), the implied specification for this residual equation will be spelled out in the following sub-section.

(b) Estimation Results

16. HOBS = -7.27 +0.273 LED +0.100 EH
 (1.98) (0.061) (0.165)
 (-3.67) (4.48) (0.61)

+0.215 TRANS +0.021 (POP ÷ 1,000) +0.510 RS
 (0.107) (0.009) (0.181)
 (2.01) (2.36) (2.81)

+0.070 AG +0.104 LVAM
 (0.040) (0.085)
 (1.74) (1.23)

$R^2 = .756$ $\bar{R}^2 = .711$ F-STAT (7,38) = 16.82

SEE = 2.038 CV = 18.663

17. OHVAM = -7.55 +0.303 MIN +0.217 HOBS
 (2.16) (0.081) (0.190)
 (-3.49) (3.76) (1.14)

+0.556 RS -1.83 NFD -1.24 UD
 (0.223) (0.89) (0.93)
 (2.39) (-2.05) (-1.33)

$R^2 = .602$ $\bar{R}^2 = .553$ F-STAT (5,40) = 12.12

SEE = 2.201 CV = 68.548

18. RS = 14.98 +1.013 (EIC ÷ 100) +0.043 URB
 (1.66) (0.288) (0.025)
 (9.04) (3.53) (1.74)

-0.084 GE
 (0.015)
 (-5.56)

$R^2 = .526$ $\bar{R}^2 = .492$ F-STAT (3,42) = 15.51

SEE = 2.124 CV = 14.288

19. EH = 3.15 +0.414 HOBS +5.05 UD
 (0.77) (0.066) (0.63)
 (4.07) (6.24) (7.96)

$R^2 = .694$ $\bar{R}^2 = .680$ F-STAT (2,43) = 48.78

SEE = 1.536 CV = 18.22%

20. LVAM = 6.09 +0.188 FISH -0.434 (W ÷ 100)
 (7.40) (0.127) (0.202)
 (0.82) (1.48) (-2.13)

+0.729 RS
 (0.298)
 (2.45)

$R^2 = 0.276$ $\bar{R}^2 = 0.224$ F-STAT (3,42) = 5.33

SEE = 4.798 CV = 56.85%

21. IR = 100 -AG -FOR -FISH -MIN -PP
 -CON -TRANS -PA -HOBS -OHVAM
 -RS -EH -LVAM

More difficulties were encountered in modelling this block than in any other section of the model, and even after various experiments, many of the equations remain quite unsatisfactory, either in terms of explanatory power or the specification itself. Undoubtedly the strongest equation in the block is that for higher order business services.

Equation 16 satisfies most of the hypotheses outlined earlier and has a fairly respectable $\bar{R}^2$. Particular note should be made of three aspects of the equation. First, when URB and POP

were included in the same regression in early OLS experiments, the former entered with the wrong sign (i.e. negative) while the latter was positive with a t-value greater than one. The evidence suggests then that relative to urbanization, population size is a more basic determinant of the employment share in HOBS. Second, perhaps surprisingly, agriculture always entered the regression with a positive and significant coefficient; the coefficient is though quite small (as would be anticipated), as a share differential of ten percentage points in agriculture generates a differential of less than one point in HOBS.

Third, as presented in Table G-16, per capita income and other high value added manufacturing were retained in the equation at the OLS stage, but dropped out in moving to 2SLS estimation. At the same time, the coefficient on the level of education increased markedly. One conclusion that can be drawn from this is that, in the OLS specification, the influence of the availability of skills was being shared between LED and EIC (and perhaps OHVAM and EH as well) but that under 2SLS all of this influence was concentrated in a single variable, LED. The latter is probably the desired result, since LED is the more basic determinant. Finally, EH was retained in the equation despite a t-value well below one. The magnitude of the EH coefficient was for the most part fairly stable under different regressions and its sign was always positive.

Turning to other high value added manufacturing, equation 17 is consistent with only some of our a priori expectations and the explanatory power of the equation is fairly weak. Among the disappointments, in early OLS experiments, LED always entered with a positive coefficient but its t-value was never greater than one. Second, urbanization -- which proved to be the stronger variable relative to population size in early OLS analysis -- ceased to be significant when the equation was estimated using Two Stage Least Squares. Finally, EH and TRANS failed to add to the equation's explanatory power.

Accordingly, we are left with an equation containing only five variables, two of which are dummies. Regarding the two dummy variables, analysis of the residuals from an intermediate equation indicated that the specification was over-estimating OHVAM in (1) Newfoundland, reflecting probably the province's distance from major Canadian markets, and (2) smaller counties whose economic base is concentrated in part on a university or other large educational or health institution. As anticipated, both the Newfoundland dummy variable and the so-called "university dummy" ^{104/} entered the regression with negative signs and t-values greater than one in absolute terms.

In contrast with the OHVAM equation, the final specification for the residential services equation was totally in accord with initial hypotheses (with urbanization turning

out to be a stronger variable than population size). The three variables together are only explaining one-half of the inter-county variation in residentiary services, but, in this regard, note should be made of the lack of variability in the RS series. There was though some concern regarding the "legitimacy" of the GE variable and, in particular, whether it was simply acting on behalf of others perhaps more "basic" determinants, especially public administration. On balance, however, it was concluded that equation (18) represents the preferable specification, despite the danger that the coefficient on GE is overstated.^{105/} One useful property of the inclusion of GE with a negative sign is that it tended to dampen the growth in residentiary services when the model was solved. This point is taken up in the next section.

Among the variables tested for the education and health equation, only one did the job, HOBS. It was found, however, that an equation with HOBS alone on the right hand side was badly under-estimating EH in less populous counties where a single education or health institution played a significant role in the economic base. Therefore, a so-called "university dummy" was constructed with a value of one for the following census divisions and counties: Census Division 9 in Newfoundland, Antigonish and Kings counties in Nova Scotia, and Gloucester, Kings, Restigouche and Victoria counties in New Brunswick.

The explanatory power of the equation improved tremendously with the inclusion of UD, the adjusted R^2 moving from .23 to .68 in the final specification.

Lower value added manufacturing turned out to be very difficult to model. Among the three primary industries, only fishing had the correct sign and a reasonable coefficient, (while RS proved to be a much stronger variable than URE). Most important, the explanatory power of the final equation is very weak, with an $\bar{R}^2$ of only .22. However, little or no success was achieved in attempting to model the three components separately. Only in the case of one component, food products, was a significant equation estimated and even then the explanatory power was very low. The remaining two variables, wood products plus textiles and clothing, proved impossible to model. Therefore, it was decided to go with the full LVAM equation above, despite the fact that its only positive feature was that it "behaved" fairly well in solving the model. If though the model is re-estimated using 1971 data, thought should be given to moving all or part of the LVAM grouping to the exogenous category.

As noted earlier, the final equation in this block is an identity defining an industry residual variable in terms of the other industry variables in the model. By doing this,

we have side-stepped the estimation problems identified by Tobin and Brainard (6) in the context of portfolio allocation. If all of the industry variables were either exogenous, or endogenous and determined by behavioural equations, the estimated coefficients would have to be constrained to ensure that the total effect of a single explanatory variable summed over all industry equations was zero. In the present exercise, this condition is satisfied through the "implicit" specification for the industry residual equation.

However, by avoiding one problem, two other problems arose. First, there is no assurance that the implied specification for the industry residual equation is economically rational. The specification can be derived by substituting equations 16 to 20 into equation 21:

$$\begin{aligned}
 21. (a) \text{ IR} = & 90.60 - 1.070 \text{ AG} - 1.00 \text{ FOR} - 1.188 \text{ FISH} - 1.303 \text{ MIN} \\
 & - 1.00 \text{ PP} - 1.00 \text{ CON} - 1.215 \text{ TRANS} - 1.00 \text{ PA} \\
 & - 0.273 \text{ LED} - 0.100 \text{ EH} - 0.021 (\text{POP} \div 100) \\
 & - 1.795 \text{ RS} - 0.104 \text{ LVAM} - 0.631 \text{ HOBS} \\
 & + 1.83 \text{ NFD} - 3.81 \text{ UD} - 1.013 (\text{EIC} \div 100) \\
 & - 0.043 \text{ URB} + 0.084 \text{ GE} + 0.431 (\text{W} \div 100)
 \end{aligned}$$

By comparing equation 21 (a) with equations 16 through 20, it can be seen that, with the exception of the exogenous industry variables,^{106/} the coefficients for a single variable sum to zero. On the other hand, it is highly doubtful that there is any economic rationale underlying this implicit specification.

Second, by leaving the industry residual "unconstrained",^{107/} problems arose in solving the model for there was no guarantee that the solution result for IR -- and thus by implication all of the other endogenous industry variables -- were economically feasible. For example, negative values were sometimes secured for the industry residual. Because of this, an iterative procedure had to be adopted in order to acquire a feasible set of solution results. In short, by avoiding problems in estimation, difficulties were encountered in solution. In re-estimating the model for 1971, thought should be given to determining all endogenous variables through behavioural equations.^{108/}

Summarizing this sub-section, the above shows clearly that greater difficulties arose here than in any other block of the model and that doubts remain regarding some of the specifications. Still, it should be noted that, abstracting from the problem discussed in the previous paragraph, these equations performed reasonably -- and perhaps even surprisingly -- well in solving the model.

10. SOLUTION RESULTS

(a) Introduction

As discussed in the Methodology section, the TROLL system has the very useful property that once the model is successfully estimated, the user can go immediately to solution. Moreover, since a Gauss-Seidel iterative technique is included in the software package, a non-linear model can be solved using this system. None of the mathematical difficulties that can arise in solving a non-linear model were encountered in solving the present model.^{109/} This is illustrated in Table 1 below where the "control solution" values for the endogenous variables are compared with their actual 1961 average values. The control solution is derived by inserting the 1961 mean values for the exogenous variables into the model. The table indicates in particular that there were no difficulties with "non-convergence":

TABLE 1

COMPARISON OF "CONTROL SOLUTION" VALUES
WITH ACTUAL VALUES FOR ENDOGENOUS VARIABLES

	(1) CONTROL SOLUTION	(2) 1961 AVERAGES	(1) (2)
EIC	659.61	667.26	98.86
ER	27.09	27.27	99.34
ELFR	94.35	94.35	100.00
W	2,261.57	2,256.94	100.21
DII	0.876	0.873	100.34
UR	5.65	5.65	100.00
PR	46.14	46.18	99.91
SE	21.91	21.93	99.91
HO	8.03	8.01	100.25
PFELE	21.98	22.18	99.10
GLFP	109.26	109.03	100.21
NMR	-5.90	-6.18	95.47
GE	101.30	100.91	100.39
URB	35.21	36.08	97.59
LED	20.60	20.82	98.94
HOBS	10.77	10.92	98.63
OHVAM	3.10	3.21	96.57
RS	14.70	14.85	98.99
EH	8.37	8.43	99.29
LVAM	8.31	8.44	98.46
IR	3.76	3.14	119.75

All but one of the control solution values is within 4.5% of its actual value while 12 of 21 are within one per cent. The one exception to the first statement is the industry residual; the errors in the five endogenous industry variables have a

cumulative effect since all of the control solution values are under-estimates of their true values. For purposes of comparison, the solution results are included in each of the following tables.

(b). Sensitivity Tests

The above indicates that the model is capable of solution. A more important step is to investigate whether it provides "reasonable" solution results in an economic sense. This was done by running a series of sensitivity tests on the model. These involve changing only one -- or in the case of industry variables only a few -- of the "instrument" variables and then comparing the computed endogenous values with the control solution values. ¹¹⁰ The instrument variables can include not only the exogenous variables but also some of the endogenous variables, particularly the endogenous industry variables, the levels of education and urbanization, and earnings (as a proxy for productivity). An endogenous variable can be altered by either changing the intercept in its equation -- a so-called additive adjustment -- or by changing the coefficients on the right hand variables -- a multiplicative adjustment. Only the former operation was attempted in the present exercise.

When working with industry variables (either endogenous or exogenous) it was necessary to change more than one in order to ensure that the industry shares that were derived

represented a feasible or "rational" solution set. In particular, attention was focussed on the solution value for the industry residual to ensure that this did not diverge too markedly from its control solution value. For some sensitivity tests, particularly those involving the endogenous industry variables, more than one solution set had to be developed before acquiring feasible values for the industry shares.^{111/} Beyond this, no checks are available concerning the viability of the results, beyond those based on the knowledge and experience of the author and the reader.

As indicated above, each test involves an alteration to a constant term in the structural equation, or a change in an exogenous variable. The 1961 means for the exogenous variables are as follows:

AG (%)	8.41	PFR (%)	19.59
FOR (%)	5.59	MIN (%)	2.34
FISH (%)	6.69	NRI (%)	22.67
WPR (%)	62.23	NFD	0.22
PA (%)	10.23	POP	41,248.37
TRANS (%)	8.25	UD	0.15
CON (%)	7.13	PP (%)	2.37

Table 2 outlines the results of the sensitivity tests while the following indicates the composition of each test and discusses the results. Unless otherwise indicated, the constant

TABLE 2
SENSITIVITY TESTS

CONTROL SOLUTION	(A)	(B)	(C)	(D)	(E)	(F)	(G)
EIC	688.17	707.56	704.23	705.90	698.13	703.42	689.63
ER	27.13	27.89	27.73	27.80	27.54	27.66	27.13
ELFR	94.37	94.95	94.70	94.79	94.85	94.32	94.15
W	2,377.66	2,397.36	2,396.69	2,397.57	2,386.30	2,398.46	2,346.19
DIT	0.876	-0.874	0.876	0.866	0.865	0.869	0.881
UR	5.65	5.05	5.30	5.20	5.15	5.68	5.85
IPR	46.14	47.21	47.05	47.13	46.66	47.13	46.23
SE	21.91	19.30	20.17	19.93	19.88	20.26	21.61
HO	8.03	3.05	7.47	7.65	7.66	-8.00	8.07
PFELF	21.98	22.78	22.40	22.30	22.24	22.63	21.06
GLEP	109.26	110.86	111.12	112.93	110.74	110.83	110.96
NMR	-5.90	-4.03	-1.72	-1.59	-4.17	-4.07	-3.91
GE	101.30	103.99	104.42	107.47	103.79	103.93	104.16
URB	35.21	41.69	40.09	39.84	39.17	35.05	44.34
LED	20.60	22.45	21.58	22.39	21.39	21.79	21.06
HOBS	10.77	12.00	11.61	11.10	11.73	11.25	10.74
OHVAM	3.10	3.66	3.50	3.23	3.50	3.47	3.07
RS	14.70	15.24	15.10	14.82	15.05	15.19	14.67
EH	8.37	8.88	8.72	8.50	8.77	8.57	8.36
LVAM	8.31	8.11	7.46	7.62	7.84	7.89	7.74
IR	3.76	1.09	2.60	3.71	2.10	2.61	4.40

142

TABLE 2 (CONT'D)

SENSITIVITY TESTS

CONTROL SOLUTION	(H)	(I)	(J)	(K)	(L)	(M)	(N)
EIC	684.86	715.86	697.00	-691.78	689.64	669.15	678.39
ER	21.16	28.86	27.76	27.61	27.41	27.55	27.37
ELFR	94.34	94.40	94.40	94.37	94.37	95.13	94.75
W	2,361.49	2,357.22	2,364.76	2,354.31	2,361.44	2,266.59	2,317.55
DII	0.856	0.832	0.863	0.833	0.856	0.877	0.878
UR	5.65	5.52	5.60	5.63	5.63	4.87	5.25
PR	46.14	49.09	47.25	47.02	46.67	46.53	46.43
SE	21.91	19.51	19.95	20.24	20.21	19.68	19.93
HO	8.03	7.39	7.83	7.99	8.02	8.03	8.04
PFELF	21.98	26.39	23.54	21.91	22.42	22.06	22.09
GLFP	109.26	112.01	111.17	112.98	111.86	111.55	111.15
NMR	-5.90	-2.67	-3.67	-1.54	-2.85	-3.21	-3.69
GE	101.30	105.93	104.50	107.55	105.67	105.15	104.47
URB	35.21	53.77	41.78	52.72	44.25	35.47	35.63
LED	20.60	25.77	22.75	21.91	21.57	-21.17	-21.21
HOBS	10.77	15.83	12.52	11.33	11.10	11.03	10.94
OHVAM	3.10	4.74	3.70	6.53	4.85	3.03	3.10
RS	14.70	15.68	15.09	15.26	15.03	14.49	14.64
EH	8.37	10.46	9.10	8.61	8.51	8.48	8.45
LVAM	8.31	8.43	7.97	8.13	7.93	10.95	9.34
IR	3.76	-3.15	3.62	2.13	4.57	4.02	5.51

TABLE 2 (CONT'D)

SENSITIVITY TESTS

	CONTROL SOLUTION	(O)	(P)	(Q)	(R)	(S)	(T)	(U)
ENC	659.61	664.92	680.25	683.90	703.60	698.21	660.27	684.83
ER	27.09	28.18	27.53	27.37	27.85	27.63	28.16	28.50
ELFR	94.35	94.35	94.36	94.36	94.59	94.51	94.64	95.92
W	2,261.57	2,202.99	2,313.22	2,340.88	2,384.77	2,379.39	2,186.18	2,259.82
DII	0.876	0.807	0.855	0.867	0.847	0.858	0.865	0.856
UR	5.65	5.65	5.64	5.64	5.41	5.49	5.35	4.08
PR	46.14	48.00	46.88	46.60	47.30	46.98	47.80	47.75
SE	21.91	20.95	20.44	20.31	18.67	19.17	26.02	21.24
HO	8.03	7.32	7.81	7.93	6.91	7.29	8.02	7.97
PFELF	21.98	29.25	24.50	23.31	24.18	23.49	26.42	22.63
GLFP	109.26	102.55	108.01	109.38	111.56	111.29	104.22	100.17
NMR	-5.90	-13.80	-7.38	-5.77	-3.20	-3.52	-11.83	-5.53
GE	101.30	90.03	99.19	101.49	105.16	104.71	101.31	100.02
URB	35.21	66.99	46.21	41.00	46.64	43.02	35.45	38.06
LED	20.60	28.95	23.85	22.52	23.69	22.87	20.66	21.30
HOB	10.77	16.35	12.70	11.78	12.19	11.74	10.83	11.26
OHVAM	3.10	7.29	4.55	3.86	3.74	3.55	3.12	3.47
RS	14.70	20.06	16.56	15.68	15.32	15.14	14.72	15.19
EH	8.37	10.68	9.17	8.79	11.96	10.77	8.40	8.58
LIVAM	8.31	12.28	9.26	8.50	8.04	7.94	8.65	8.67
IK	3.76	-14.68	-0.239	3.39	0.75	2.84	3.29	1.83

TABLE 2 (CONT'D)
SENSITIVITY TESTS

	CONTROL SOLUTION	(V)	(W)	(X)	(Y)	(Z)
EIG	659.61	667.37	638.24	831.25	672.10	728.86
ER	27.09	27.56	26.47	27.29	27.84	28.07
ELFR	94.35	94.38	93.96	93.57	94.40	94.43
W	2,261.57	2,258.51	2,219.46	2,971.40	2,256.64	2,470.45
DII	0.876	0.863	0.886	0.838	0.855	0.860
UR	5.65	5.62	6.04	6.43	5.60	5.57
PR	46.14	46.91	45.27	46.87	47.39	47.77
SE	21.91	21.72	22.99	22.84	21.60	18.14
HO	8.03	7.85	8.30	7.64	7.73	7.74
PFELF	21.98	23.18	21.72	26.03	23.92	24.20
GLFP	109.26	109.62	106.47	103.86	109.85	112.80
NMR	-5.90	-5.48	-9.19	-12.26	-5.22	-1.75
GE	101.30	101.91	96.61	92.22	102.29	107.25
URB	35.21	40.26	32.85	53.00	43.35	44.58
LED	20.60	21.88	19.47	24.99	27.66	28.95
HOB	10.77	12.16	10.96	13.68	13.02	13.23
ORVAM	3.10	3.53	3.18	5.54	3.80	3.97
RS	14.70	14.95	14.78	17.96	15.10	15.31
EH	8.37	8.95	8.45	9.57	9.30	9.39
LVAM	8.31	8.50	8.55	7.63	8.62	7.48
IR	3.76	0.91	3.09	-5.38	-0.84	5.62

TABLE 2 (CONT'D)

SENSITIVITY TESTS

	CONTROL SOLUTION	(A-1)	(B-1)	(C-1)	(D-1)	(E-1)	(F-1)	(G-1)	(H-1)
EIC	659.61	662.44	718.62	685.85	688.67	668.05	644.57	646.81	668.29
ER	27.09	27.38	27.61	27.39	27.13	27.62	26.26	26.38	27.30
ELFR	94.35	94.35	94.38	95.14	94.22	94.35	93.41	93.55	94.36
Diff	2,261.57	2,251.64	2,465.80	2,347.33	2,379.78	2,256.45	2,262.56	2,262.41	2,281.38
	0.876	0.870	0.875	0.868	0.838	0.869	0.887	0.887	0.881
UR	5.65	5.65	5.62	4.86	5.78	5.65	6.59	6.45	5.64
PR	46.14	46.62	47.00	46.26	46.26	47.04	45.17	45.31	46.49
SE	21.91	21.95	18.49	16.41	22.06	21.93	22.31	22.25	21.81
HO	8.03	7.99	8.03	7.97	5.97	8.01	8.07	8.07	8.12
PFELF	21.98	22.41	22.35	22.61	22.67	27.20	21.59	21.65	21.13
GLFP	109.26	110.83	113.80	108.44	108.35	109.01	114.71	113.90	113.21
NMR	-5.90	-4.06	-0.57	-6.87	-6.98	-6.21	-5.38	-0.45	-1.26
GE	101.30	103.94	108.93	99.92	99.76	100.87	102.06	101.95	107.94
URB	35.21	47.09	48.26	37.98	38.23	36.17	33.51	33.77	31.49
LED	20.60	21.07	22.35	21.29	21.35	20.84	20.18	20.25	19.68
HOBS	10.77	11.10	11.30	11.23	11.26	10.94	10.47	10.52	10.11
OHVAM	3.10	3.34	3.50	3.47	3.51	3.22	2.87	2.91	2.60
RS	14.70	15.02	15.22	15.20	15.25	14.87	14.41	14.46	14.08
EH	8.37	8.51	8.59	8.56	8.57	8.44	8.25	8.27	8.10
LVAM	8.31	8.58	7.43	8.31	8.20	8.45	8.10	8.13	7.77
IR	3.76	2.45	8.95	2.23	2.20	3.09	4.90	4.73	6.34

terms and the exogenous variables are unchanged from the estimated coefficients and the above mean values respectively. The results for the endogenous variables are grouped in accordance with the five blocks outlined in Chapters 5 through 9. In discussing the results, particular attention will be focussed on the three main endogenous variables: per capita income, earnings and the degree of inequality in income, especially the first two.

SENSITIVITY TESTS

For Tests (A) through (C): TRANS from 8.25 to 10.25
and PP from 2.37 to 3.37

Test (A): AG from 8.41 to 5.41

Test (B): FOR from 5.59 to 2.59

Test (C): FISH from 6.69 to 3.69

For Tests (D) through (S): AG from 8.41 to 7.41, FOR from 5.59
to 4.59, FISH from 6.69 to 5.69

Test (D): PA from 10.23 to 13.23

Test (E): TRANS from 8.25 to 11.25

Test (F): PP from 2.37 to 5.37

Test (G): CON from 7.13 to 10.13

Test (H): MIN from 2.34 to 5.34

Test (I): Add 3.0 to intercept in HOBS equation

Test (J): Add 1.0 to intercept in HOBS equation

Test (K): Add 3.0 to intercept in OHVAM equation

Test (L): Add 1.5 to intercept in OHVAM equation

SENSITIVITY TESTS (CONT'D)

- Test (M): Add 3.0 to intercept in LVAM equation
- Test (N): Add 1.5 to intercept in LVAM equation
- Test (O): Add 3.0 to intercept in RS equation
- Test (P): Add 1.0 to intercept in RS equation
- Test (Q): Add 0.5 to intercept in RS equation
- Test (R): Add 3.0 to intercept in EH equation
- Test (S): Add 2.0 to intercept in EH equation
- Test (T): PFR from 19.59 to 100; all other X variables
and intercepts unchanged
- Test (U): NRT from 22.67 to 11.33; all other variables unchanged
- Test (V): POP from 41,248.37 to 82,496.74
- Test (W): PA from 10.23 to 6.23, TRANS from 8.25 to 10.25,
CON from 7.13 to 9.13
- Test (X): Add 700 to intercept in W equation;
all other variables unchanged
- Test (Y): Add 5.0 to LED intercept
- Test (Z): Add 5.0 to LED intercept; AG from 8.41 to 6.41,
FOR from 5.59 to 3.39 and FISH from 6.69 to 4.69
- Test (A-1): Add 10.0 to URB intercept
- Test (B-1): Add 10.0 to URB intercept; AG from 8.41 to 6.41,
FOR from 5.59 to 3.59, and FISH from 6.69 to 4.69
- Test (C-1): Add 5.0 to SE intercept
- Test (D-1): Add 2.0 to HO intercept
- Test (E-1): Add 5.0 to PFELF intercept
- Test (F-1): Add 5.0 to GLFP intercept
- Test (G-1): Add 5.0 to NMR intercept
- Test (H-1): Add 5.0 to GE intercept.

Tests (A) through (C) compare the impacts of three per cent declines in employment share in the three primary industries; in each test, the shares in TRANS and PP increase by two and one per cent respectively. Table 2 indicates that the income gains are much greater for forestry (B) and fishing (C) than for agriculture, reflecting for the most part the fact that the shift out of agriculture generated almost no improvement in the employment ratio. In short, in shifting out of agriculture, productivity gains are realized but there are no improvements in the rate of labour force utilization except for those associated with the decreases in the hours and seasonality variables. Comparing Tests (B) and (C), the results are quite comparable, except that in (C), the decrease in DII is much sharper, reflecting in turn the more substantial decline in HO.

Turning to the other endogenous variables, improvements are indicated in almost all cases. For example, the suggestion is that the growth in the labour force population and employment would have been higher, and net migration less negative, if the Atlantic Region had been less dependent on the primary activities in 1961. Also, higher employment shares are indicated for the three "endogenous" service industries plus other high value added manufacturing, reflecting among other factors the increases in income, urbanization and

and education. On the other hand, LVAM declines in each case, the result mainly of the increase in the wage rate (and for (C) the decrease in fishing)

In Tests (D) through (S), declines of one percentage point are assumed for each primary industry, and then tests conducted for the remaining industries, both exogenous and endogenous. Tests (D) to (H) compare the five remaining exogenous industries. The greatest income and wage gains are achieved from increasing the employment share in public administration (Test D) and pulp and paper (Test F) with improvements occurring in both overall productivity and labour force utilization (including seasonality and hours). In the case of public administration, sharp improvements are also indicated in the "change over time" block, illustrating the high growth experienced between 1951 and 1961 by most Atlantic counties dominated by public administration. On the other hand, only modest improvements are indicated for urbanization, education, and the urban-oriented industries in the industry block, reflecting the relatively undeveloped forward linkages out of the public administration sector.^{112/}

In the case of transportation (Test E), a fairly substantial increase in employment income per capita is realized, based both on a rise in earnings and an increase in

the employment ratio. The results for Test (E) also point out fairly large increases in share for higher order business services and other high value added manufacturing, (both larger than in (D) and (F)), indicating the impact of linkages out of transportation. One concern of the author is that the model, as presently specified, understates the positive influence on income and earnings of such industries as HOBS and OHVAM. If this is true, then the model is also underestimating the gains in income and earnings resulting from shifts into transportation.

Compared with Tests (D) through (F), only modest increases in income and earnings are realized from shifts into construction (G) and mining (H), reflecting for the most part only nominal increases in the employment ratio. Moreover, in the case of mining, the growth performance has deteriorated. One aspect of the results from Test (H) suggests that the model is providing intuitively plausible solution values. Compared with the control solution values, the decline in the growth in employment is somewhat greater than the decline in the growth in the working population. This in turn tended to have a dampening effect on the employment ratio. This can best be illustrated by comparing the results for Test (H) with those for Test (E). The favourable effects on ER from improvements in HOBS, PFELF and SE are roughly comparable between the

two solution sets, but in (II) these favourable effects tended to be offset by the deterioration in growth performance and more particularly the sharper decline in GE relative to GLFP. In short the changes in these endogenous variables seem to be internally consistent, suggesting in turn that, in this respect at least, the structural form of the model is properly specified. Tests (I) through (S) attempt to analyze the effects of stimulating employment in the five endogenous industries. At first, 3.0 was added to the intercept term in the relevant structural form equation. In almost every case, however, the solution value for IR was not realistic. Accordingly, further tests were attempted with smaller amounts added to the intercept term.

These difficulties reflect the high degree of interaction among the industries in the industry block, plus the fact that most of the changes are in the same direction (that is, increases in HONS lead to gains in RS, EH and so on). This problem may also reflect a weakness in the specification of the residentiary services equation. Employment income per capita enters the RS equation with a positive coefficient. On the one hand, this seems plausible in light of the high income elasticity for services, and the local-market orientation of this activity. On the other hand, this positive coefficient tends to contradict the likely tendency for entrepreneurs in residentiary services to economize on labour as rates of pay

increase; this reflects in turn the fact that residential services is a highly labour intensive activity. Accordingly, perhaps as a consequence of these conflicting tendencies, Table 4 points out that in 1961 residential services accounted for about the same proportion of employment in the Atlantic, Ontario and Canada as a whole. In solving the present model, the only dampening effects on residential services came from the inclusion of the growth in employment in the RS equation. The dampening impact of GE on PS, while very useful in itself, was insufficient to prevent the generation of "non-feasible" solution values for the industry block. 113/

Tests (I) and (J) show the effects of stimulating employment in high order business services. Focussing on the more realistic set of results (Test J), Table 2 indicates a fairly substantial advance in per capita income, approximately on a par with Test (E) for transportation. Compared with Test (E), the increase in the employment ratio is somewhat higher (the result of fairly substantial advances in labour force participation and the proportion of females in the employed labour force) while the advance in earnings is lower. Further, as might be anticipated, stimulating employment in HOBS leads to fairly large increases in urbanization and the level of education (resulting from the increased demands for skills).

Finally, following from the previous paragraph, both the structure of the model and the sensitivity tests suggest that higher gains in earnings, labour force utilization -- and therefore income -- would likely be realized from stimulating employment in HOBS if the associated growth in residentiary services could be further dampened.

Tests (K) and (L) point out the impact of stimulating employment in other high value added manufacturing (mainly secondary manufacturing). Both solution sets are fairly reasonable. Moreover, they are highly comparable, despite the quite large difference in the solution values for the industry residual. Focussing on Test (L), this can perhaps best be compared with Test (J), for in -- terms of the industry block -- the only essential difference between them lies in higher (lower) share in OHVAM (HOBS) in Test (L). Our comparisons point out that marginally higher income gains accrue from stimulating higher order business services relative to secondary manufacturing (reflecting the higher level of labour force utilization associated with the former), but that OHVAM leads to a more equal distribution of income (as pointed out by the lower value for DII). Therefore, the possibility exists for conflicts between average income and income distribution. Such conclusions, however, must be offered very tentatively, the consequence of the limitations associated with the equation specification for DII.

Tests (M) and (N), particularly the former, suggest that only very meager gains in income and earnings can be realized by shifting employment out of the primary industries and stimulating employment in lower value added manufacturing. Here again, the results from stimulating JOBS (i.e. Test J) provide the most useful benchmark for comparison. In Test (M), the advance in per capita income is only \$9.54 compared with \$37.39 in Test (J), reflecting for the most part the negligible rise in earnings in (M) compared with an increase of better than \$100 in (J). However, the employment ratio also makes some contribution to this differential, rising by .44% in Test (M) and .67% in Test (J). Also noteworthy is the fact that the degree of inequality in income in Test (M) is above the control solution value, compared with a fairly large decline in the case of Test (J). All in all then, the results suggest that relatively small economic gains can be achieved from shifting employment out of the primary industries and stimulating employment in lower value added manufacturing. (On the other hand, such shifts may be defensible if they are the only alternative to economic stagnation and depopulation. In this context, it might be significant that these shifts do lead to some increase in the growth in employment.) 114/

Similar to the tests for LVAM, Tests (P) and (Q) indicate relatively small economic gains from stimulating employment in residential services, although in Test (Q), the increases in income and earnings are somewhat greater than in Test (M). As Test (Q) suggests, adding 3.0 to the intercept in the PS equation had the most explosive effect of any alteration on the remaining industries in the industry block, reflecting the large degree of interaction between residential services and the other endogenous industries. The background to this was discussed earlier.

Finally, Tests (R) and (S) point out fairly strong gains in income, earnings and labour force utilization from stimulating employment in education and health. Particularly noteworthy here, relative to most previous tests, is the sharp improvements in the seasonality and hours variables; this probably accounts for most of the advance in earnings. The upshot is that the income advance in Test (S) is the highest of any of the "comparable" (and/or "feasible") solution sets for the endogenous industry variables (the others being Tests (J), (L), (M) and (O)).

With one exception, Tests (T) through (H-1), involve "shocking" one of the non-industry variables in the model (most of which could be viewed in certain contexts as instrument

variables). In Test (T), the model is solved under the assumption that the Atlantic is 100% Francophone. Per capita income remains unchanged, as the advance in the employment ratio (based mainly on the increase in the participation rate)^{115/} is balanced off by the decline in earnings resulting from the increase in seasonality. Also noteworthy is the sharp advance in the net out-migration rate, leading to a sharp decrease in GLFP and through that variable to improvements in UR and PR. Test (U) shows the effect of cutting the rate of natural increase in half. As might be anticipated, the growth in the working population fell sharply, leading to improvements in the labour force participation rate and unemployment rate and, through these variables, to an increase in per capita income. All in all, this test seemed to provide internally consistent results, although the economic rationale underlying some of the changes in the industry block, -- in particular the advance in LVAM -- is not totally clear.

Test (V) indicates very modest economic gains from doubling the population of the "average" Atlantic county. The suggestion is therefore that the model is not sufficiently picking up the effects of economies of scale and agglomeration, and/or as of 1961 such economies had yet to be developed in the Atlantic Region. Test (W) returns us briefly to the industry block by analyzing the impact of a sharp decline in the share

of employment accounted for by public administration and defense, with the surplus workers going to construction and transportation. This allows us to focus on the longer-term impact of, for example, the closure of most of the major military bases in the Region. As indicated there, almost all of the changes in the more important economic indicators are unfavourable, despite the dampening effects of increases in the endogenous industries.

In Test (X), a 30% increase in productivity is simulated by adding \$.700 to the intercept in the earnings equation. This generated a non-feasible solution set in the sense that the industry residual is negative. This reflects for the most part the sharp advance in RS associated with the increase in per capita income. The advance in RS then led to the strong decline in the growth in employment. Because of these problems, it is very difficult to interpret these results. However, the solution set does suggest one general conclusion, that sharp advances in productivity can be expected to lead to a very different economic structure in the Atlantic relative to that in existence in 1961. The implication for the present exercise is that the structural form of the model would be very different at sharply higher levels of productivity. On the one hand, this serves to point out the limitations of a cross-sectional model of the type

developed here. Clearly, the solution results become less reliable, the greater -- and more basic -- the shock to the system. This was anticipated by the author. On the other hand, these difficulties serve the very useful purpose of pointing out the "general equilibrium" nature of any economic system. In the present example, a significant increase in productivity causes all of the actors in the economic system -- producer and consumer, manager and worker -- to make adjustments, leading in turn to changes in the structural coefficients of the model. In a more ambitious project, it might be possible to account at least partially for these adjustments. One possibility would be to estimate separate models for all five regions in Canada, and then for example, compare the coefficients in the Ontario and Atlantic models in order to provide some insights into future changes in the Atlantic coefficients. A further -- or perhaps complementary -- approach would be to re-estimate the Atlantic model for 1971 and then compare the coefficients with those for the 1961 Atlantic model.

In Test (Y), 5.0 was added to the intercept in the education equation, while in Test (Z), further adjustments were made in order to generate a positive value for TK. As Test (Y) indicates, the increase in LED generates considerably higher shares for the endogenous industry variables but, because these include both high and low wage activities, the gain in

per capita income is relatively modest; moreover there is a slight decline in earnings. It is only when declines in the primary industries are included, as in Test (2), that strong economic gains are realized. The implicit assumption in Test (2) is that the growing "endogenous" industries will attract labour only from the primary activities; this of course may not be the case. Tests (A-1) and (B-1) indicate that similar results were acquired through shocking the urbanization variable, first without and then with declines in the primary variables. In this case, the first set is the more reasonable, as the increase in URB generates relatively small increases in the endogenous industry variables. Similar to the test for POP, these results suggested that, as of 1961, agglomeration economies were relatively under-developed in the Atlantic Region.

Tests (C-1) and (D-1) illustrate the effects of decreasing first seasonality and then hours by about 25%, holding the remaining variables constant. The gains in income are roughly comparable and fairly substantial in both tests. In the case of seasonality, both earnings and the employment ratio are affected while, for hours, the impact is concentrated on the earnings variable.

One possible anomaly in these results should be pointed out. In both tests per capita incomes increase, but the growth in employment declines, resulting in turn in an increase in the rate of out-migration.^{116/} On the one hand, this could reflect a possible trade-off between income and employment in the Atlantic Region. The rationale here is that higher earnings and income force entrepreneurs to economize on labour, leading to fewer employment opportunities. On the other hand, this result could be the consequence of "specification error" to the extent that the dampening effect of increases in residentiary services on the growth in employment is overstated.^{117/}

In Test (E-1), the value for the percentage of females in the employed labour force was increased by about a quarter. Small income gains resulted, reflecting the advance in the participation rate which offset the small decrease in earnings. The fall in overall productivity would seem to be an intuitively plausible result -- Women's Lib notwithstanding -- but the decline generated here is probably too small to be significant.

In the final three sensitivity tests, shocks were given to the three change over time variables. In general, the results seem reasonable. The increase in the growth in

the working population leads to a decrease in the employment ratio -- reflecting a "softer" labour market -- generating in turn a fall in per capita income. Similar results are realized from a strong decrease in the rate of net out-migration. In contrast, moving the growth in employment upward leads to some improvement in the labour market -- although less than might be anticipated because of the resulting advance in GNP -- and to an increase in earnings (reflecting for the most part the decrease in RS). These work together to generate a small advance in per capita income.

The sensitivity tests placed into sharper focus the limitations inherent in the estimation technique used in the industry block, related in particular to the inclusion of an industry residual identity. They also suggested areas where changes in equation specification might be required. In general, however, the results from the sensitivity tests tended to increase the confidence of the author in the structure of the model (and hopefully they will have a similar effect on the reader). For the most part, the results seem to be economically plausible, and add further weight to the author's belief that a cross-section model of this type can tell us a great deal about the income determination process in the Atlantic Region.

(c) Alternative Scenarios

The author then developed some alternative "scenarios" for the Atlantic Region. Relative to the sensitivity tests, these involved alterations to a number of parameters. The results are outlined in Table 3. The first five are only illustrative in nature while the last four represent an attempt to analyze how well the model predicts probable developments in the Atlantic Region over the last decade.

TABLE 3

ALTERNATIVE SCENARIOS

	1961 CONTROL SOLUTION	(A)	(B)	(C)	(D)	(E)
RIC	659.61	802.94	792.56	803.92	781.26	769.74
ER	27.09	28.89	28.57	30.14	28.82	28.16
ELFR	94.35	95.62	95.32	94.41	94.32	94.28
W	2,261.57	2,706.18	2,691.04	2,607.88	2,623.57	2,631.52
DIJ	0.876	0.839	0.848	0.770	0.825	0.853
UR	5.65	4.38	4.68	5.59	5.68	5.72
PR	46.14	48.54	48.16	51.30	49.10	48.00
SE	21.91	15.44	15.62	15.44	15.85	16.05
HO	8.03	6.85	7.11	7.08	7.56	7.80
PREF	21.98	22.99	22.86	28.88	25.67	24.06
GLFP	109.26	117.12	116.41	118.10	116.19	115.24
NMR	5.90	3.33	2.50	4.49	2.24	1.12
GE	101.30	114.50	113.32	116.15	112.94	111.34
URB	35.21	49.04	46.77	80.22	59.57	49.22
LED	20.60	24.06	23.71	29.42	26.01	24.31
HOBS	10.77	12.87	12.48	18.23	14.65	12.85
OHVAM	3.10	4.08	3.94	8.91	6.16	4.79
RS	14.70	15.64	15.54	16.85	16.01	15.58
EH	8.37	9.24	9.08	11.46	9.97	9.23
LVAM	8.31	6.45	6.44	7.76	7.07	6.73
IR	3.76	0.68	3.48	-5.25	4.09	8.78

TABLE 3 (CONT'D)

ALTERNATIVE SCENARIOS

1961 CONTROL SOLUTION		MEAN VALUES	1971 SIMULATION			
			(F)	(G)	(H)	(I)
EIC	659.61	989.57	775.14	845.07	908.03	912.00
ER	27.09	30.04	30.13	31.20	31.29	31.73
ELER	94.35	95.66	94.69	94.53	94.46	94.61
W	2,261.57	2,979.00	2,495.58	2,684.59	2,920.84	2,898.61
DII	0.876	--	0.849	0.798	0.802	0.785
UR	5.65	4.34	5.31	5.47	5.54	5.39
PR	46.14	46.18	46.79	48.53	48.70	49.32
SE	31.91	--	17.68	17.68	17.51	17.32
HO	8.03	--	8.29	7.83	7.84	7.71
PFEUF	21.98	27.58	21.98	26.65	26.26	27.10
GLFP	109.26	112.98	106.34	105.92	107.31	107.98
NMR	-5.90	17.95	-3.49	-3.98	-2.34	-1.55
GE	101.30	117.39	103.71	103.01	105.35	106.47
URB	35.21	42.20	39.21	58.87	57.04	62.99
LED	20.60	31.04	22.49	31.22	30.78	31.68
HOB	10.77	14.77	11.55	14.92	14.88	15.84
OHVAM	3.10	5.79	3.71	6.35	5.98	6.80
RS	14.70	15.15	15.84	17.45	16.82	17.02
EH	8.37	11.00	8.68	10.08	10.06	10.46
LVAM	8.31	8.44	8.00	8.36	7.38	8.12
IR	3.76	3.14	10.50	1.12	3.18	0.64

SCENARIO (A): AG from 8.41 to 4.21, FOR from 5.59 to 2.80, FISH from 6.69 to 3.35, PA from 10.23 to 14.23, TRANS from 8.25 to 13.25, PP from 2.37 to 3.73.

SCENARIO (B): Same as (A) for AG, FOR, FISH and PP; PA from 10.23 to 13.23, TRANS from 8.25 to 12.25.

SCENARIO (C): Same as (A) for AG, FOR, FISH and PP; add 4.0 to intercept in HOBS equation, and 3.0 to intercept in OHVAM equation.

SCENARIO (D): Same as (A) for AG, FOR, FISH and PP; add 2.0 to intercept in HOBS equation, and 1.5 to intercept in OHVAM equation.

SCENARIO (E): Same as (A) for AG, FOR, FISH and PP; add 1.0 to intercept in HOBS, and add .75 to intercept in OHVAM.

SCENARIO (F): 1971 Mean Values for the X variables.

SCENARIO (G): 1971 Mean Values for the X variables; add 200 to intercept in W equation, add 5.0 to intercept in LED equation, add 1.0 to intercept in OHVAM equation.

SCENARIO (H): 1971 Mean Values for the X variables; add 400 to intercept in W equation, add 5.0 to intercept in LED equation, add 0.5 to intercept in HOBS equation, add 1.0 to intercept in OHVAM equation, deduct 1.0 from intercept in RS equation, add 0.5 to intercept in LVAM equation.

SCENARIO (I): 1971 Mean Values for the X variables; add 400 to intercept in W equation, add 5.0 to intercept in LED equation, add 1.0 to intercept in HOBS equation, add 1.5 to intercept in OHVAM equation, deduct 1.0 from intercept in RS equation, add 1.0 to intercept in LVAM equation.

and where the 1971 means for the exogenous variables are as follows:

AG	(%)	3.00	PFR	(%)	18.75
FOR	(%)	3.59	MIN	(%)	1.74
FISH	(%)	6.00	NRI	(%)	16.67
WPR	(%)	68.01	NFD		.22
PA	(%)	9.23	POP		44.273
TRANS	(%)	8.25	UD		.15
CON	(%)	2.40	PP	(%)	2.50

The alterations involved in Scenarios (A) and (B) were limited to the exogenous industry variables. Reductions of one-half are assumed for the three primary industries, with the shares being distributed among three industries, public administration, transportation and pulp and paper. In Scenario (A), all of the decreases in share are redistributed, while in (B), two per cent is left undistributed in order to move the industry residual closer to the 1961 control solution figure. As would be anticipated from the sensitivity tests, this redistribution results in a substantial advance in income and earnings. For Scenario (B), the increase is around a fifth for both variables. All of the labour force utilization measures have also improved significantly, and urbanization and education showed their anticipated increases. Further, the three growth measures moved sharply upward, with net

3

5

OF / DE

migration going from negative to positive. Finally, reasonable changes are indicated for the five endogenous industries, with all but LVAM showing a small increase.

For the reasons outlined earlier, greater difficulties were encountered when attention was focussed on stimulating employment in two endogenous industries, HOBS and OHVAM (in place of PA and TRANS). Three solution sets were computed with the middle one -- Scenario (D) -- providing the most realistic results. Comparing Scenarios (D) and (B), the income gains in the former are somewhat smaller, the result in part of the greater increase in residentiary services associated with the larger advances in HOBS and OHVAM. As noted earlier, this may reflect a weakness in specification in the RS equation. At the same time, this result points out the possibility that the income gains from stimulating employment in higher wage activities may be less than anticipated unless, at the same time, productivity gains can be encouraged in the generally lower wage residentiary activities.

In Scenarios (F) through (I), attempts were made to simulate the 1961 to 1971 growth performance of the Atlantic Region. Unfortunately, full 1971 Census data is not yet available. Particularly lacking is full labour force information^{118/} such as industrial distribution, labour force

utilization and so on. Therefore, estimates had to be made based on the highly aggregated industry data, plus the labour force utilization information, provided by the Labour Force Survey of Statistics Canada. Further, employment income per capita and earnings were estimated from the National Accounts data. The estimates were generally derived by applying the 1961-71 percentage point changes -- or in the case of Y and W the percentage increase deflated by the increase in prices^{119/} -- to the 1961 means of the exogenous and endogenous variables. Further adjustments were required to take account of the fact that the 1971 figures are "weighted" averages^{120/} whereas the 1961 figures are simple unweighted county averages. Estimates could be developed for all exogenous variables and all but three of the endogenous variables; thus, the model could be solved. It should be emphasized, however, that these estimates are only indicative of rough orders of magnitude.

Scenario (F) simply involved plugging in the 1971 exogenous variables into the 1961 model. This lead to a good "prediction" for the employment ratio, but the predicted values for earnings and therefore per capita income were well below the actual 1971 values. Perhaps most importantly, the solution set is not "feasible" in the sense that the industry residual is much too high. These results though were far from disappointing, since this scenario fails to take account

of two considerations. First, overall productivity gains were experienced over the last decade, that is, productivity gains over and above those associated with changes in industry mix. This implies changes to the earnings equation. Second, government policy interventions such as education expenditures and industrial incentives programs probably brought about important alterations to the structure of the Atlantic economy over the last decade.

Scenario (G) represents an attempt to take account of some of these issues. The alterations have been limited to so-called "additive" factors.^{121/} First, 200 was added to the intercept in the wage equation, implying approximately a one per cent per year advance in overall productivity as defined above (likely an under-estimate). Second, 5.0 was added to the intercept in the education equation; third, 1.0 was added to the constant term in the OHVAM equation, to take account of the impacts of ADIA, RDIA and the provincial incentives programs. The results for this scenario exhibited the expected characteristics, with more favourable values for most variables relative to Scenario (F) (and of course the 1961 control solution). Still, earnings and income continued to be under-estimated; further, the "prediction" for RS was much too high, and that for EH too low.

Accordingly, further fine tuning seemed warranted. Two slightly different scenarios were computed, Scenarios (H) and (I). The results were roughly comparable, but Scenario (H) represents the more feasible set. The alterations made in (H) were as follows: first, 400 was added to the W equation intercept, implying overall productivity gains of about two per cent per year; second, 0.5, 1.0 and 0.5 were added to the intercepts in the HOBS, OHVAM and LVAM equations respectively, in order to capture both the direct and indirect effects of incentives programs; third, 1.0 was deducted from the intercept in the RS equation, to take account of probable productivity improvements in that sector; finally, 5.0 was added to the intercept in the LED equation to take account of increases in education.

The results for Scenario (H) are quite encouraging. Even with further fine-tuning, the model continues to underestimate earnings and per capita income, but in both cases it does cover two-thirds or more of the distance between the actual 1961 and 1971 values. This is a fairly creditable performance for a cross-section model. The model also does fairly well with respect to labour force utilization, with an underestimate in ELFR offsetting an over-estimate in PR. Finally, with the inclusion of the "add" factors, the predictions for the level of education and the five endogenous industries are all fairly close to their actual values.

There are though two areas where the model has not performed too well. First, the results in the change over time block are quite far off the mark; moreover, only for GE is the prediction in the right direction. In the case of the growth of employment, the suggestion is that not all of the high growth industries in the Atlantic during the 1951-61 period remained in that category over the 1961-71 decade (and similarly for the low growth industries). For example, this is doubtless true for public administration and more specifically military employment where cutbacks occurred during the sixties ~~coupled~~ with substantial increases during the preceding period. Also noteworthy is that the model "predicts" a much lower rate of out-migration for the 1961-71 period, whereas the rate of out-migration was in fact higher than in the 1951-61 period. The implication is that, for a given economic structure and growth in employment, the Atlantic people showed a higher propensity to migrate out of the Region during the sixties relative to the fifties. This is probably the anticipated result, given the improvements in transportation and communications and the retraining and resettlement programs of the Department of Manpower and Immigration.

The second drawback indicated by Scenario (II) is that the model is over-estimating the level of urbanization in the Atlantic. In this case however, the predicted value may be closer to the truth than the "actual" value. Many analysts have noted the large increase in recent years in ribbon development and exurban sprawl surrounding the larger urban centers in the Region. People living in these areas are probably categorized as "rural non-farm" in the Census, whereas in terms of economic function they are urban. Also, because distances tend to be short and because of recent improvements in the highway network, the amount of commuting from rural home to urban job is probably on the increase. In short, in terms of economic function, the level of urbanization is probably higher than the actual figures would suggest.

Finally, the analyst should neither be surprised nor discouraged by the fact that this cross-section model does not totally predict the course of the Atlantic economy over the last decade. Undoubtedly, basic structural changes have occurred over this period, implying alterations to the coefficients in the model's structural form. For example, reflecting compositional changes during the sixties, secondary manufacturing (i.e. OHVAM) may be closer to a "high wage" activity than was the case a decade ago. Such changes

can only be fully taken into account through the use of "multiplicative" factors, i.e. changes to the β coefficients in the structural form. In the final analysis however, the best approach would be to re-estimate the model using 1971 data and then compare the two estimations. For the present exercise, it is perhaps sufficient that a cross-section model of this type both predicts the direction of change over the next decade and covers a fair proportion of the distance between the actual 1961 and 1971 values. From this point of view, the present model is generally successful.

(d) Ontario and Canada Simulations

As a final exercise, simulations were run using the 1961 averages for Ontario and Canada in order to analyze the extent to which the endogenous values for Ontario and Canada can be "predicted" from an Atlantic model. Essentially, the objective is to analyze whether or not the Atlantic economy is simply a part of a larger economic system; that is, whether or not the coefficients in the structural form of the Atlantic model are not markedly different from those for a similar model that might be estimated for Ontario or the total country. If this were the case, the implication would be that lower earnings and incomes in the Atlantic are largely explained by

the Region's unfavourable industry mix, with intra-industry productivity differentials and economies of scale and agglomeration, among other factors, playing relatively minor roles. The results of these simulations are outlined in Table 4.

Before continuing, however, it should be noted that the averages in the table greatly exaggerate the size of the gaps between the Atlantic and the remainder of the country, since the Ontario and Canada means are "weighted" averages whereas the Atlantic averages are unweighted. This in turn reflects the tendency for the more populated counties to also be the more highly developed, for example, Halifax county in Nova Scotia, York county (i.e. Toronto) in Ontario. This difference between the data sets may distort the results to some extent, but it should not totally negate the usefulness of the exercise.

TABLE 4

CANADA AND ONTARIO SIMULATIONS

CANADA SIMULATION								
	ATLANTIC CONTROL SOLUTION	MEAN VALUES	(A)	(B)	(C)	(D)	(E)	(F)
EIC	659.61	1,158.00	72.58	784.55	802.23	891.82	904.10	919.13
ER	27.09	34.32	29.65	30.39	31.49	30.68	30.77	31.19
ELFR	94.35	96.14	94.32	94.37	94.43	93.92	93.94	93.98
N	2,261.57	3,191.00	2,522.45	2,513.25	2,499.61	2,903.41	2,950.05	2,972.76
DAI	9.876	0.755	0.859	0.827	0.779	0.798	0.765	0.761
UR	5.65	3.86	5.68	5.63	5.57	6.08	6.06	6.02
PR	46.14	54.04	47.60	48.76	50.49	49.46	49.60	50.25
SE	21.91	12.93	16.91	16.71	16.42	17.29	17.10	16.75
HO	8.03	7.49	8.47	8.22	7.85	7.95	8.14	8.04
PELFP	21.98	27.61	24.55	26.21	28.69	28.93	27.18	27.47
GLFP	109.26	121.07	111.05	112.19	113.90	109.61	114.56	116.30
NMR	-5.90	7.74	-3.58	-2.25	-0.24	-5.28	0.54	2.58
GE	101.30	120.19	105.48	107.39	110.26	103.06	111.37	114.29
URB	35.21	69.63	35.89	47.66	65.28	64.35	70.95	71.95
LED	20.60	29.96	24.51	26.28	28.92	29.22	27.34	27.61
HOBS	10.77	17.11	12.49	14.33	17.08	16.30	14.96	15.74
OHVAM	3.10	11.99	4.24	5.90	8.38	8.53	10.52	10.39
RS	14.70	14.40	15.53	15.99	16.69	18.16	16.87	16.32
RH	8.37	8.87	8.57	9.33	10.46	10.14	9.59	9.91
LVAM	8.31	8.15	6.64	7.02	7.58	6.89	5.77	5.28
FR	3.76	3.55	16.61	11.51	3.88	4.04	6.36	6.44

TABLE 4 (CONT'D)

CANADA AND ONTARIO SIMULATION

	ATLANTIC CONTROL SOLUTION	ONTARIO SIMULATION				
		MEAN VALUES	(G)	(H)	(I)	(J)
EIC	659.61	1,380.00	827.34	840.14	865.28	949.03
ER	27.09	37.27	30.56	31.31	32.83	31.62
ELFR	94.35	96.66	94.15	94.19	94.28	93.75
W	2,261.57	3,419.00	2,666.35	2,656.64	2,637.51	3,051.83
DII	0.876	0.653	0.847	0.815	0.750	0.775
UR	5.65	3.34	5.85	5.81	5.72	6.25
PR	46.14	56.87	47.87	49.03	51.36	49.74
SE	21.91	11.26	13.58	13.38	12.99	13.97
HO	8.03	8.18	8.59	8.34	7.85	8.07
PFELE	21.98	29.14	24.69	26.37	29.72	29.13
GLFP	109.26	123.69	112.66	113.77	116.00	111.14
NMR	-5.90	14.90	-0.03	1.28	3.90	-1.81
GE	101.30	125.60	108.05	109.91	113.66	105.10
URB	35.21	77.34	39.71	51.58	55.25	68.46
LED	20.60	33.16	26.28	28.07	34.63	31.06
HOB	10.77	17.69	13.55	15.40	19.10	17.41
OHVAM	3.10	17.98	4.86	6.53	9.86	9.19
RS	14.70	14.87	16.03	16.51	17.47	18.71
EH	8.37	8.48	8.75	9.52	11.05	10.35
LVAM	8.31	7.32	6.30	6.69	7.47	6.59
IR	3.76	2.39	20.24	15.07	4.78	7.48

TABLE 4 (CONT'D)

CANADA AND ONTARIO SIMULATION

ONTARIO SIMULATION						
	ATLANTIC CONTROL SOLUTION	MEAN VALUES				
		(K)	(L)	(M)	(N)	
EIC	659.61	1,380.00	963.33	961.56	977.06	1,020.51
ER	27.09	37.27	32.38	31.71	32.55	32.18
ELFR	94.35	96.66	93.79	93.76	93.80	93.55
W	2,261.57	3,419.00	3,041.21	3,092.32	3,079.03	3,278.85
DII	0.876	0.653	0.742	0.753	0.712	0.699
UR	5.65	3.34	6.21	6.24	6.20	6.45
PR	46.14	56.87	50.91	49.88	51.17	50.74
SE	21.91	11.26	13.78	13.77	13.59	14.12
HO	8.03	8.18	7.82	8.25	7.99	8.07
PFELF	21.98	29.14	30.86	29.18	29.23	29.25
GLFP	109.26	123.69	112.21	116.09	117.51	116.32
NMR	-5.90	14.90	-0.56	4.00	5.67	4.28
GE	101.30	125.60	107.29	113.80	116.19	114.20
URB	35.21	77.34	80.50	75.08	89.99	95.77
LED	20.60	33.16	32.89	29.18	31.13	31.19
HOB9	10.77	17.69	19.29	16.07	18.04	17.44
OHVAM	3.10	17.98	10.88	11.18	13.44	14.93
RS	14.70	14.87	19.22	17.43	18.02	18.88
EH	8.37	8.48	11.13	9.80	10.61	10.36
LVAM	8.31	7.32	7.01	5.48	5.97	5.73
IR	3.76	2.39	2.20	9.78	3.64	2.39

- Simulation (A): Canadian mean values for the X vector.
- Simulation (B): Canadian mean values for the X vector; plus add 1.0 to intercept in HOBS equation, and 1.0 to intercept in OHVAM equation.
- Simulation (C): Canadian mean values for the X vector; plus add 3.0 to intercept in HOBS equation and 2.5 to intercept in OHVAM equation.
- Simulation (D): Canadian mean values for the X vector; plus add 400 to intercept in W equation, 1.0 to intercept in HOBS equation; and 2.0 to intercept in OHVAM equation.
- Simulation (E): Canadian mean values for the X vector; plus add 400 to intercept in W equation, 1.0 to intercept in HOBS equation, 5.0 to intercept in OHVAM equation and deduct 1.0 from intercept in RS equation.
- Simulation (F): Canadian mean values for the X vector; plus add 400 to intercept in W equation, add 2.0 to intercept in HOBS equation, add 5.0 to intercept in OHVAM equation, and deduct 1.5 from intercept in RS equation.
- Simulation (G): Ontario mean values for the X vector.
- Simulation (H): Ontario mean values for the X vector; plus add 1.0 to intercept in HOBS equation, and 1.0 to intercept in OHVAM equation.
- Simulation (I): Ontario mean values for the X vector; plus add 3.0 to intercept in HOBS equation, and 3.0 to intercept in OHVAM equation.
- Simulation (J): Ontario mean values for the X vector; plus add 400 to intercept in W equation; 1.0 to intercept in HOBS equation, and 2.0 to intercept in OHVAM equation.
- Simulation (K): Ontario mean values for the X vector; plus add 400 to intercept in W equation, 2.0 to intercept in HOBS equation, and 3.0 to intercept in OHVAM equation.

Simulation (L): Ontario mean values for the X vector; plus add 400 to intercept in W equation, 1.0 to intercept in HOBS equation, 5.0 to intercept in OHVAM equation and deduct 1.0 from intercept in RS equation.

Simulation (M): Ontario mean values for the X vector; plus add 400 to intercept in W equation, 2.0 to intercept in HOBS equation, 6.5 to intercept in OHVAM equation, and deduct 1.0 from intercept in RS equation.

Simulation (N): Ontario mean values for the X vector; plus add 600 to intercept in W equation, 1.0 to intercept in HOBS equation, 7.64 to intercept in OHVAM equation, and deduct 1.0 from intercept in RS equation.

and where the 1961 values for the exogenous variables are as follows:

		CANADA	ONTARIO
AG	(%)	9.90	7.05
FOR	(%)	1.68	0.75
FISH	(%)	0.56	0.10
WPR	(%)	66.04	67.80
PA	(%)	7.46	7.57
TRANS	(%)	6.22	5.01
CON	(%)	6.66	6.43
PFR	(%)	28.09	6.82
MIN	(%)	1.88	1.78
NRI	(%)	22.45	20.74
NFD		.04	.00
POR		76.631	115.483
UD		.05	0
PP	(%)	1.57	1.58

In simulations (A) and (G), the 1961 values for the exogenous variables were simply plugged into the Atlantic model, while in (B) and (C) in the Canada simulations, and (H) and (I) in the Ontario simulations, small additive adjustments were included in order to (1) learn the properties of the model when different sets of exogenous values are applied, and (2) bring the solution values for the industry residual closer to their actual values. In all six solutions, the "predicted" endogenous values fall far short of their actual magnitudes, but the changes are consistently in the correct direction and, more important, generally cover a significant proportion of the distance between the Atlantic means and the Canada -- or Ontario -- values. Also, the predicted endogenous values are generally more favourable in Ontario than in the country as a whole, which is the desired result.

In summary however, the findings suggest that the basic structure of the Atlantic economy -- as summarized by industry productivity levels, economies of scale and agglomeration, and linkages between industries -- differs from that in the remainder of the country. The implication is that differences in natural resource endowment, distances from major markets, and perhaps topography, culture and history, have probably resulted in important divergences in basic economic

structure between the Atlantic and Canada as a whole. The further implication that could be drawn from this is that simply replicating the Ontario industry mix may not necessarily lead to the highest possible incomes for the Atlantic Region.

In the remaining simulations, experiments were made to investigate the extent to which the gaps could be closed through further additive adjustments, particularly to the wage equation. The endogenous values did move closer to their counterpart Canadian and Ontario means, but in general fairly large gaps remained. The indications are therefore that more fundamental adjustments are required, involving changes to the coefficients in the structural form (i.e. multiplicative adjustments).

To sum up, in most respects, the model seems to be working fairly well. Both its structure and solution results provide useful insights into the income determination process in the Atlantic Region. Also, the solution results have indicated that important implications for government policy can be derived from a cross-sectional model of the form estimated in this project. Finally, the model generally does a good job in predicting the direction of change and a significant proportion of the required magnitude. The evidence strongly suggests therefore that a similar model estimated for 1971 should prove to be a useful policy tool in the years ahead.

11. FINAL CONCLUSIONS

This final section will summarize the major conclusions that can be drawn from the estimation and solution results developed during the course of this exercise. Many of these have been noted previously in the text but are of sufficient importance to be noted again in this concluding section. The chapter will end with a brief discussion of areas for further research and/or of "roads not taken" during this exercise because of time constraints.

First, note should be made of the major conclusions regarding the econometrics applied during this work. These are largely drawn from the attached appendices.

(1) Comparisons between the simple correlation (Appendix C) and multiple regression (Appendix E) results pointed out as many differences as similarities. Clearly then, simple correlation can offer very misleading results and should only be applied as a last resort, i.e. when data or other constraints rule out the use of more sophisticated techniques.

(2) Principal components (Appendix F) proved to be a very useful analytical device, particularly in pointing out linkages between economic activities. However, in the present

exercise, their usefulness in regression analysis was minimal, both because of the difficulties encountered in interpreting the estimated coefficient on a component, and because the components taken together explained no more of the variability in the dependent variable than the actual variables themselves.

(3) In comparisons between the Ordinary Least Squares and Two Stage Least Squares results (Appendix G); it became very clear that single equation estimation bias can be a very serious problem in a cross-section model. The implication is that applying OLS in a cross-section model can lead to highly unreliable results and thus should be avoided if at all possible. There are therefore no easy "short-cuts" in estimating a cross-section model; the application of simultaneous equation estimation techniques must be viewed as a necessity. This diverges markedly from the conclusions of analysts working with time-series models, for in these cases the OLS and 2SLS results were generally quite comparable.

(4) Moreover, Two Stage Least Squares generally provided "neater" results in the sense that the final 2SLS specifications generally included fewer explanatory variables than the final OLS specifications (Appendix G). In some instances, the

suggestion was that the variability shared among two or more variables in the OLS results tended to become concentrated in a single explanatory variable under 2SLS estimation.

Of greater importance are the economic conclusions that can be derived from the estimation results -- both single and simultaneous equation -- and the solution results from the econometric model. Some of these are general in nature while others refer to specific industries and other variables.

(1) The estimation and solution results for the model indicate that: (i) a cross-sectional model of this type can provide very useful insights into the wage and income determination process in the Atlantic Region and (ii) a similar model based on 1971 data should prove to be a useful policy tool for the Department of Regional Economic Expansion and other government departments interested in regional development. While improvements could doubtless be made to some equations, the specification of the model seems essentially sound and for the most part is consistent with (i) the hypotheses offered by the theory and literature, and/or (ii) known characteristics of the Atlantic economy. Also, no mathematical problems were encountered in solving the model.

while the solution results were generally reasonable in an economic sense, abstracting from the problems with the industry block.

In the 1971 Atlantic scenario and the Ontario and Canada simulations, it was found that the model -- with some "fine tuning" -- generally predicted the correct direction of change and a significant proportion of the required magnitude. The Ontario and Canada simulations also suggested that the structure of the Atlantic economy differs significantly from the economic structure of the remainder of the country, but this result was far from unexpected.

Perhaps most importantly, the structure of the model and the solutions based on it show very clearly the "general equilibrium" nature of the economic relationships underlying the income determination process and, in particular, how the alteration in one variable can lead, either directly or indirectly, to important changes in a wide range of other variables. This indicates that, in implementing a regional development policy (involving an alteration to an "instrument" variable), attention should be focussed not only on the "target" variable but also other variables, for the favourable changes in the target variable may be offset -- in total or in part -- by unfavourable changes in the others. The implication is that more than one policy initiative -- involving

changes in more than one instrument variable -- may be required before the desired or "optimum" results can be achieved.

(2) The estimation and solution results strongly illustrate the all pervasive role played by industrial structure in the generation of earnings and income differentials in the Atlantic Region. Industry mix has a strong impact both directly through the wage equation and indirectly in particular through the manpower utilization equations. It can be concluded therefore that a county's relative position with respect to earnings and income is determined in large part by its industrial structure.

These results would seem to be at least partially consistent with "traditional" economic theory (and/or the so-called competitive hypothesis), for they suggest an important role for inter-industry differences in production functions in the generation of intercounty earnings differentials. It should be emphasized however that these industry effects may also reflect unionization and other "non-economic" forces that are industry specific.

Finally, the findings suggest that the emphasis of regional planners on industrial structure has not been misplaced. At the very least, our results are consistent with the planner's working hypothesis that industry mix represents

the most appropriate vehicle for breaking into the "circle of causation" underlying regional disparities.

(3) On the other hand, the results for some industries -- particularly with regard to estimation -- are surprising and perhaps disquieting. Other high value added manufacturing, failed to enter the earnings equation with its anticipated positive coefficient, indicating that as of 1961 secondary manufacturing was not a high-wage industry in the context of the Atlantic Region. Further to this, the early OLS results suggested that this high-wage role had been taken over by certain segments of the tertiary sector, in particular higher order business services. This hypothesis, however, did not hold up under Two Stages Least Squares estimation. The upshot is that, in the final specification of the earnings equation, only one industry has a positive coefficient. Therefore, the positive influences of industry mix on earnings are mainly indirect in nature, through in particular the seasonality and hours variables.

As noted previously in the text,^{123/} the lack of positive industry coefficients may reflect one of three considerations: first, it could be based on specification error; second, it could result from the fact that all of the industries not included in the equation are paying

slightly above average wages, but that none are far enough above the average to be picked up in regression analysis; third, for the missing industries (particularly in the tertiary sector), within industry earnings differentials between countries (related perhaps to intra-industry production function differences) may be too large to allow a significant coefficient to be estimated -- the implication is that an industry may be a low-wage activity in one part of the Atlantic and a high wage activity in another. The final consideration suggests that, despite our overall success with industrial structure in estimating the model, this approach may break down with respect to particular industrial activities, especially in the non-goods producing sectors.

Placed in a regional planning context, the implications of the above are two-fold. On the one hand, the structure of the model indicates clearly that shifting employment out of the primary industries and primary manufacturing can lead to substantial gains in earnings and income in the Atlantic Region. On the other hand, the model's structure offers few solid answers concerning which industrial activities should receive government encouragement (although some implications can be drawn from the non-wage equations in the model). Fortunately, some of these questions were resolved once the model was solved. The conclusions based on the solution results are outlined below.

(4) The structure of the model indicates that some industries can have conflicting effects on per capita income and to a lesser extent earnings. For example, the coefficient on agriculture in the wage equation is negative, reflecting its below average productivity, but agriculture has a positive impact on the employment ratio through the participation rate and unemployment rate equations. The net effect of these conflicting tendencies -- which turned out to be negative -- could only be discerned through solving the model. The model's structure then can only be a partial guide to the characteristics of the economy being modelled.

(5) Very few socio-cultural variables have been successfully introduced into the model, but one does appear in three different equations, the percentage of Francophones in the total population. It is interesting to note that, in two out of three cases, the implications to be drawn from the PFR coefficient are contrary to conventional wisdom, for the coefficients in both the net migration rate and per cent female in the employed labour force equations suggest that the Francophone community in the Atlantic is more responsive to economic stimuli than the Anglophone community. The solution results indicated that the net effect of ethnicity on income was negligible.

(6) Some surprises were attached to the estimation and solution results for other non-industry variables in the model. In particular, the structure of the model indicated a much smaller role for urbanization and population size than was originally anticipated. More importantly, the solution results suggested that sharp increases in urbanization and population size will have only a small effect on average per capita incomes. Two different conclusions could be drawn from this. One possibility is that the economies of agglomeration had not yet developed in the Atlantic Region as of 1961; the implication of this is that, given probable developments in the Atlantic Region over the last decade, the results for a 1971 Atlantic model would be quite different, or, at the very least, the results from a model for Ontario would diverge considerably from the above. A second possibility is that the importance of the economies of agglomeration have been greatly exaggerated, at least in relation to per capita income. The author leans toward the former conclusion but the latter cannot be totally ignored.

(7) The solutions hint at possible trade-offs between average income and the size distribution of income, but a better specification for the degree of income inequality equation -- together with a better measure of income concentration -- would likely be required before the solution results could be viewed as more than illustrative.

/...

(8) Previously, it was noted that a model of this type can provide some useful insights into the income determination process. At the same time, because the data and the model are thirteen years out of date, it would likely be a mistake to derive "hard and fast" policy prescriptions from the model discussed here. Still, a number of possible policy implications should be noted:

(i) the model indicated that strong economic gains can be realized from shifting employment out of the primary industries, but it matters greatly which industries are expanded;

(ii) for example, the solution results suggest that little or no gains in income accrue from shifting employment into primary manufacturing; supporting such activities must be justified with respect to other objectives, such as, improving productivity in the sector (and thus changing the coefficients on LVAM in the model) or to prevent depopulation; unfortunately, with respect to the latter, the model suggests only a very weak positive association between per capita income and population size;

(iii) similarly, our results indicate that relatively small economic advances are realized from supporting secondary manufacturing of the form already present in the Atlantic; the main emphasis therefore should be placed on re-structuring secondary manufacturing in the Atlantic Region (and thereby increasing the coefficients on OHVAM in the model);

(iv) fairly substantial income gains are suggested from encouraging certain types of service industries, in particular, transportation, higher order business services, and public administration (perhaps through the decentralization of federal government employment); therefore, to the extent that some service industries are sufficiently "foot-loose" to be attracted to a given location, our results suggest that industrial incentives programs and industry promotion campaigns should not stop with secondary manufacturing;

(v) our experiences with the residentiary services variable suggest that, together with supporting higher-wage industry, attempts should also be made to improve productivity in lower wage activities;

(vi) the solution results indicate that, even with no alterations in industrial structure, fairly substantial economic gains can be realized from lowering the degree of seasonality in the work force and/or increasing the average number of hours worked; accordingly, lowering the degree of underemployment in the Atlantic should be viewed as an important instrument variable;

(vii) the results suggested that only small gains in income accrue from increasing the level of education, unless explicit attempts are also made to alter industrial structure and thus the demands for labour;

(viii) efforts at lowering the rate of net out-migration will only lead to lower per capita income unless this is done in conjunction with job creation programs and changes in industrial structure.

This paper will close with a brief discussion of areas for further research and/or issues that would have been investigated further in this exercise if time had permitted.

(1) The author would have preferred that there be sufficient time to investigate a number of respecifications of equations, and reformulations of the variables. A partial listing would include:

- the inclusion of further variables in the residentiary services equation to dampen the growth in RS when the model was solved;

- reformulation of the $HOBS$ variable in order to find a measure that would stand up under 2SLS estimation;

- further investigation of the less comprehensive measure for urbanization;

- the use of the GINI coefficient to measure the degree of inequality in income;

- moving all or part of lower value added manufacturing to the exogenous variables;

- inclusion of other variables in the growth in employment equation in order to better pick up the effects of differential changes in comparative advantage;

It is recommended that all of these respecifications and reformulations -- together with others noted in the text -- be investigated if a 1971 model is estimated.

(2) Our difficulties with the industry block when the system was solved indicate that all endogenous industry variables should be determined by behavioural equations. This would require the application of "Restricted Least Squares" along the lines discussed in the Tobin - Brainard paper (6).

(3) It is recommended that a set of principal components derived from all relevant variables (i.e. industry, occupation and "other") be computed before attempting to specify the 1971 model. One possibility haunting the present author is that a key variable has, through oversight, been left out of the model. Such a set of principal components would minimize this danger. These results could also be of great assistance in equation specification.

(4) Some insights into the more dynamic aspects of the income determination process could be discerned by using the growth in median earnings equation in Appendix E as the basis for the construction of a "change over time" model of the Atlantic Region, using changes from 1961-71 by county as the variables in the model.

(5) Some dynamic elements could be added to the cross-sectional model outlined in this thesis by (i) comparing the 1961 and 1971 Atlantic models once the latter is estimated, and/or (ii) comparing the Atlantic model with models estimated for the other regions in Canada (particularly Ontario) and the country as a whole. Such comparisons should provide some useful insights regarding the future course of the Atlantic economy. This of course would be a highly ambitious project and would certainly be beyond the scope of a single researcher.

FOOTNOTES

1/ The major Canadian studies are those of Denton (17) and Chernick (8) conducted for the Economic Council. These will be discussed later in the report.

2/ That is, all members of the work force with previous work experience.

3/ A preferable measure of the size distribution of income would likely be the Gini coefficient but the author was not able to acquire this data, while it was concluded that its computation would be too time consuming for the present project.

4/ Later in the work, the problem of multicollinearity was solved to some degree by adding together industry variables which were highly related, both statistically and economically. However, the problem of insufficient degrees of freedom constantly plagued the regression work and provided one of the arguments for the development of a full econometric model.

5/ As will be outlined in the literature section, Holmes and Munro (29) developed a small cross-section model but this included only four equations and almost no industry detail.

6/ See pages 19-26 of the present study.

7/ One methodological point raised by the Chenery study is the possibility that a log-linear specification should be preferred in estimating the relationship between per capita income and industrial structure. The author's original intention was to attempt some log-linear specifications, but this was ruled out by the large number of zero elements in the matrix of explanatory variables. This was particularly true of some of the industry variables, e.g. pulp and paper where only a relatively few counties possess any activity at all.

8/ Hanna argued that correlation analysis could not be applied because of the large number of different categories that required investigation, for example, in the case of industrial structure. He noted that, when the number of independent variables exceeded six, the computation became unwieldy. This however was written when the computer was in its infancy.

9/ See Technical Appendix D for a fuller description of the methodology.

10/ In fact, he calculated that the independent effects of industry composition covered only nine of the 72% of the total variation explained by industry, with the remainder being accounted for by what he calls 'correlated effects'. Related to the latter is the tendency for states with a favourable industrial structure (i.e. a structure skewed towards higher paying industries) to pay higher than average within-industry rates. Hanna suggests that this tendency can be easily explained by economic forces, specifically by high wage industries 'pulling up' the wages of low wage industries in states where the higher wage industries are predominant. The suggestion is therefore that a large proportion of the secondary or correlated effects should also be attributed to industrial structure.

11/ Both limitations make it very difficult to derive economic and policy implications from his work, for example, regarding which industries should be supported to reduce regional disparities.

12/ Kuznets (33 and 34) analyzed the association between industry structure and per capita income in a pair of articles published in 1957 and 1958. The first article dealt with international comparisons while the second looked at patterns within the United States. Both cross-section and time trend analysis was applied, while industrial structure was measured both in terms of income and employment. Emphasis here will be placed on his United States findings using the industrial distribution of employment, but in general there was close agreement.

between his international and intra-national results, and between his findings using the different industrial distribution measures. Because of the imprecise character of the estimates of the variables used in the Kuznets' analysis (particularly in his international study), he found it necessary to use direct comparison rather than simple or rank correlation techniques in his research.

- 13/ The association between per capita income and the share of employment accounted for by the S sector was particularly weak when industrial distribution was measured in terms of income.
- 14/ Kuznets defined secondary manufacturing to include clothing, machinery and equipment and miscellaneous manufacturing, while raw-materials oriented industries included food and tobacco, textiles, lumber, furniture, paper, chemicals, petroleum, and stone, clay and glass products. As will be outlined later, somewhat different definitions are used in the present analysis.
- 15/ As suggested in an earlier footnote, the correspondence between his cross-section and time-series findings provides a useful backdrop to the present study.
- 16/ This finding also points out the difficulty of conducting this type of analysis at the national level, where production functions -- and thus skill levels and capitalization -- within an industry can vary tremendously from one section of a country to another. For example, in this sense, agriculture in the Atlantic is a very different industry from agriculture in the Prairie provinces.
- 17/ Implicit here is the assumption that the tenets of neo-classical theory hold to at least some degree and thus a rough correspondence exists between output per worker and earnings per worker. As will be discussed later, this fundamental principle is now coming under increasing criticism.

18/

This point can perhaps best be illustrated with a numerical example. Assume a simple two-country, two sector and two period model with different but converging industrial structures and with the observed tendencies outlined in the text. (Rather than output per worker, the following could be assumed to be in terms of earnings per worker.)

	Period (0)		Period (1)	
	<u>Country</u>		<u>Country</u>	
	(1)	(2)	(1)	(2)
A Sector Share (%)	50	25	25	15
<u>Output Per Worker:</u>				
All Industries	100	200	200	300
A Sector	50	150	150	250
M & S Sectors	150	217	217	309

As the differences in industrial structure diminish, intra-sectoral productivity differentials decline markedly, from 300% to 167% in the A sector, and from 145% to 138% in the M and S sectors. While there might be some question concerning the direction of causation (although industrial structure as the explanatory variable seems to be more reasonable), the inter-relationship between industrial structure and intra-sectoral productivity differentials is readily apparent. This suggests that industrial structure can take on added significance when second and higher order effects are taken into account. For the present analysis, the significance of this can be seen when regressing per capita income -- or earnings per wage earner -- on industrial structure. The coefficient on the industry share variable will reflect not only the relative capital or skill intensity of the industry but also the higher order effects suggested here, namely, the association between intra-sectoral differences in earnings per worker and differences between countries in industrial structure. (One further point should be noted parenthetically. The possibility exists that different conclusions may have emerged from the Kuznets analysis if he had been given the opportunity to work with more disaggregated data.)

19/ This point is further discussed in footnote 43/ below.

20/ See Appendix D.

21/ However, this tendency is somewhat stronger when the analysis is limited to the nine pre-1949 provinces. Chernick also notes that the degree of dispersion is much lower when a 'weighted' coefficient of variation is applied. This suggests to him that the resources required to increase the degree of inter-regional participation would not involve a large proportion of the total national income. Finally, he found that the degree of dispersion or trends over time are not significantly altered when income is measured in constant rather than current dollars. This result may be misleading however, since he used as a deflator the price indexes for the eleven regional cities across Canada. Thus, no account is taken of the oft-noted possibility that prices in rural areas are significantly below those in urban centres. Since the low-income provinces are generally more rural, the inclusion of rural prices would likely lower the degree of dispersion among provincial incomes.

22/ Based on this finding, Chernick concluded that "high average provincial income levels not only reflect a relatively uniform degree of participation in the economic activity of the country as a whole but also represents a uniform degree of participation among the sub-regions as well" (page 40).

23/ Denton's methodology is similar to that of Hanna and generally involved the 'standardization' or weighting of average earnings of wage-earners in each region according to a uniform weighting pattern. For example, average earnings in each region were recalculated on the assumption that the region had the same 'mix' of industries as the country as a whole and the results compared with the original or 'unstandardized' averages. However, Hanna interpreted his results somewhat differently from Denton. The latter simply compared the actual and standardized series and found that standardization did not decrease the variability of the series to any appreciable extent. In contrast, Hanna went one step further by

regressing the actual series on the standardized series, in this case, 'rate-constant' earnings; his interpretation was that all of the remaining variability could be attributed either directly or indirectly, to industrial structure.

24/ While not particularly germane to the present exercise, probably Denton's major contribution involved the development of a model to study the impact of changes in the overall Canadian unemployment rate on relative income levels among the five regions. The regional response model contained 41 structural equations estimated from annual data from 1947 to 1964; the main endogenous variables were earned income, unemployment, the price level, the labour force participation rate, and the labour force source population, for each Canadian region, while the critical exogenous variables were the national unemployment rate and the national consumer price index. The model was used to simulate the behaviour of relative income levels over a twenty year period under markedly different national unemployment rates. It was found that the overall picture of regional disparities in Canada remained essentially unchanged under high and low rates of unemployment at the national level. Specifically, Denton discovered that the income gap between the Atlantic Region and the country as a whole did not narrow by operating the model at a low national unemployment rate. One point should be noted however. A possible weakness in the model was the failure to 'endogenize' the growth in population which, in the simulations, was simply assumed to grow at the prevailing regional rates between 1947 and 1964. To the extent that rates of migration out of the Atlantic are highly responsive to national -- and particularly Ontario -- unemployment rates, the population of the Atlantic would be larger under a high national unemployment rate and accordingly the income gap between the Atlantic and Canada as a whole would be larger. At the very least, these considerations cast some doubt on Denton's otherwise very interesting results.

25/ Two other Canadian studies should be mentioned in passing. After examining the lack of convergence in regional per capita incomes in Canada, McInnis (39) concluded that the relative constancy in regional differentials has perhaps been the result of offsetting forces. Factors leading to convergence, such as migration and the marked reductions in regional

differences in age-sex composition, have been balanced off by changing technology -- which has probably favoured high incomes areas -- and differences in the quality of labour. However, an alternative argument could be offered, that is, that migration out of the low income areas has been one of the forces preventing large-scale convergence, since out-migration often involves the youngest and most skilled elements of the work force. To this extent, migration and the quality of the work force may be inter-related.

Also, in the Report of the Royal Commission on Bilingualism and Biculturalism (54), regression analysis was applied to explain earnings differentials among 100,000 male wage and salary earners in Montreal. They found that ethnic origin was a significant explanatory variable, although much less important than under-employment, age, schooling and occupation. This result suggests that ethnic origin -- as measured by the percentage of the population that is Francophone -- might be a useful explanatory variable in wage and income equations estimated for the Atlantic Region.

- 26/ North also suggested that the composition of the export industries is a prime determinant of the cyclical sensitivity of a region.
- 27/ Accordingly, from page 319 of his article, "the resulting pattern of regional growth represents an equilibrium process which will tend to eliminate geographic differences in the returns to resources."
- 28/ Implicit here is the assumption that, within a given industry, not all regions adopt the best available production process.
- 29/ Which Borts defines as the marginal physical product of capital multiplied by the prices of the region's export goods; with the resulting product divided by the price of imported goods.

29 a/ Page 343. The real wage (and thus the marginal product of labour) in a region is a function of the marginal physical product of labour and the prices of its imports.

30/ On the other hand, interstate migration was consistent with traditional theory, i.e. from low to high wage states, although the amount was insufficient to bring about wage convergence.

31/ See pages 21 and 22.

32/ This approach may also be reasonable in a statistical sense since, by definition, the variation among counties in the employment share of residuary industries should be relatively small.

33/ Similar results were obtained for the American States.

34/ In this regard, compare Munro's results on per capita income and the Gini coefficient by province, with Chernick's findings on average income and the degree of dispersion among counties.

35/ The initial rise is associated with a number of factors such as barriers to inter-regional trade and factor and information flows, selective migration and capital flows from poorer regions and the emphasis of central government policy on overall development. The later convergence would reflect less selective labour migration, the development of more efficient capital markets and the reorientation of government policy towards less developed regions.

36/ Note should be made of the fact that there is not total agreement among researchers concerning the association between income and inequality. In a study of regions in the United Kingdom, Devine (16) discovered a small but

positive correlation between the Gini coefficient and average income. However, his time series findings were consistent with the above, for the degree of inequality declined in all regions between 1949-50 and 1964-65, a period of some economic growth in the United Kingdom.

37/ With skill margins likely being lower where the average level of educational attainment is higher.

38/ The larger the Gini coefficient, the greater the degree of inequality within a region.

39/ In the estimated model, mean income was positively related with the employment ratio and education and negatively related with the employment share in agriculture. Income concentration was found to be positively associated with investment income and negatively associated with the employment ratio and education. With respect to the employment ratio, positive coefficients were estimated for mean income, agriculture's share and median age, while the employment ratio was negatively correlated with the unemployment rate. Finally, investment income was positively associated with mean income, income concentration and median age.

The specification for the income concentration equation is very similar to that estimated by Al-Sammari and Miller (2) in their analysis of state differences in income concentration (also measured by the Gini coefficient) in the United States. They found that, in 1959, five-sixths of the state variation in income concentration could be explained by five factors: the share of property income, the relative weight of agriculture, education, per cent of non-white in the total population and the employment ratio. Racial composition -- not used, for perhaps obvious reasons, in the Canadian study of Munro and Holmes -- was found to be the most important right hand variable.

40/

Unfortunately, their study was marred somewhat by a final conclusion that possessed only a tenuous link with their previous analysis. Their final conclusion was that greater attention should be given to increasing education expenditures in rural areas and the Maritimes in order to facilitate migration out of these areas. However, migration was not one of the variables in their model, and it is far from clear what impact migration has on average income and income distribution in the source area.

As Richardson (55) noted, neo-classical general equilibrium theory predicts that, given inter-regional wage differentials, labour will migrate from low to high wage regions until real wages are equalized. He further notes, however, that migration is a dynamic process, and that we cannot be sure that it will be equilibrating in its effects. Among other possibilities, the out-migrants may be largely composed of the most skilled elements of the source area's population. Given the 'non-neutrality in production' hypothesis outlined earlier, their leaving may tend to lower wages and income in the source area.

41/

It should be noted that average value productivity is strictly correlated with the marginal value product for only certain types of production functions. These production functions will generally be assumed in the following analysis.

42/

That is marginal productivity theory in the context of market equilibrium and maximizing behaviour.

43/

An article by Reder (53) is probably the best example of this genre. Reder offered the proposition that, in the long run and under competitive conditions, any industry will pay the same price for the same grade of labour as any other industry, abstracting from non-pecuniary attractions of different industries and locations; thus, real wage differences among industries should only reflect differences in the skill mix, while differences in money wages between location for a given skill should be no

greater than can be explained by differences in living costs. Reder applied various tests to analyze this proposition. Based on all of his results, his conclusion was that the competitive hypothesis is a fairly good predictor of skill margins and inter-industry differentials over a long period of time, but not a very reliable guide in explaining shorter term trends. It can be argued however that the "rationalizations" he offers to justify the competitive hypothesis are somewhat strained, and that his quantitative methodology -- involving the computation of rank correlation coefficients -- was too crude to allow his findings to be more than illustrative.

44/ Pages 40-41.

45/ Page 769. Scully suggested that the industry dummies may not only be measuring the impact of industry mix, but may also be picking up such influences as inter-regional variations in technology or the vintage of the capital stock

46/ Page 770. He indicated previously that the rate of return on human capital is higher in the South, reflecting its relative scarcity in that region.

47/ See footnote 43.

48/ The Reder article (53) is a very good example of this.

49/ Much of this discussion is taken from Richardson (55) Chapter 12. Underlying this discussion is the extent to which factors of production are mobile between regions. Samuelson has shown that, at least theoretically, factor mobility is not necessary for factor price equalization as long as there is free trade in goods. But many of the impediments to the free flow of factors between regions also work against the free flow of goods, such

as transportation costs. Thus, in the real world, neither perfect factor mobility nor free trade in goods prevails, even intra-nationally. These issues will be returned to later in the body of the report.

50/ Although, as suggested earlier, low participation rates are probably more related to economic factors, such as 'unfavourable' industrial structure and below average earnings.

51/ In order to avoid confusion, Region with a capital (R) will be used to refer to the Atlantic, while region with a small (r) will be used in the generalized sense of a sub-national or sub-provincial geographic (or economic) entity.

52/ The second statement in point (5), and point (7) relates to some work of Reder (53). In discussing skill margins and industry differentials, he reported that two low-wage industries in the United States, agriculture and laundering, showed a relative decline in wages over the longer-term, rather than converging to the all-industry average as would be predicted from the competitive hypothesis. He suggested that these trends reflect the fact that both are declining industries and that following from this, the necessary adjustments to reduced employment levels had not yet been completed. He further argued that, since the remaining workers were generally older, the relative declines in wages may also be associated with a relative decrease in skills. Relating this to the present study, the possibility exists that low earnings and/or lack of convergence in particularly depressed counties may be the consequence of a similar set of 'disequilibrium' forces. These effects might be picked up by the unemployment rate or the average age of the work force, among other variables.

53/ Vanderkamp (63), in a cross-section study of migration in the Maritimes, found that on balance out-migration led to some improvement in unemployment but that its effects were less than "overall" figures might suggest, since on average for every five unemployed people leaving the Maritimes, two previously employed persons became

unemployed. These "second-order" effects of migration further cloud the association between migration and in particular average earnings.

54/ This is particularly true for Chapter 5. For Chapters 6 through 9, it will be necessary to further elaborate upon the literary and theoretical antecedents for the equation specifications.

55/ Even if there were sufficient degrees of freedom, it is not possible to include all industries in the 'employment shares' matrix in the regression estimation, since the employment shares add to one. This would make the X matrix 'singular' -- since the employment shares would sum to the column of ones for the constant term -- and thus prevent the inversion of the $X'X$ matrix in the computation of the coefficients. This is a special case of the more general problem of multicollinearity.

56/ See Johnston, page 322.

57/ Which in a cross-section model is synonymous with the number of exogenous variables.

58/ The problem of identification is discussed in Chapter 12 of Johnston (32); while a comprehensive treatment of the identification problem is provided by F.M. Fisher, The Identification Problem, McGraw-Hill, New York, 1966.

59/ See Johnston, Chapters 12 and 13.

60/ Johnston (32), pages 344-346. Indirect Least Squares involves first deriving the "reduced form" of the model from the structural (or original) form and then using OLS to estimate the reduced form. The reduced form expresses each endogenous variable as a function of all of the exogenous variables in the system. Since each equation is just identified, the structural form coefficients can be derived uniquely from the reduced form coefficients.

61/ Johnston (32), pages 380-384. Basically, Two Stage Least Squares is Ordinary Least Squares in two steps. In the first step, each endogenous (Y) variable is regressed on all of the exogenous variables in the model and the fitted Y values recovered from each regression. (This involves therefore the estimation of the reduced form of the model.) Then, in the second step, each structural form equation is estimated using the fitted Y values in place of the actual Y values when an endogenous variable appears on the right hand side of a regression. The key here is that, through step one, the endogenous variables are "purged" of their correlation with the error term.

62/ Forty-six observations are used in this estimation and, accordingly, there are 35 degrees of freedom or more in each equation. Unfortunately, insufficient work has been done to date in econometrics to say with certainty that small sample bias will not arise in the present context. Beyond this possible difficulty, insufficient degrees of freedom did prove a hindrance in estimation, for in most equations an endogenous variable was a function of a relatively large number of explanatory variables. One "rule of thumb" that is sometimes applied in econometrics is that there should be at least ten observations for each estimated parameter in a regression equation. In general, it has not been possible to follow this rule in the present exercise.

63/ First, the structural form is expressed in the following terms:

$$AY = BX + u$$

Then, the reduced form can be derived through inverting the A matrix, that is:

$$\begin{aligned} Y &= A^{-1} BX + A^{-1} u \\ &= \pi X + v \end{aligned}$$

In a linear model there exists a unique solution for a given set of predetermined values. Since an algorithm for solving non-linear systems was not available to the

author at the time of the preliminary report, the initial specification of the model was solved by (1) deriving "linear approximations" for the non-linear equations in the model, and (2) utilizing the linear algebra outlined above. The solution results were outlined in a supplementary paper to the preliminary report. Appendix H offers a comparison between (1) the initial and final specifications of the model, and (2) the initial and final solution results.

64/ See Appendix A.

65/ For a discussion of the problems encountered in solving non-linear systems of equations, see Peter Henrici, Elements of Numerical Analysis, John Wiley and Sons, New York, 1964, Chapter 5.

66/ However, where theoretical and literary antecedents do exist, they are noted in the first sub-section of Chapters 5 to 9.

67/ Beyond the standard errors of the coefficients, the t-statistics, and the adjusted and un-adjusted multiple correlation coefficients (R^2 and R^2), three further statistics are noted in outlining the regression results. The first is the F-Statistic (F-STAT) for the overall significance of the regression. This statistic is significant as long as at least one of the coefficients on the explanatory variables is significantly different from zero. At the 95% confidence level, the critical value for the F-Statistic comes to 2.12 for 10 and 35 degrees of freedom, and 2.45 for 5 and 40 degrees of freedom. The second is the Standard Error of the Estimate (SEE). Third, the coefficient of variation (CV) is noted, this measure is simply the Standard Error of the Estimate divided by the mean of the endogenous variable.

68/ Multicollinearity relates to the intercorrelation among the right hand variables. The presence of multicollinearity provides an upward bias to the standard errors of the coefficients, and thus tends to lower the t-statistics.

69/ The presence of multicollinearity can be detected in part through the application of the Farrow-Glauber equations which basically involves regressing each explanatory variable in turn on every other explanatory variable in the same equation and then comparing the F-Statistics from each regression. One set of Farrow-Glauber equations was run in the present exercise; these involve the final OLS specification for the wage determination equation. These results are outlined in Appendix E.

70/ Some researchers -- in particular Holmes and Munro (29) -- have suggested that the application of 2SLS offers a partial solution to the problem of multicollinearity. Their conclusion is strongly questioned in Appendix G.

71/ One of the critical assumptions of OLS is that the error terms are pair-wise independent. See Johnston, Chapter 8.

72/ In the present study, the counties are simply ordered alphabetically by province -- that is, as they are ordered in the Census publications -- and thus have not been ordered in terms of population size, per capita income, location or some other manner appropriate to the detection of autocorrelation in a cross-section model. Accordingly, the Durbin-Watson statistic is of limited utility and is not included with the regression results presented here. However, a DW statistic diverging markedly from two could be indicative of a large number of errors of similar sign and magnitude within a given province. In general though, analysis of the residuals themselves will be necessary for the detection of possible autocorrelation, or more generally for the detection of similar error terms among counties with similar characteristics. This analysis is often conducted in the present exercise in order to improve the specification of an equation and in particular to test whether a key explanatory variable is missing from the equation.

73/ See Johnston (32), pages 214-221. This problem especially arises when the size of the explanatory variable varies greatly among observations. The expectation is that some degree of heteroscedasticity is present in this system, since for example, employment shares in a given industry can vary greatly from county to county, sometimes as much as ten-fold. Our results there are probably clouded by the presence of heteroscedasticity, but little can be done to solve this problem. Its ultimate solution requires the application of Generalized Least Squares for which no computer programs are presently available in Ottawa.

74/ The methodology adopted by the author for the initial estimation of the model provided "correct" estimates for the coefficients but did not provide correct estimates of the diagnostic statistics (SEE, R^2 , etc.). This point is further discussed in Appendix H.

75/ In the smaller counties, the share of employment in construction was the result in large part of shorter term factors such as whether a large project was being constructed in the year under review. For the latter, it was thought that pulp and paper would largely be a function of the size of the forestry sector but, as indicated in the principal components appendix, no such association was found to exist, at least on an individual county basis.

76/ Low and high value added manufacturing are defined in Appendix B and further discussed in Appendix E.

77/ As discussed in Appendix E, a proxy measure had to be used for the level of capitalization.

78/ See page 9 of Appendix C.

79/ The analysis of residuals from equation one suggested that the equation tended to over-estimate employment income per capita in areas where the "modernized" sector -- and thus wage earners as a proportion of total employment --

tends to be small (and alternatively tended to underestimate income in the more highly developed counties). The implication was that W is perhaps an over-estimate of earned income per employee in the less developed counties. This tendency could perhaps be corrected by including wage earners as a proportion of total employment as a further adjustment factor in equation one. However, when this specification was attempted, wage earners as a proportion of total employment entered the equation with a t -statistic below one and, moreover moved the coefficient on wages further away from -- rather than closer to -- unity.

^{80/} Or at least with t -statistics greater than one. <

^{81/} See Table G-4 in Appendix G. There is some possibility that the degree of multicollinearity between the RS and HOBS variables increased in moving from OLS to 2SLS estimation. However, the sharp (absolute) decline in both t -statistics is also explained in part by the marked absolute decreases in the coefficients. It was decided therefore that the two coefficients were not sufficiently "robust" to allow their inclusion in the final specification, while experiments attempted with each variable separately proved unproductive.

^{82/} As one possible specification error, 2SLS specifications were attempted including two occupation variables, managerial and professional and clerical and sales, as right hand variables. While both variables were positive and significant using Ordinary Least Squares (see Appendix E), they failed to add to the explanatory power of the 2SLS wage equation.

^{83/} See for example, Evans (20), page 260, and McCracken (38), page 279.

^{84/} This hypothesis is often investigated in developing time series models. See for example, Helliwell (27), page 106 and McCracken (38), page 279. There is though a counter hypothesis used in modelling the participation rate for women. This is the "additional worker"

hypothesis which assumes a positive coefficient for the unemployment rate. This was used in the Candide model (McCracken, page 66). Perhaps our ultimate lack of success with the unemployment rate variable stems from these two countervailing effects offsetting each other.

- 85/ In some time-series models, for example, the Candide model (McCracken, page 65), the labour force is divided into its primary and secondary components. The former includes males, generally twenty-five to fifty-four, while the latter is composed of younger and older males and females. Thus, females generally account for the majority of the secondary component.
- 86/ Income variables are used in the participation rate equations of both the Candide and RDX 2 models (see McCracken, page 65 and Helliwell, page 105). In the present context, the implicit assumption is that, at the relatively low income levels in the Atlantic, the supply curve of labour is upward sloping over the relevant portion, rather than "backward bending" which would be the result if the income effect swamped the substitution effect.
- 87/ Admittedly, this is a very cumbersome variable to work with. The more familiar measure for hours is average hours worked per week. However, the hours data from the Census is too aggregated to allow us to compute this average.
- 88/ The specification attempted here therefore has features in common with that used in the Wharton School model, which includes an index of capacity utilization and the yearly wage rate as right hand variables. See Evans (20), page 260.
- 89/ A similar interpretation could be placed on the negative coefficient for low value added manufacturing.

90/ Despite a t-statistic well below one, residential services has been kept in the equation. Two reasons can be offered for this: first, the sign and size of the coefficient on RS seem defensible; second, the results indicated a high degree of collinearity between RS and both GE and GLFP and that this collinearity tended to increase in moving from OLS to 2SLS estimation.

91/ In the OLS estimation, the correlation coefficient between GE and GLFP came to .93. A correlation of this magnitude is generally indicative of "harmful" multicollinearity but, in this case, it was not a problem.

92/ The possibility exists then that, in the Atlantic Region, the income and substitution effects are of approximately equal weight. This might be an area for further research.

93/ In early OLS results, forestry and agriculture possessed their anticipated signs but their coefficients were very small. Therefore, as Table G-10 indicates, fishing was the only primary industry appearing in the final OLS specification. However, in moving to 2SLS estimation, the t-statistic on fishing fell below one and therefore, the variable was dropped from the equation. Further, LVAM, which was also in the final OLS specification, stopped adding to the adjusted R^2 when PFR was included in the specification.

94/ The coefficient on PFR is quite small but for extreme cases it can be quite significant. For example, the coefficient suggests that, other things being equal, the percentage of females in the employed labour force is 5.4 percentage points higher in a county that is 100% Francophone relative to one that is totally Anglophone.

95/ For a fairly complete discussion of the factors associated with migration, see Richardson (56), pages 89-103. Richardson notes that migration models can be divided into two groups: The first are essentially economic models where the role of income or job opportunities is

crucial. The second group is more varied, including gravity models, information flow models and non-economic theories, but all stress the influence, either direct or indirect, of space and distance. The specifications attempted in the present exercise generally reflect the first category although one non-economic variable (PFR) was later included.

96/ Shift/share analysis is discussed in Richardson (55) pages 342-347, while it is used extensively in the Perloff, etc. text (49).

97/ There are two difficulties with using the 1951-61 averages: first, it would involve the introduction of totally new variables into the model; second, since the 1951 Census information is less detailed, it would require the use of more aggregated industry groupings. On balance then, the 1961 industry variables are perhaps the preferred right hand arguments given the above constraints.

98/ This term was coined by Perloff et al (49) in their 1960 text and relates to a region's "access" to basic inputs and markets. They argued that "inter-regional differences in access to inputs and markets can be expected to bring about different patterns of regional growth behaviour" and conversely that "the various growth patterns displayed can be explained by identifying the relative advantages and disadvantages of the region with regard to input and output access for the major types of economic activities". The latter statement probably underlies the useful role played by urbanization in the growth in employment equation.

99/ The concept of agglomeration economies is discussed at length in both Richardson texts (55 and 56).

100/ Perhaps for our purposes, a more precise definition of urbanization would be preferred, such as the proportion of the population in centres of ten thousand and over. However, we had little success in using an urbanization variable based on this definition as a right hand

argument in the wage equation. In other words, this definition did no more for us than the more standard one. Moreover, there is some problems attached to this alternative definition, for example, the large number of zero elements among the observations which make the variable "less continuous" than the original urbanization variable. Therefore, the decision was made to go with the more standard definition in developing the model.

101/ One possible drawback to the 2SLS results, however, was the sharp increase in the size of the coefficient on other high value added manufacturing (Table G-14). This may have lead to some difficulties in solving the model.

102/ The qualifying phrase is necessary because the simple correlation coefficient between the level of education and the net migration rate is positive and almost significant at the 95% confidence levels, using a one-tail test. However, as indicated in the text, the relationship turns negative (although not significant) once industrial structure -- and therefore the demand for skills -- is taken into account.

103/ Perloff (49), page 76 and pages 89-93.

104/ To be defined below when the EH estimation results are discussed.

105/ Equation 13 indicates a fairly strong positive association between GE and PA. Therefore, an alternative hypothesis for the RS equation is that RS and PA are negatively related, reflecting the fact that many military bases tend to be self-contained. However, perhaps because of the high collinearity between PA and GE, attempts at including both in the RS equation did not prove fruitful. It should be noted though that the absolute magnitude of the GE coefficient may be inflated by its collinearity with PA.

106/ For the exogenous industry variables, the coefficients sum to minus one.

107/ That is, by not determining IR through a behavioural equation.

108/ And thus adopting the Tobin - Brainard estimation methodology.

109/ These were outlined in the Methodology section, page 76 and include the problems of non-convergence and multiple solutions.

110/ An instrument variable is one that at least potentially can be altered by government policy intervention.

111/ For some solution sets negative values were obtained for the industry residual. On the other hand, in deriving the solution results, it was thought sufficient if the industry residual simply approached its control solution value. Demanding total accuracy would have used up too much computer time. Moreover, the results indicated that small changes in the industry shares did not have a very large impact on the main endogenous variables, i.e. earnings, per capita income and the degree of inequality in income. Still, it should be remembered that little or no economic meaning can be attached to the solution value for the industry residual, while the solution results for for other five endogenous industry variables are generally approximations only.

112/ At least for the sector as a whole, there could be fairly well developed linkages for some of the components in the sector, such as federal civilian employment.

113/ One specification that was tested involved including both EIC and W in the RS equation, with the hope that they would operate in opposite directions. However, the estimation results proved unsatisfactory, perhaps reflecting again the probability that the hourly wage rate rather than annual earnings is the more appropriate right hand variable.

114/ It would be a mistake however to place too much emphasis on this. As Table G-13 points out, LVAM entered the original OLS specification with a negative coefficient. However, in moving to 2 SLS estimation, its t-value dropped below one, although the coefficient remained negative and in fact increased somewhat. As a consequence, LVAM was dropped from the specification. An argument could be made that equation 13 (b) is to be preferred over 13 (d), despite the low t-statistic on LVAM. If 13 (b) had been used in the model, the growth in employment in Test (M) would likely have been at least two percentage points lower.

115/ The advance in the participation rate is a consequence in turn of the increase in PFELF.

116/ Whereas viewing per capita income and migration alone, our a priori expectation would likely be that increased per capita income would lead to a drop in out-migration.

117/ Again, as suggested in footnote 114/, equation 13 (b) might be the preferable specification for the GE equation (see Table G-13).

118/ At the time that the final draft of this thesis was being completed, some industrial distribution data from the Census was becoming available. However, it was too aggregated to fully meet our needs, as it includes only twelve industry groups. It was decided therefore that this additional information was not sufficient to warrant re -- computing scenarios (F) through (I).

- 119/ The implicit GNE prime deflator was used to deflate the increases in per capita income and earnings. Being a national index, this can only provide an approximation for price trends in the Atlantic Region. Moreover arguments could be made for applying a different national index, for example, the consumer price index. With respect to the first problem, no comprehensive price index exists on the regional level, while for the second, the GNE deflator seems the more appropriate measure.
- 120/ That is, with greater weight given to the more populous counties, which also tend to be the more highly developed.
- 121/ "Add" and "mult" factors -- or adjustments -- are called in the trade "fine tuning".
- 122/ The adjustments made in Simulations (G) through (I) are exactly the same as those for Simulations (A) through (C). Further, Simulation (J) is the same as Simulation (D) and Simulation (L) is the same as Simulation (E).
- 123/ See page 96 of Chapter 5.

BIBLIOGRAPHY

- (1) ALLEN, R.G.D., Macro-economic Theory, MacMillan and Company, London, 1967.
- (2) AL-SAMMARIE, Ahmad and Miller, HERMAN, P., "State Differentials in Income Concentration," American Economic Review, March, 1967, pp. 59-71.
- (3) BORTS, G.H., and STEIN, J.L., Economic Growth in a Free Market, Columbia University Press, New York, 1964.
- (4) BORTS, G.H., and STEIN, J.L., "Regional Growth and Maturity in the United States: A Study of Regional Structural Change," Regional Analysis: Selected Readings, Lionel Needleman, Editor, Penguin Books, Harmondsworth, England, 1968, pp. 159-197.
- (5) BORTS, G.H., "The Equalization of Returns and Regional Economic Growth," American Economic Review, June, 1960, pp. 319-347.
- (6) BRAINARD, William C., and TOBIN, James, "Pitfalls in Financial Model Building," American Economic Review, May, 1968, pp. 99-122.
- (7) CHENERY, H.B., "Patterns of Industrial Growth," American Economic Review, September, 1960, pp. 624-654.
- (8) CHERNICK, S.E., Inter-regional Disparities in Income, Staff Study, No. 14, Economic Council of Canada, Queen's Printer, Ottawa, 1966.
- (9) CHIANG, Alpha C., Fundamental Methods of Mathematical Economics, McGraw-Hill Incorporated, New York, 1967.

- (10) CHISWICK, Barry R., "Earnings Inequality and Economic Development," Quarterly Journal of Economics, June, 1971, pp. 21-39.
- (11) CORINA, John, Labour Market Economics. A Short Survey of Recent Theory, Heinemann Educational Books, London, 1972.
- (12) COURCHENE, Thomas J., "Inter-provincial Migration and Economic Adjustment," Canadian Journal of Economics, November, 1970, pp. 550-576.
- (13) CULLEN, Donald E., "The Inter-industry Wage Structure," American Economic Review, June, 1956, pp. 353-369.
- (14) CZAMANSKI, Stanislaw, Regional Income and Product Accounts of Northeast Nova Scotia, Regional Studies, No. 1, Institute of Public Affairs, Dalhousie University, Halifax, Nova Scotia, Canada, 1968.
- (15) CZAMANSKI, Stanislaw, An Econometric Model of Nova Scotia, Regional Studies Series, No. 2, Institute of Public Affairs, Dalhousie University, Halifax, Nova Scotia, Canada, 1968.
- (16) CZAMANSKI, Stanislaw, Regional Science Techniques in Practise: The Case of Nova Scotia, D.C. Heath and Company, Lexington, Massachusetts, 1972.
- (17) DENTON, Frank T., An Analysis of Inter-regional Differences in Manpower Utilization and Earnings, Staff Study, No. 15, Economic Council of Canada, Queen's Printer, Ottawa, 1966.
- (18) DEVINE, P.J., "Inter-regional Variations in the Degree of Inequality of Income Distribution," United Kingdom 1949-1965, The Manchester School of Economic and Social Studies, June 1969, pp. 141-160.

- (19) DUNLOP, John T., "Productivity and the Wage Structure, pp. 341-362. Income, Employment and Public Policy, W.W. Norton, New York, 1948.
- (20) EVANS, Michael K., Macro-economic Activity: Theory, Forecasting and Control, Harper and Row, New York, 1969.
- (21) FUCHS, Victor R., The Service Economy, National Bureau of Economic Research, New York, 1968.
- (22) FULMER, John L., "Measurement of Agricultural Income of Counties," Regional Income, Studies in Income and Wealth, Volume 21, National Bureau of Economic Research, Princeton University Press, Princeton, 1957, pp. 343-376.
- (23) GARBARINO, Joseph W., "A Theory of Inter-Industry Wage Structure Variation," Political Journal of Economy, May, 1950, pp. 282-305.
- (24) HANNA, Frank A., "Analysis of Interstate Income Differentials: Theory and Practice," Regional Income, Studies in Income and Wealth, Volume 21, National Bureau of Economic Research, Princeton University Press, Princeton, 1957, pp. 113-194.
- (25) HANNA, Frank A., State Income Differentials, 1919-1954, Duke University Press, Durham, North Carolina, 1959.
- (26) HAY, Keith A.J., "Trends in the Location of Industry in Ontario," Canadian Journal of Economics and Political Science, August, 1965, pp. 368-381.
- (27) HELLIWELL, John F. and Associates, The Structure of RDX2, Part I, Bank of Canada, Ottawa, 1971.
- (28) HERSHKOWITZ, M., and SUSSMAN ZVI, "Growth, Induced Changes in Final Demand, Educational Requirements and Wage Differentials," The Review of Economics and Statistics, May, 1971, pp. 169-175.

- (29) HOLMES, R.A., and MUNRO, J.M., "Regional Non-farm Income Differences in Canada: An Econometric Study," Journal of Regional Science, April, 1970, pp. 65-74.
- (30) ISARD, Walter, Methods of Regional Analysis, Technology Press of the Massachusetts Institute of Technology, and John Wiley and Sons, New York, 1960, Chapter 4.
- (31) JOHNSON, George E., "Wage, Theory and Inter-regional Variation," Industrial Relations, May, 1967, pp. 321-338.
- (32) JOHNSTON, J., Econometric Methods, McGraw-Hill Book Company, New York, 1972.
- (33) KUZNETS, Simon, "Quantitative Aspects of the Economic Growth of Nations: Part II, Industrial Distribution of National Product and Labour Force," Economic Development and Cultural Change, July, 1957, (Supplement), pp. 3-111.
- (34) KUZNETS, Simon, "Quantitative Aspects of the Economic Growth of Nations: Part III, Industrial Distribution of Income and Labour Force by States, United States, 1919-1921 to 1955," Economic Development and Cultural Change, July 1958, pp. 1-128.
- (35) KUZNETS, Simon, "Quantitative Aspects of the Economic Growth of Nations: Part VIII, Distribution of Income by Size," Economic Development and Cultural Change, January, 1963, pp. 1-80.
- (36) LITHWICK, N.H., Economic Growth in Canada, University of Toronto Press, Toronto, 1969.
- (37) LOGAN, M.F., "The Spatial Dimensions of Economic Development: The Case of the Upper Midwest," Regional Studies, May, 1970, pp. 117-125.

- (38) McCRACKEN, M.C., Candida Project Paper No. 1: An Overview of Candida Model 1.0, Information Canada, Ottawa, 1973.
- (39) McINNIS, Marvin, "The Trend of Regional Income Differentials in Canada," Canadian Journal of Economics, May, 1968, pp. 440-470.
- (40) MANSFIELD, Edwin, "City Size and Income," Regional Income, Studies in Income and Wealth, Volume 21, National Bureau of Economic Research, Princeton University Press, Princeton, 1957, pp. 271-318.
- (41) MASTERS, Stanley H., "An Inter-Industry Analysis of Wages and Plant Size," Review of Economics and Statistics, August, 1969, pp. 341-345.
- (42) M'RONEY, J.R. and WALDER, J.M., "A Regional Test of the Heckscher-Ohlin Hypothesis," Regional Analysis: Selected Readings, Lionel Needleman Editor, Penguin Books, Harmondsworth, England, 1968, pp. 252-273.
- (43) MUNDELL, R.A., "International Trade and Factor Mobility," American Economic Review, June, 1957, pp. 320-335.
- (44) MUNRO, John M., The Measurement of Inter-regional Income Disparities in Canada, Paper presented to the Annual Meeting of the Canadian Political Science Association, Ottawa, June 8, 1967.
- (45) NORTH, Douglas C., "Location Theory and Regional Economic Growth," Journal of Political Economy, June, 1955, pp. 243-258.
- (46) PAFOLA, T.S. and BHARADWAJ, V.P., "Dynamics of Industrial Wage Structure: An Inter-country Analysis," Economic Journal, March, 1970, pp. 72-90.

- (47) PERLOFF, Harvey S., "Assessing Regional Progress", Regional Income, Studies in Income and Wealth, Volume 21, National Bureau of Economic Research, Princeton University Press, 1957, pp. 35-68.
- (48) PERLOFF, Harvey S., "Inter-relations of State Income and Industrial Structure", Regional Science Association, Papers, II, 1956, pp. 70-85.
- (49) PERLOFF, Harvey S., and Associates, Regions Resources and Economic Growth, Resources for the Future Inc., John Hopkins Press, Baltimore, 1960.
- (50) PODOLUK, Jenny R., Incomes of Canadians, 1961 Census Monograph, Dominion Bureau of Statistics, Ottawa, Canada, 1968.
- (51) POLZIN, Paul E., "State and Regional Wage Differences", Southern Economic Journal, January, 1972, pp. 371-372.
- (52) REDER, M.W., "The Theory of Occupational Wage Differentials", American Economic Review, December, 1955, pp. 833-852.
- (53) REDER, M.W., "Wage Differentials: Theory and Measurement", from Aspects of Labour Economics, National Bureau of Economic Research, Princeton, 1962.
- (54) Report of the Royal Commission on Bilingualism and Biculturalism, Book III, The Work World, Queen's Printer, Ottawa, 1969, Chapter 5.
- (55) RICHARDSON, Harry W., Regional Economics, Weidenfield and Nicholson, London, 1969.
- (56) RICHARDSON, Harry W., Regional Growth Theory, the MacMillan Press Ltd., London, 1973.

- (57) SAMUELSON, P.A., "International Trade and Equalization of Factor Prices," Economic Journal, June, 1948, pp. 163-184.
- (58) SCULLY, Gerald W., "Interstate Wage Differential: A Cross Section Analysis," American Economic Review, December, 1969, pp. 757-773.
- (59) SLATER, David W., "Decentralization of Urban Peoples and Manufacturing Activity in Canada," Canadian Journal of Economics and Political Science, February, 1961, pp. 72-84.
- (60) STABLER, J.C., "Exports and Evolution: The Process of Regional Change," Land Economics, February, 1968, pp. 11-23.
- (61) TIEBOUT, Charles M., "Exports and Regional Economic Growth," Journal of Political Economy, April, 1960.
- (62) THOMPSON, Lorin A., "Appraisal of Alternate Methods of Estimating Local Area Incomes," Regional Income, Studies in Income and Wealth, Volume 21, National Bureau of Economic Research, Princeton University Press, Princeton, 1957, pp. 319-342.
- (63) VANDERKAM, John, "The Effect of Out-Migration on Regional Employment," Canadian Journal of Economics, November, 1970, pp. 541-549.
- (64) WALSH, F., "Education in Production," American Economic Review, September, 1969, pp. 35-39.
- (65) WEISS, Leonard W., "Concentration and Labour Earnings," American Economic Review, March, 1966, pp. 96-117.

- (66) WILLIAMSON, J.G., "Regional Inequality and the Process of National Development: A Description of the Patterns," Regional Analysis: Selected Readings, Lionel Needleman, Editor, Penguin Books, Harmondsworth, England, 1968, pp. 99-158.
- (67) WOYTINSKY, W.S., Employment and Wages in the United States, Twentieth Century Fund, New York, 1953, Chapter 39-44.

APPENDICES

- APPENDIX A: THEORETICAL FOUNDATIONS
FOR THE WAGE AND INCOME EQUATIONS
- APPENDIX B: MNEMONICS
- APPENDIX C: SIMPLE CORRELATION ANALYSIS
- APPENDIX D: STANDARDIZATION ANALYSIS:
METHODOLOGY AND RESULTS
- APPENDIX E: SINGLE EQUATION ESTIMATION:
EARNINGS AND INCOME PER CAPITA
- APPENDIX F: PRINCIPAL COMPONENTS
- APPENDIX G: COMPARISON BETWEEN THE ORDINARY LEAST
SQUARES AND TWO STAGE LEAST SQUARES RESULTS
- APPENDIX H: COMPARISON OF PRELIMINARY REPORT
RESULTS WITH FINAL RESULTS
- APPENDIX I: DATA SOURCES

APPENDIX A:
THEORETICAL FOUNDATIONS
FOR THE WAGE AND INCOME EQUATIONS

APPENDIX A:
THEORETICAL FOUNDATIONS
FOR THE WAGE AND INCOME EQUATIONS

As the review of the literature indicates, the present analysis has a very clear empirical orientation. Following from this, it was not anticipated that a totally new theory of wage and income determination would originate from this work. Rather, the objective of the study is to develop a fruitful approach for empirically estimating how wages and incomes are determined in a given region, and how various economic variables interact to generate earnings and income differentials between areas within that region. Still, underlying this work is a solid foundation of traditional economic theory. This Appendix reviews the relevant economic theory and outlines the inferences and implications that can be drawn from it. Moreover, attention is focussed on how the underlying theory could be tested directly if all of the requisite input and output data were available. This provides a useful backdrop to the discussion of (1) the actual methodology that is applied in the present exercise (Chapter 3) and (2) the estimation results for the earnings and income block (Chapter 5).

/...

As discussed at the end of Chapter 2, the perspective of the regional planner has been adopted for this exercise, with analysis first focussing on the industrial dimension and then going on to other variables. This approach moreover is consistent with traditional "marginal productivity" theory which emphasizes differences in the amounts and types of the demands for and supplies of labour. Accordingly, a review of the traditional theory should begin with a study of production functions and the differing marginal products of labour -- and thus earnings -- implied by different production functions.

Expanding on this, the theory would suggest that, given the availability of all of the requisite input and output data by county and by industry, the study of income and earnings differentials should begin with the analysis of the production functions of those industries that make up a region's economic base.

Before going further however, it is necessary, to outline how average earnings and average per capita income are computed mathematically (rather than determined economically which of course is the objective of the present exercise). Average earnings (w_j) in a given county is simply the weighted sum of the wages (w_{ij}) of the individual industries (i) in the county, where the weights are the employment shares of industry (i) in county (j), that is:

/...

$$(1) w_j = \sum_{ij} ES_{ij}$$

$$\text{where } ES_{ij} = \frac{E_{ij}}{E_{Tj}}$$

and w_j = average earnings in county (j)

w_{ij} = earnings of industry (i) in county (j)

ES_{ij} = the employment share of industry (i) in county (j)

E_{ij} = employment (i.e. the number of the employed labour force) of industry (i) in county (j)

E_{Tj} = total employment in county (j)

Accordingly, the ES_{ij} (i.e. the weights) sum to one.

Then, based on data for 26 industries, average earnings for all 46 counties can be computed through the following matrix multiplication:

$$(2) AW^a = W \times ES \text{ or}$$

$$\begin{bmatrix} w_1 \\ w_2^a \\ \vdots \\ w_3^a \\ \vdots \\ w_{46}^a \end{bmatrix}_{[46 \times 46]} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} & \cdots & w_{26,1} \\ w_{12} & w_{22} & w_{32} & w_{42} & \cdots & w_{26,2} \\ w_{13} & w_{23} & w_{33} & w_{43} & \cdots & w_{26,3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{1,46} & w_{2,46} & w_{3,46} & w_{4,46} & \cdots & w_{26,46} \end{bmatrix}_{[46 \times 26]} \times \begin{bmatrix} ES_{11} & ES_{12} & ES_{13} & \cdots & ES_{1,46} \\ ES_{21} & ES_{22} & ES_{23} & \cdots & ES_{2,46} \\ ES_{31} & ES_{32} & ES_{33} & \cdots & ES_{3,46} \\ ES_{41} & ES_{42} & ES_{43} & \cdots & ES_{4,46} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ES_{26,1} & ES_{26,2} & ES_{26,3} & \cdots & ES_{26,46} \end{bmatrix}_{[26 \times 46]}$$

The elements along the diagonal of the AW^a matrix are actual average earnings for each county, while the remaining elements in the matrix could be used in standardization analysis. (Accordingly, the

/...

elements down a column of AW^a show the effects of holding the employment shares -- or weights -- constant and allowing only wages to vary; these therefore are what Polzin (51) calls 'industry constant' earnings; similarly, the elements along a row indicate 'rate constant' earnings.)

As the second step, earned or employment income per person can be viewed as a multiplicative function of, earned income (i.e. earnings) per employee, employment as a proportion of the labour force (i.e. one minus the unemployment rate), the labour force as a proportion of the labour force source population (the participation rate), and the source population as a proportion of the total population (one minus the dependency ratio). That is:

$$(3) \quad \frac{Y^e}{N} = \frac{Y^e}{E} \times \frac{E}{L} \times \frac{L}{N'} \times \frac{N'}{N}$$

where Y^e is earned income

N is total population

E is total employment

L is the labour force

N' is the labour force source population
(population 15 and over excluding inmates
of institutions and Indians on reserves).

$$\text{and therefore } \frac{Y^e}{N} = \frac{Y^e}{E} \times \frac{E}{N}$$

where E/N is what Denton calls the size of the employment base.

/...

Accordingly, bringing equations 2 and 3 together, it can be seen that earned income per person is a function of the industry mix and the three labour force utilization variables. A specification of this form is discussed in the single equation estimations outlined in Appendix E.

Turning next to production function analysis, given complete data by industry and county concerning output (or more precisely value added) plus the labour, physical capital and human capital (i.e. skills) employed in the production process, it would be possible to compute production functions of the following general form:

$$(5) \quad Q_i \equiv f(L_i, K_i, HC_i)^{2/}$$

where Q_i is total output or value added in industry (i)

L_i is 'homogeneous' labour employed in industry (i)

K_i is the physical capital in industry (i)

HC_i is the 'human capital' employed in industry (i), however defined and measured.

Various production functions could be fitted to the data, such as the usual Cobb-Douglas production function with constant returns to scale, or a modified Cobb-Douglas allowing increasing and decreasing returns to scale, or a constant

/...

elasticity of substitution production function which allows the elasticity of substitution to differ from unity. The production function that provided the best fit with the data -- the highest R^2 -- would then be utilized in the further analysis of that industry. For exposition purposes, the Cobb-Douglas production function of the following form will be analyzed here:

$$(6) \quad Q_i = AK_i^\alpha HC_i^\beta L_i^\gamma$$

where $\alpha + \beta + \gamma = 1$

This production function has the following 'nice' properties:

- (1) constant returns to scale, i.e. the restriction $\alpha + \beta + \gamma = 1$ is imposed;
- (2) following out of the first property, Euler's theorem holds, i.e. if each input factor is paid the amount of its marginal product, the total product will be exhausted exactly by the distributive shares for all input factors;
- (3) unitary elasticity of substitution.^{3/}

The next step would be to regress Q_i on the amounts of the inputs in order to estimate the above coefficients. For the above Cobb-Douglas function, a log-linear specification would be required of the form:

$$(7) \ln Q = \ln A + \alpha \ln K_1 + \beta \ln HC_1 + \delta \ln L_1 + \varepsilon_1$$

A 'restricted' least squares regression would be estimated in order to ensure that the condition $\alpha + \beta + \delta = 1$ was met.

Once the production function was estimated, the marginal products of each factor could then be derived through partial differentiation. For the Cobb-Douglas, there would be:

$$(8) \quad MP_K = \frac{\partial Q}{\partial K} = \alpha A K^{\alpha-1} HC^{\beta} L^{\delta} = A \alpha K^{\alpha-1} HC^{\beta} L^{\delta-1}$$

$$= A \alpha K^{\alpha-1} HC^{\beta} L^{\delta-1} = A \alpha \left(\frac{K}{L}\right)^{\alpha-1} \left(\frac{HC}{L}\right)^{\beta}$$

$$(9) \quad MP_{HC} = \frac{\partial Q}{\partial HC} = A K^{\alpha} HC^{\beta-1} L^{\delta} = A K^{\alpha} HC^{\beta-1} L^{\delta-1}$$

$$= A K^{\alpha} HC^{\beta-1} L^{\delta-1} = A \left(\frac{K}{L}\right)^{\alpha} \left(\frac{HC}{L}\right)^{\beta-1}$$

$$(10) \quad MP_L = \frac{\partial Q}{\partial L} = A K^{\alpha} HC^{\beta} L^{\delta-1} = A K^{\alpha} HC^{\beta} L^{\delta-1}$$

$$= A (1-\alpha-\beta) \left(\frac{K}{L}\right)^{\alpha} \left(\frac{HC}{L}\right)^{\beta}$$

However, since this is now a three factor production function, the average earnings of (heterogeneous) labour would no longer be equal to the marginal product of homogeneous labour.

Rather, since human capital is embodied in labour, average earnings in a given industry would be a function of the marginal product of homogeneous labour, the marginal product of human capital, plus the (average) number of units of human capital per unit of homogeneous labour in the industry. For

the Cobb-Douglas function, the wage equation would take the following form:

$$\begin{aligned}
 (11) \quad W^C &= MP_L + \frac{HC}{L} \times MP_{HC} \\
 &= A(1-\alpha-\beta) \left(\frac{K}{L}\right)^\alpha \left(\frac{HC}{L}\right)^\beta + \left(\frac{HC}{L}\right) A\beta \left(\frac{K}{L}\right)^\alpha \left(\frac{HC}{L}\right)^{\beta-1} \\
 &= A(1-\alpha-\beta) \left(\frac{K}{L}\right)^\alpha \left(\frac{HC}{L}\right)^\beta + A\beta \left(\frac{K}{L}\right)^\alpha \left(\frac{HC}{L}\right)^\beta \\
 &= (1-\alpha-\beta+\beta) \left(\frac{K}{L}\right)^\alpha \left(\frac{HC}{L}\right)^\beta \\
 &= A(1-\alpha) \left(\frac{K}{L}\right)^\alpha \left(\frac{HC}{L}\right)^\beta
 \end{aligned}$$

This will be defined here as the 'competitive' wage W^C since it is the average earnings implied by marginal productivity theory. Accordingly, as outlined below comparisons could be made between competitive and actual earnings in a given industry.

There are various ways in which the above framework could be applied to test aspects of the competitive hypothesis. Essentially, these would involve testing three inter-related issues: (1) whether or not production conditions in a given industry can be summarized by an estimated production function; (2) the extent to which labour, including embodied human capital, is being paid in accordance with marginal productivity theory; and (3) the extent to which equilibrium has been achieved within and among the different industries and/or sub-regional labour markets in the Atlantic Provinces.

Three possible methods for testing the competitive hypothesis will be outlined here. First, the framework could be used to analyze the forces underlying inter-industry differentials within the Atlantic Region. Once the production functions were estimated and the wage equations derived for each industry, the average values (ie. for the Region as a whole) of the capital to labour and human capital to labour ratios could be inserted into each industry wage equation in order to compute the competitive wage for that industry. Then, actual wages for the 26 industries could be regressed on competitive wages to indicate the extent to which actual wages are explained by the underlying technology in each industry (and thus implicitly the competitive hypothesis) and the degree to which the residual variation can be explained by such variables as the institutional setting in each industry (that is, unionization and the extent of monopoly power in product markets), the urban orientation of the industry, and perhaps discrimination eg. the percentage of females in the industry labour force.

Second, this theoretical framework could be used to analyze intra-industry differentials in the Atlantic Region. For a given industry, the capital to labour and human capital to labour ratios in each county would be inserted into the industry wage equation in order to compute the 'competitive' wage by county by industry. To the extent that an

'equilibrium' wage has been determined in each industry, the competitive wage, and the K/L and HC/L ratios, should be the same across all counties, and the competitive wage should equal the actual wage in each county. These comparisons by industry would provide a test of the 'strong' competitive hypothesis in which all three of the above conditions are met. Alternatively, we could perhaps test a 'weak' competitive hypothesis by noting that HC/L and K/L ratios together with competitive and actual wages, can differ across counties -- implying only partial adjustments among counties to initial differences in factor proportions -- but that the actual wage should equal the competitive wage in each county. The extent to which this 'weak' hypothesis is realized can be tested by regressing actual wages for the 46 counties on competitive wages in a given industry. Other variables would then be added to the regression equation in an attempt to explain the residual variation; these would include the level of urbanization (either specific to the industry or county-wide) by county and possibly within industry differences in the extent of unionization.

As a final intra-industry test, the above information would allow us to replicate the work of Scully (58) for the Atlantic Region. This would involve estimating 26 industry equations through regressing actual industry wages by county

on the capital and human capital to labour ratios in each county, plus other factors such as the degree of industry unionization in each county and possible discrimination variables. Both log-linear and linear specifications would be tested. The best specification together with the size of the intercept term and the coefficients on the capital-labour and human capital-labour ratios would offer a rough guide to the underlying production process in each industry, since implicitly we would be estimating the wage equation (11) which was derived above for the Cobb-Douglas production function. This would provide a more indirect test of the competitive hypothesis but it would suggest whether longer term forces or both long and short term factors are the primary determinants of earnings in a given industry. According to Reder (53) the longer term competitive hypothesis would suggest that the human capital to labour ratios should be the dominant explanatory variables in the wage determination equations, while if both long and short term forces are operative, both the K/L and HC/L ratios would be significant in the regression equations. Similarly, a regression could be run for all 26 industries, using actual earnings by industry (the Atlantic average) as the dependent variable, and the (all-county) average K/L and HC/L ratios as the main explanatory variables. To the extent that Reder's longer term competitive hypothesis

/...

holds, the HC/L ratio should be the dominant variable in this equation, further suggesting that actual earnings differentials among counties would be largely explained by differences among industries in the skill mix (ie. the human capital to labour, ratio).

Third and most important for the present exercise, the above theoretical framework could be applied to test whether or not inter-regional wage differentials in the Atlantic Region can be largely explained by differences in production conditions among industries and thus implicitly by the competitive hypothesis. This could be done in the following manner: (i) competitive wages by industry by county would be derived by inserting the actual K/L and HC/L ratios by county into the derived industry wage equations; (ii) the competitive wages would be formed into a matrix (W^C) to replace matrix W in equation (2) above; (iii) through matrix multiplication, average 'competitive' earnings would be computed for each county, using equation (2); (iv) actual average earnings by county would then be regressed on 'competitive' average earnings by county in order to discover to what extent actual earnings are explained by the 'average' production conditions in each county, that is, the average production conditions arising out of a county's industry mix; (v) as the final step, other variables, such as the level of

/...

unionization, the level of urbanization and the rate of labour force utilization, could be added to the equation to attempt to explain the residual variation in actual earnings. As discussed in the Methodology chapter, the actual methodology employed in this exercise represents an attempt to approximate this procedure.

Before closing this Appendix, mention should be made of a further strand of theory that runs through this analysis. This is the Heckscher-Ohlin theory of international and interregional trade and production or more precisely the Heckscher-Ohlin theory as further developed by Samuelson (57) and Mundell (43). The Heckscher-Ohlin hypothesis states that a region or country will produce and export those goods that uses most intensively its relatively abundant factor of production. Thus, in its most trivial sense, the theory would predict that, in areas of the Atlantic where natural resources are most abundant, primary products would dominate their export mix, while in areas where skilled labour is most plentiful, their exports would be largely made up of skill-intensive goods and services.

Beyond these 'trivial' predictions, the theory offers some useful predictions concerning factor prices and factor movements. In this regard, Mundell showed that, given perfect factor mobility but imperfect mobility of commodities

/...

and inequality in factor prices between countries (or regions), factors would move in such a way as to bring about factor price and commodity price equalization, and to eliminate factor movements. Alternatively, Samuelson has indicated that, under the following assumptions: (i) identical, linear, homogeneous production functions for each good between countries; (ii) diminishing returns along isoquants for each good; (iii) irreversible factor intensities at all possible factor price ratios; (iv) perfect competition, so that factor returns are valued according to their marginal products; and (v) the number of commodities exceeds the number of factors; factor-price equalization will result from trade, even in the absence of factor movements, by establishing a unique relationship between relative factor and commodity prices. Placing the Mundell and Samuelson analyses in an inter-regional context, Richardson (55) argued that, between regions, there would be a higher degree of, but imperfect, mobility of factors compared with between countries, and a close approximation to free trade. Thus, he suggested that commodity flows between regions may compensate for the imperfections in factor markets.

There are three conditions for equilibrium under the factor price equalization theorem: (a) equal factor prices between regions; (b) equal marginal physical products of each factor in the production of each good in every region;

/...

- 15 -

(c) with zero transport costs, equal prices of identical goods in all regions. The first two conditions are particularly relevant to the present exercise, for, given the requisite data, they could be tested applying the theoretical framework sketched out above. In fact, satisfying the first two conditions would be equivalent to satisfying the 'strong' competitive hypothesis analyzed previously, that is, that differences in wages between different regions and industries should be no greater than can be explained by skill (ie. human capital) differentials.

Richardson offers several reasons why the factor-price equalization model is more relevant to an inter-regional system than an international one, beyond the obvious reason of freer commodity and factor flows. These would include (a) more similar production functions across regions than across countries; (b) reflecting smaller variation in relative factor prices, less danger of factor intensity reversals, even if the relevant production functions are of the C.E.S. variety; (c) more comparable demand functions. However, Richardson concludes that "...acceptance of the view that the assumptions of the factor-price equalization model bear a closer approximation to the realities of a regional than an international system does not imply that price equalization will in fact occur.^{5/} The factors underlying this conclusion are discussed in sub-section 2 (e) of the text. If a direct

/...

test were possible for the Atlantic, it would be most surprising to find the data totally consistent with the factor-price equalization model, despite the relative 'homogeneity' of the Region.

Because of severe data constraints, it is not possible to directly test the theoretical framework outlined in this section. No comprehensive output data is available by industry and county -- and for some service industries, many of which play an important role in the Atlantic economy, there is little agreement concerning how output should be measured -- while data on capital and human capital by industry and county is totally nonexistent on an industry by county basis. And with regard to earnings, only all-industry averages by county are available, there being no comprehensive data by industry by county. The only available information is employment (and thus employment shares) by county for 26 industry groups. This information does however allow us to approximate the third procedure discussed above. This is discussed at the beginning of the Methodology section.

FOOTNOTESAPPENDIX A

^{1/} See Denton (17) and Appendix C. Rather than usual matrix multiplication, the W^a for the (j) counties could also be computed by calculating the inner product for each county.

^{2/} A production function of the form:

$$Q_i = f(K_i, E_{i1}, E_i, \dots, E_{in})$$

where the $E_{i1} \dots E_{in}$ represent different skill levels is probably a more realistic formulation, but it would also be much more difficult to work with. Therefore, it is assumed that the different types of skills can be aggregated into a single human capital measure.

^{3/} Whereas the C.E.S. production function assumes constant returns to scale but non-unitary (although constant) elasticity of substitution. For a three factor production function, the C.E.S. function would take the following form"

$$Q_i = A_i [\alpha K_i^\rho + \beta HC_i^\rho + \delta L_i^\rho]^{-\frac{1}{\rho}}$$

where $\alpha + \beta + \delta = 1$

and (the elasticity of substitution) = $\frac{1}{1+\rho}$

Thus, when $\rho = 0, \rho = 1$ and we again have the usual Cobb-Douglas function.

FOOTNOTES (CONT'D)

APPENDIX A

4/ The regression to be estimated would be derived as follows; starting from:

$$Q = AK^\alpha HC^\beta L^{(1-\alpha-\beta)}$$

and dividing through by L

$$\frac{(Q)}{(L)} = \frac{AK^\alpha HC^\beta L^{(1-\alpha-\beta)}}{L}$$

$$\text{then } \frac{(Q)}{(L)} = AK^\alpha HC^\beta L^{-\alpha-\beta}$$

$$= A \frac{(K)}{(L)}^\alpha \frac{(HC)}{(L)}^\beta$$

Therefore, the regression to be estimated would take the following form:

$$\ln \frac{(Q)}{(L)} = \ln A + \alpha \ln \frac{(K)}{(L)} + \beta \ln \frac{(HC)}{(L)}$$

5/ Richardson (55), page 287.

APPENDIX B:

MNEMONICS

APPENDIX B:

MNEMONICS

The following provides a listing of the mnemonics used in this exercise. The mnemonics are divided into five categories: The first four focus on the "level" variables used in the cross-section regressions for 1951, 1961 and 1966. The final category deals with the "change over time" variables between 1951 and 1961.

EARNINGS AND INCOME VARIABLES^{1/}

DIFAMW:	Degree of Inequality in Family Earnings (Inter-Quartile Range Divided by the Median).
DII:	Degree of Inequality in Income Per Non-Farm Family (Inter-Quartile Range Divided by the Median)
DIW:	Degree of Inequality in Earnings (Inter-Quartile Range Divided by the Median).
EIC:	Employment Income Per Capita (\$)
FAMW:	Earnings Per Family (\$)
FW:	Average Female Earnings (\$)
ICNF:	Income Per Capita in the Non-Farm Sector (\$)
INFF:	Income Per Non-Farm Family--(\$)
IRNF:	Income Per Income Recipient, Non-Farm (\$)

MEFW: Median Female Earnings (\$)
MEMW: Median Male Earnings (\$)
MEW: Median Earnings of All Wage-Earners (\$)
MW: Average Male Earnings (\$)
PIC: Personal Income Per Capita (\$)
W: Average Annual Earnings of All Wage Earners (\$)

INDUSTRIES (PERCENTAGE OF TOTAL COUNTY EMPLOYMENT)

AG: Agriculture
COMM: Communications
CON: Construction
EH: Education and Health
FIRE: Finance, Insurance and Real Estate
FISH: Fishing and Trapping
FOR: Forestry
HE: Health
HOBS: Higher Order Business Services (Communications;
Wholesale Trade; Finance, Insurance and Real
Estate; Hotels and Restaurants; and other
Business Services)
HVAM: High Value Added Manufacturing (Pulp and Paper,
Primary Metals, Metal Fabricating and other
Manufacturing)
IR: Industry Residual (Utilities and "Industry not
Specified")
LVAM: Low Value Added Manufacturing (Food Products,
Wood Products and Textiles and Clothing)

MAN: Total Manufacturing

MIN: Mining

OHVAM: Other High Value Added Manufacturing (HVAM
excluding Pulp and Paper)

OS: Other Services (Mainly Business Services)

PA: Public Administration and Defense

PESE: Personal Services (Hotels and Restaurants, and "Other")

PP: Pulp and Paper

PRIM: Primary (Agriculture, Forestry and Fishing)

RS: Residentiary Services (Retail Trade and Other
Personal Services)

SEC: Secondary (Mining, Manufacturing and Construction)

TERT: Tertiary (Transportation, Trade, Services
and Public Administration)

TOSE: Total Services (Personal, Business and Education
and Health)

TOTR: Total Trade (Wholesale and Retail)

TRANS: Transportation and Storage

TTRANS: Total Transportation (Transportation, Storage
Communications and Utilities)

UT: Utilities

WHTR: Wholesale Trade

OCCUPATIONS (PERCENTAGE OF TOTAL COUNTY EMPLOYMENT)

CLE: Clerical

CLESA: Clerical and Sales

CRPW: Craftsmen and Production Workers

SSEC: Shift in Secondary
STERT: Shift in Tertiary
STOSE: Shift in Total Services
STORTR: Shift in Total Trade
STOWC: Shift in Total White Collar
STTRANS: Shift in Total Transportation

FOOTNOTE

APPENDIX B:

^{1/} All Earnings and Income Figures are Annual Averages.

NFD: Newfoundland Dummy Variable (Value of One for Ten Census Divisions in Newfoundland, Zero for all Others)

NWVAEM: Non-Wage Value Added Per Employee in Manufacturing (Proxy for Physical Capital Per Employee) (\$/Employee)

PFELF: Per Cent Female in Employed Labour Force (%)

PFR: Francophones as Percentage of Total Population (%)

POP: Total Population

PR: Labour Force Participation Rate (Labour Force as Percentage of Labour Force Source Population) (%)

RSC: Retail Sales Per Capita (\$/Capita)

SE: Seasonality (Per Cent of Wage Earners Working 26 Weeks or Less Per Year) (%)

UD: University Dummy Variable (Value of One for Census Division Nine in Newfoundland, Antigonish and Kings Counties in Nova Scotia; Gloucester, Kings, Restigouche and Victoria Counties in New Brunswick; Zero for all Others)

UR: Unemployment Rate (%)

URB: Per Cent Urban (Urban Population as Percentage of Total) (%)

VACM: Value Added Per Capita in Manufacturing (\$/Capita)

VAEM: Value Added Per Employee in Manufacturing (Proxy for Physical and Human Capital Per Employee) (\$/Employee)

WPR: Working Population Ratio (Labour Force Source Population as Percentage of Total Population)

CHANGE OVER TIME VARIABLES: 1951 - 1961

A: Percentage Change or Index

- GE: Growth in Employment (Index, no change = 100)
- GLFP: Growth in Labour Force (Source) Population
(Index, no change = 100)
- GMEFW: Growth in Median Female Earnings (%)
- GMEMW: Growth in Median Male Earnings (%)
- GMEW: Growth in Median Earnings of all Wage Earners (%)
- NMR: Net Migration Rate (Net Migrants Divided by the
1951 Population) (%)
- NRI: Natural Rate of Increase (Births Minus Deaths
Divided by 1951 Population) (%)

B: "Shift" Variables (1961 Value Minus 1951 Value)

- CER: Change in Employment Ratio
- CSE: Change in Seasonality
- CURB: Change in Urbanization
- SAG: Shift in Agriculture
- SCON: Shift in Construction
- SFIRE: Shift in Finance, Insurance and Real Estate
- SFISH: Shift in Fishing
- SFOR: Shift in Forestry
- SMAN: Shift in Total Manufacturing
- SMIN: Shift in Mining
- SPRIM: Shift in Primary

- 7 -

SSEC: Shift in Secondary
STERT: Shift in Tertiary
STOSE: Shift in Total Services
STORTR: Shift in Total Trade
STOWC: Shift in Total White Collar
STTRANS: Shift in Total Transportation

FOOTNOTE

APPENDIX B:

¹/All Earnings and Income Figures are Annual Averages.

APPENDIX C:

SIMPLE CORRELATION ANALYSIS

APPENDIX C:

SIMPLE CORRELATION ANALYSIS

As the review of the literature emphasizes, most quantitative work in this field -- in particularly the work of Kuznets (33 and 34) and Perloff (48 and 49) -- has been limited to the simple pair-wise associations between earnings of per capita income on the one hand and other variables -- industry, occupation etc. -- on the other. Accordingly, simple correlation analysis offered an appropriate starting point for the present analysis.

The simple correlation results also provides a useful backdrop to the analysis in the text and the following appendices for it gives the reader some "feel" for the underlying relationships. However, as has been suggested earlier, simple correlations can often be misleading and thus must be interpreted with care.

Table C-1 presents the simple correlation coefficients between average annual earnings and a broad range of other economic variables. Further information regarding the mnemonics used here is outlined in Appendix B.

TABLE C-1

SIMPLE CORRELATION COEFFICIENTS
BETWEEN AVERAGE ANNUAL EARNINGS, W (\$), AND OTHER VARIABLES

CORRELATION
COEFFICIENTS

Other Earnings, Etc. Variables

Average Male Earnings : MW (\$)	.99
Average Female Earnings: FW (\$)	.90
Income Per Recipient, Non-Farm: IRNF (\$)	.93
Income Per Capita, Non-Farm INCF (\$)	.64
Earnings Per Family: FAMW (\$)	.97
Income Per Non-Farm Family: INFF (\$)	.84
Personal Income Per Capita: PIC (\$)	.69
Employment Income Per Capita: EIC (\$)	.79
Dispersion Indices: the Inter-Quartile Range Divided by the Median:	

Earnings: DIW	-.73
Family Earnings: DIFAMW	-.85
Income Per Non-Farm Family: DII	-.43
Median Earnings: MEW (\$)	.99
Median Male Earnings: MEMW (\$)	.99
Median Female Earnings: MEFW (\$)	.89

Industries (Percentage of Total County Employment)

Primary: PRIM	-.84
Agriculture: AG	-.35
Forestry: FOR	-.36
Fishing and Trapping: FISH	-.64
Secondary: SEC	.11
Mining: MIN	.31

CORRELATION
COEFFICIENTS

Total Manufacturing: MAN	-.09
Low Value Added Manufacturing: LVAM	-.55
High Value Added Manufacturing: HVAM	.39
Pulp and Paper: PP	.25
Other High Value Added Manufacturing: OHVAM	.32
Construction: CON	.06
Tertiary: TERT	.81
Transportation and Storage: TRANS	.05
Total Trade: TOTR	.29
Wholesale Trade: WHTR	.37
Finance, Insurance and Real Estate: FIRE	.50
Total Services: TOSE	.13
Health: HE	.26
Personal Services: PESE	.10
Other Services (Mainly Business): OS	.27

Three-Part Breakdown Used in Model:

Higher Order Business Services: Communications; Wholesale; Finance, Insurance and Real Estate; Hotels and Restaurants; and other Business Services: HOBS	.43
Residential Services: Retail and other Personal Services: RS	.15
Education and Health: EH	.05
Public Administration and Defense: PA	.59

CORRELATION
COEFFICIENTSOccupations (Percentage of Total
County Employment)

Total White Collar: TOWC	.57
Managerial and Professional: MAPR	.22
Managerial: MA	.02
Professional (including Technical): PROF	.31
Clerical and Sales: CLESA	.69
Clerical: CLE	.74
Sales: SA	.41
Service and Recreation: SR	.53
Transport and Related Occupations: TRO	.07
Loggers and Fishermen: LF	-.68
Farmers and Farm Workers: FAFW	-.34
Miners: MI	.23
Craftsmen and Production Workers: CRPW	.20
Labourers: LAB	.13

Other Variables

Per Cent Urban (%); Urban Population as Per Cent of Total: URB	.68
Employment Ratio (%); Employment as Per Cent of Total Population: ER	.40
Participation Rate (%): PR	.58
Unemployment Rate (%): UR	-.25
Seasonality (%); Per Cent of Wage Earners Working 26 Weeks or Less Per Year: SE	-.79
Hours (%); Per Cent of Wage Earners Working Less than 35 Hours a Week: HO	-.66
Level of Education (%); Per Cent of Population 5 and over not attending school with at least two years of secondary schooling: LED	.65

CORRELATION
COEFFICIENTS

Value Added Per Employee in Manufacturing: VAEM	.48
Capital Per Farm: CF	.41
Growth in Employment: GE	.66
Per Cent Female in Employed Labour Force: PFELF	.22
Per Cent French; Francophones as Per Cent of Total Population; PFR	-.20

The first part of the table lists the coefficients between average annual earnings and the other earnings and income variables that could be used in regression analysis. As might be expected, the correlation coefficients are very high, i.e. .9 or above for the other "productivity" measures, such as male and female earnings, income per recipient in the non-farm sector and median earnings.^{1/} The very close association between average and median earnings is very encouraging, since median earnings is the only income or earnings variable available for the 1951 and 1951-61 regression analysis. On the other hand, the coefficients are lower -- although still highly significant at the usual confidence levels^{2/} -- in the case of the "welfare" measures such as personal and employment income per capita. For example, the coefficient indicates that, acting on its own, earnings "explains" only 62 per cent of the intercounty variation in employment income per capita (a variable that is also used in the model). Accordingly,

there remains considerable variation to be explained by the size of the employment base.

Finally, consistent with the findings of Williamson (66), Kuznets (35) and others, the table indicates a strong negative association between earnings and the degree of inequality in earnings and income. The association is though much stronger for the two earnings measures than the one income measure which is the one applied in the model to indicate the degree of inequality in income. This is perhaps not surprising for the latter includes not only the effects of productivity differentials but also the effects of government transfers and intercounty variations in the size of the employment base. Accordingly, the coefficient between employment income per capita and the degree of inequality in the income per non-farm family series is somewhat higher, coming to $-.63$.

Turning to the industry coefficients, there are a number of results here that differ from those of previous authors.^{3/} Consistent with previous findings, there is a strong negative association in the Atlantic between earnings and the share of employment in primary activities, with all three primary industries showing a strong negative relationship. On the other hand, the correlation between earnings and the proportion of employment accounted for by the secondary sector is very weak, with a coefficient of $.11$ which is not

statistically significant. (The secondary sector here coincides with the (M) sector of Kuznets.). This result differs from those of Kuznets, Perloff et al. It should be emphasized that this is mainly a reflection of the strong role played by lower value added (mainly primary) manufacturing within the Atlantic economy; the association between this sub-sector and earnings is strongly negative (which accords with previous findings). The association between earnings and the remaining segments of the secondary sector is generally positive, although it is never very strong.

Finally, the table suggests a very strong positive association between earnings and the share of employment in tertiary activities. This again diverges from the results of other authors where the association -- while generally positive -- was in the main fairly weak. Looking at the different components of the tertiary sector, the association was particularly strong in the case of public administration and defense. These results would suggest therefore that the service industries are the "high-wage" activities in the Atlantic Region, whereas in most developed economies that role is generally played by the secondary sector. It should be emphasized again, however, that these are only pair-wise associations and may not hold up under multiple regression analysis.

4.5

OF/13

- 8 -

The simple correlation results for the occupation categories generally accord with those for the industry groupings. The coefficients are significantly positive for the services oriented occupations such as the professional, clerical, sales and service and recreation groups. In contrast, the coefficient is strongly negative for loggers and fishermen, and farmers and farm workers, while the association between earnings on the one hand, and miners plus craftsmen and production workers on the other, is positive but not statistically significant. One noteworthy result is the negligible association between earnings and the managerial group which, at first glance, would seem to be counter-intuitive. This likely reflects the fact that smaller establishments tend to predominate in the less developed parts of the Atlantic, resulting in those counties in a relatively higher proportion of management to total employment. (Many of these establishments might be family firms with only one or two workers.)

There are probably few surprises in the coefficients for the other variables. As would be expected, the coefficients are significantly positive for the level of urbanization, the employment ratio, the participation rate (which is a major part of the employment ratio), the level of education, the growth in employment and the two capital

/...

- 9 -

proxies, value added per employee in manufacturing (which ideally is a measure of both physical and human capital per employee)^{4/} and capital per farm. On the other hand, they are significantly negative for the degree of seasonality and the hours variables. These are the expected results since earnings here is an annual figure; thus both negative coefficients reflect structural or compositional differences between counties, or more specifically inter-county differences in the size of the employment base when the latter is broadly defined to include seasonality and below average hours. One result that may not have been anticipated is the fact that the coefficient between earnings and the unemployment rate, while negative, is only significant at the 90 per cent confidence level. This could reflect one of two factors. First, in highly agricultural areas such as on Prince Edward Island, both earnings and unemployment tend to be low, the major problem there being underemployment. Second, in areas such as those centred on Corner Brook, Grand Falls and other pulp and paper towns, the high levels of unionization could be associated with both higher earnings and above average rates of unemployment.

Some simple correlation results are also available for changes over time between 1951 and 1961. In the following table, the "dependent" variable is the increase in median

- 10 -

earnings over that decade while the other variable is generally the percentage shift in the share of employment of a given industry or occupation category. For example, for Census Division One in Newfoundland, the share of employment accounted for by primary activities declined from 15.09 per cent in 1951 to 5.21 per cent in 1961; therefore, for the shift in primary variable the value of Census Division One's observation would be -9.88 per cent. The categories below are generally more aggregated than in the previous table, reflecting the fact that the 1951 Census offers much less detailed information.

TABLE C-2

SIMPLE CORRELATION COEFFICIENTS
BETWEEN THE GROWTH IN MEDIAN EARNINGS: GMEW; 1951 - 1961
AND CHANGES IN OTHER VARIABLES

	<u>CORRELATION</u> <u>COEFFICIENTS</u>
<u>Other Earnings Variables</u>	
Growth in Median Male Earnings: GMEW	.98
Growth in Median Female Earnings: GMEFW	.50
<u>Shifts in Employment Shares:</u>	
<u>Industry and Occupation</u>	
Primary: PRIM	-.67
Agriculture: SAG	.02
Forestry: SFOR	-.30
Fishing and Trapping: SFISH	-.56

CORRELATION
COEFFICIENTS

Secondary: SSEC	.38
Mining: SMIN	.19
Total Manufacturing: SMAN	.32
Construction: SCON	.23
Tertiary: STERT	.31
Total Trade: STOTR	-.23
Total Services: STOSE	.31
Total Transportation: STTRANS	.09
Finance, Insurance and Real Estate: SFIRE	-.14
Total White Collar: STOWC	-.09
<u>Others</u>	
Change in Per Cent Urban: CURB	.51
Change in the Employment Ratio: CER	.06
Change in Seasonality: CSE	-.77

For the reason outlined in the previous paragraph, this listing is less detailed than the one in Table C-1, but it can still be used for comparisons with the cross-section analysis. Similar to the latter, Table C-2 indicates a strong negative association between the growth in earnings and the shift in share experienced by primary activities. The implication is then that the greater the decline in share of the primary industries, the larger the increase in median earnings. This very much agrees not only with our own cross-section work but with the findings of others.^{5/} It should be

noted, however, that, within the sector, the coefficients are significantly negative for only two of the three industries, forestry and especially fishing.

In contrast, there is some divergence between Tables C-2 and C-1 for the remaining two sectors. Table C-2 indicates a significantly positive -- although still fairly weak -- correlation between the growth in earnings and the increase in share in the secondary sector, whereas in cross-section the coefficient was not significant. Within the sector, the change over time coefficient for manufacturing is positive and significant -- although perhaps still fairly small -- compared with negative (although insignificant) for cross-section. Moreover, while the coefficient between growth in earnings and the shift in tertiary is positive and statistically significant, it is much smaller than the cross-section coefficient. This reflects the negative and almost significant coefficient for total trade in the change over time analysis, compared with a significantly positive correlation in cross-section work. Following from this, on a pair-wise basis, no significant association was found between white-collar and earnings in the change over time analysis, whereas these two were highly positive in Table C-1. All of this may be inter-related with the surprisingly small correlation between the growth in the median earnings of all wage earners and growth

- 13 -

in earnings for female wage earners, with a coefficient of only .80. (The interrelationship flows from the fact that most working females are employed in the tertiary sector.)

Finally, Table C-2 indicates a very strong positive association between the growth in earnings and the increase in urbanization, and a strong negative relationship between earnings growth and the change in seasonality. Both results are in accord with the cross-sectional findings. In general then, the simple pair-wise comparisons indicate a rough correspondence between the cross-sectional and time series findings, but important divergences do exist.

/...

FOOTNOTESAPPENDIX C:

- 1/ It was decided that the correlation between average earnings and the other "productivity" measures was sufficiently high to make unnecessary the estimation of separate regressions utilizing these other measures as left hand variables. The one exception was median earnings which was used in a 1961 regression in order to provide a "benchmark" for a 1951 regression utilizing median earnings as the dependent variable.
- 2/ With 45 degrees of freedom, the correlation coefficient must have an absolute value of .28 or more to be statistically significant at the 95 per cent confidence level.
- 3/ In particular, Kuznets (33 and 34) and Perloff (48 and 49). It should be noted that here the "dependent" variable in the pair-wise correlations is average earnings whereas in previous studies it was generally personal or employment income per capita. However, as Table C-1 indicates, the correlation between earnings and employment income per capita is quite high. The expectation is then that the results would not differ greatly if employment income per capita -- or some other "welfare" measure -- were used in the analysis, particularly with respect to the industry and occupation variables.
- 4/ Since it includes both wages, i.e. the returns to "homogeneous" labour and human capital, and profits, i.e. the returns to physical capital. This measure therefore should be higher the higher the average wage -- which should be a positive function of human capital intensity -- and the higher the "relative" profit (i.e. profit per employee) which should be a positive function of physical capital intensity.
- 5/ In particular the "time series" work of Kuznets (34 and 35) and Borts (4 and 5).

APPENDIX D:

STANDARDIZATION ANALYSIS:

METHODOLOGY AND RESULTS

APPENDIX D:

STANDARDIZATION ANALYSIS: METHODOLOGY AND RESULTS

Two standardization calculations were completed for this exercise, those for rate constant earnings for all wage earners and for male rate constant earnings. The former for a given county (j) is computed by multiplying wages (W) in each industry (i) in the Atlantic as a whole by the industry's employment shares (ES) in county (j), and then summing through all (k) industries i.e:

$$\begin{aligned}(1) \quad RCW_j &= W_{1A} ES_{1j} + W_{2A} ES_{2j} + \dots + W_{kA} ES_{kj} \\ &= \sum_{i=1}^k W_{iA} ES_{ij}\end{aligned}$$

Rate constant male earnings are calculated in a similar manner except that male wages by industry and employment shares for the male labour force are used.

Industry constant earnings (not computed here because of the lack of data) are the opposite of rate constant earnings and essentially involve applying the distribution of employment in the Atlantic as a whole to industry earnings in each county. Accordingly, industry constant earnings (ICW) in county (j) are:

/...

$$(2) \quad ICW_j = W_{1j} ES_{1A} + W_{2j} ES_{2A} + \dots + W_{kj} ES_{kA}$$

$$= \sum_{i=1}^k W_{ij} ES_{iA}$$

After calculating "rate constant" earnings for both all wage earners and male wage earners, these two measures were regressed on their counterpart "unstandardized" series. The coefficient of determination (i.e. the square of the simple correlation coefficient) came to .79 for the all-wage earners regression, and was marginally higher, .82 between actual male earnings and rate constant male earnings. The suggestion is therefore that the intercounty variation in earnings declines very little by assuming the same rates of pay in a given industry across all counties in the Atlantic. Following from Hanna's (24 and 25) arguments, the contention could be made that the remaining variation can be accounted for by industrial distribution, including both first and higher order effects. As suggested in the text, this conclusion is perhaps a little too strong, but still the above results imply an important role for differences in industry structure in generating earnings differentials within the Atlantic Region. This judgement received further support from the multiple regression results and the specification of the model.

APPENDIX E:

SINGLE EQUATION ESTIMATION:

EARNINGS AND INCOME PER CAPITA

APPENDIX E

SINGLE EQUATION ESTIMATION: EARNINGS AND INCOME PER CAPITA

Originally, single equation estimation using Ordinary Least Squares was to provide the focal point for the present exercise. As the Introduction indicates, however, this work has broadened considerably, with the primary objective becoming the development of a full econometric model. Still, single equation estimation, particularly in relation to a wage (i.e. earnings) determination equation, played an important role in this exercise, for it provided the necessary first step in the specification of the model. Also, estimating such an equation proved to be one of the more difficult tasks in the entire exercise. For these reasons, it was thought necessary to include an appendix discussing in full the different equations attempted for the wage determination equation. This appendix will also outline the other single equation work that was attempted in the course of this project.

The regression strategy outlined in the Methodology section was applied in estimating a wage determination equation. At the outset, attempts were made to specify an equation using industry alone. The simple correlation analysis provided some useful guidelines concerning the variables that might be included. First, the primary industries would doubtless play a critical role. Second, the correlation coefficients

/...

suggested that dividing the manufacturing sector into low and high value added (or wage) segments would probably be necessary. This was done using value added per employee data provided by the 1961 Census of Manufacturers. For the Atlantic Region, value added per employee (on a total activity basis) for the eight manufacturing industry groups in the (Population) Census came to the following:

Value Added Per
Employee: 1961 (\$)

Low Value Added Manufacturing

Food Products	5,382
Textiles and Clothing	4,132
Wood Products	4,305

High Value Added Manufacturing

Pulp and Paper	11,211
Printing	6,381
Primary Metals	5,967
Metal Fabricating	6,250
Other Manufacturing	6,707

The division outlined above was applied in further work.

Third, the correlation work suggested that some segments of the tertiary sector -- such as wholesale trade, finance, insurance and real estate, and public administration and defense -- could be major factors in the determination of wages in the Atlantic Region. What was ultimately done was to group together those activities that are largely local

/...

market oriented (or residentiary) and those that, at least potentially, could be "exported" across county and provincial borders. This will be further discussed below.

An early specification that held some promise was as follows:^{1/}

(1) $\hat{W}$ =	2860.15 (599.02) (4.77)	-25.15 PRIM (6.82) (-3.69)	-21.31 LVAM (8.79) (-2.42)
	+12.35 HVAM (9.87) (1.35)	+7.59 TRANS (13.63) (0.56)	-0.75 TOTR (15.48) (-.05)
	+61.62 FIRE (87.30) (.71)	+2.66 HE (23.00) (.12)	-40.58 PESE (24.00) (-1.69)
	+4.40 OS (74.21) (.005)	+12.05 PA (7.66) (1.57)	

SEE = 206.94 $R^2 = 0.876$ $\bar{R}^2 = 0.840$ CV = 9.17%

where the variables are defined in Table A-1. The co-efficient on the PRIM variable can be interpreted as follows: given a one percentage point difference in primary's share of employment between two counties, average annual earnings would (on average) be expected to differ by \$25.15, other things being equal.

Interpreting the results, the signs on most co-efficients seemed to be defensible, while the t - statistics were greater than one for five of the ten explanatory variables. Perhaps most important for an initial specification, the X variables were

explaining 84 per cent of the intercounty variation in earnings, which at least suggested that we were on the right track.

One further possibility indicated by equation (1) was that it might be useful to split services between what came to be called higher order business services and residentiary services. The positive sign on FIRE suggested that the former would likely be positively related with earnings, while the negative sign on PESE suggested perhaps a negative association between earnings and residentiary services.

Higher order business services were defined to include wholesale trade, finance, insurance and real estate, hotels and restaurants, other business services and ultimately communications. The preliminary definition also included education and health but this was later split out. Residentiary services is made up of retail trade and other personal services.

Based on this, further regression were run to improve on the specification in equation (1). Table E-1 outlines those that are perhaps most representative.

/....

TABLE E-1

DEPENDENT VARIABLE: AVERAGE ANNUAL EARNINGS (\$) 5

INDEPENDENT VARIABLES: INDUSTRY SHARES

EQ. NO.	CONSTANT	AG	FOR	FISH	LVAM	HVAM	EH	HORS	RS	PA
(2)	2,910.4 (346.9) (8.39)	-25.88 (5.33) (-4.85)	-29.91 (8.58) (-3.49)	-27.89 (6.48) (-4.31)	-26.16 (8.56) (-3.06)	8.74 (8.00) (1.09)	-0.66 ^{1/} (8.98) (0.07)			9.89 (5.11) (1.93)
	F-STAT (7,38) =	32.91	SEE = 211.97	R ² = .858	R ² = .832	CV = 9.39%				
(3)	3,300.8 (390.1) (8.46)	-28.66 (5.45) (-5.26)	-33.31 (7.95) (-4.19)	-31.35 (6.07) (-5.16)	-26.98 (7.71) (-3.49)	8.85 (7.52) (1.18)			-19.79 (16.67) (-1.19)	6.35 (5.04) (1.26)
	F-STAT (7,38) =	34.33	SEE = 208.16	R ² = .863	R ² = .838	CV = 9.22%				
(4)	3,160.6 (409.6) (7.72)	-27.49 (5.54) (-4.96)	-28.45 (9.09) (-3.13)	-27.65 (6.93) (-3.99)	-24.67 (7.98) (-3.09)	11.08 (7.77) (1.43)		17.98 ^{2/} (16.43) (1.09)	-28.42 (18.40) (-1.54)	7.96 (5.24) (1.52)
	F-STAT (8,37) =	30.35	SEE = 207.62	R ² = .868	R ² = .839	CV = 9.20%				
(5)	3,400.7 (451.1) (7.54)	-28.65 (5.58) (-5.13)	-28.29 (9.02) (-3.13)	-28.21 (6.90) (-4.09)	-28.09 (8.40) (-3.34)	9.64 (7.80) (1.24)	-19.30 (15.67) (-1.23)	27.47 ^{2/} (18.04) (1.52)	-34.03 (18.84) (-1.81)	5.19 (5.67) (0.92)
	F-STAT (9,36) =	27.52	SEE = 206.19	R ² = .873	R ² = .841	CV = 9.14%				

^{1/} i.e. EH + HORS but with HORS not including communications.

^{2/} But excluding communications.

TABLE E-1 (CONT'D)
 DEPENDENT VARIABLE: AVERAGE ANNUAL EARNINGS (\$) $\bar{Y}$
 INDEPENDENT VARIABLES: INDUSTRY SHARES

EQ. NO.	CONSTANT	PPM	AG	FOR	FISH	LVAM	HVLM	PP	OHVAM	EH	HOBS	RS	PA	MAN
(6)	3,300.6 (458.4) (7.20)		-27.59 (5.67) (-4.87)	-27.43 (9.18) (-2.99)	-28.38 (6.81) (-4.16)	-27.40 (8.42) (-3.25)	10.40 (7.86) (1.32)			-17.42 (15.07) (-1.16)	24.74 ^{3/} (15.56) (1.59)	-32.27 (18.15) (-1.78)	6.03 (5.71) (1.06)	
	F-STAT (9,36) = 27.69 SEE = 207.62 $R^2 = .874$ $\bar{R}^2 = .842$ CV = 9.11													
(7)	3,296.6 (436.3) (7.56)	-27.76 (5.25) (-3.66)				-27.72 (7.57) (-3.66)	10.47 (7.63) (1.37)			-17.38 (14.63) (-1.19)	25.14 ^{3/} (13.53) (1.86)	-32.32 (17.61) (-1.84)	6.05 (5.48) (1.10)	
	F-STAT (7,38) = 37.55 SEE = 200.21 $R^2 = .874$ $\bar{R}^2 = .874$ CV = 8.87													
(8)	3,068.8 (553.6) (5.54)		-32.18 (6.77) (-4.75)	-20.06 (10.96) (-1.82)	39.03 (7.71) (-5.06)					-5.37 (18.01) (-0.30)	19.84 (18.87) (1.05)	-22.47 (21.90) (-1.03)	7.98 (6.92) (1.15)	-6.25 (8.34) (-0.75)
	F-STAT (8,37) = 19.48 SEE = 250.07 $R^2 = .808$ $\bar{R}^2 = .767$ CV = 11.08													
(9)	3,424.4 (449.7) (7.62)		-29.82 (5.63) (-5.29)	-33.78 (9.56) (-3.53)	-29.21 (6.82) (-4.41)	-30.41 (8.33) (-3.65)	13.50 (7.81) (1.73)	-11.88 (14.46) (-0.82)		-23.26 (14.96) (-1.55)	30.93 (15.47) (2.00)	-30.24 (17.64) (-1.71)	3.67 (5.69) (0.65)	
	F-STAT (10,35) = 26.84 SEE = 199.38 $R^2 = .885$ $\bar{R}^2 = .852$ CV = 8.83													

TABLE E-1 (CONT'D)

DEPENDENT VARIABLE: AVERAGE ANNUAL EARNINGS
INDEPENDENT VARIABLES: INDUSTRY SHARES

EQ. NO.	CONSTANT	AG	FOR	FISH	LVAM	PP	OHVAM	EH	HOBS	RS	PA
(10)	3,683.6 (201.9) (18.25)	-32.33 (4.06) (-7.97)	-37.42 (7.67) (-4.88)	-31.79 (5.23) (-6.08)	-33.63 (6.63) (-5.08)	11.20 (6.90) (1.62)	-16.24 (12.68) (-1.28)	-27.25 (13.52) (-2.01)	29.99 (15.27) (1.96)	-35.12 (15.82) (-2.22)	
	F-STAT (9,36) = 30.26	SEE = 197.76	R ² = .883	R ² = .854	CV = 8.76%						
(11)	3,718.1 (201.8) (18.42)	-31.13 (3.98) (-7.82)	-34.88 (7.48) (-4.66)	-31.36 (5.26) (-5.96)	-32.31 (6.60) (-4.89)	12.50 (6.88) (1.82)		-25.79 (13.59) (-1.90)	27.37 (15.27) (1.79)	-42.65 (14.81) (2.88)	
	F-STAT (8,37) = 33.26	SEE = 199.47	R ² = .878	R ² = .852	CV = 8.84%						
(12)	*3,290.22 (417.08) (7.89)	-27.65 (4.95) (-5.58)	-30.18 (8.46) (-3.57)	-27.42 (6.22) (-4.41)	-27.75 (7.64) (-3.63)	15.61 (7.34) (2.13)		-20.06 (14.38) (1.39)	30.00 (15.35) (1.95)	-31.65 (17.48) (-1.81)	5.86 (5.00) (1.17)
	F-STAT (9,36) = 30.01	SEE = 198.48	R ² = .882	R ² = .854	CV = 8.79%						

Table E-1 illustrates the "groping" of the analyst towards a satisfactory specification for a wage determination equation. A few general comments should be made before looking at each equation individually. First, to the extent that pairwise correlations are a good indicator, there would seem to be few problems of multicollinearity with the industry variables as defined in the final equations in Table E-1. The highest simple correlation coefficients between any two X variables in equation 12 are .67 for higher order business services and residentiary services, and -.56 between HOBS and fishing. Further, in equation 10, fairly high correlations were indicated for other high value added manufacturing on the one hand, and HOBS and RS on the other, with coefficients of .59 and .57 respectively. These coefficients are much too low to indicate a problem, as it is generally thought that a simple correlation must be .9 or above for multicollinearity to be harmful. Still, given the large number of explanatory variables in these equations, any conclusions regarding multicollinearity must be viewed as tentative until the Farrar-Glauber equations are estimated. These were run for the final specification of the wage determination equation.

Second, among the independent variables, the three primary industries plus LVAM proved to be the most robust. Each possessed its anticipated sign; moreover, the size of each co-efficient varied relatively little from one equation to the next. Finally, the values of the coefficients were remarkably similar among the four industries. This suggested that the three primary industries - and perhaps LVAM as well - could be added together, thus "restricting" the coefficients on the three industries to the same value. This was done in equation seven with the desired results, that is, the coefficient on PRIM was essentially the same as those on the three primary industries in equation six, and more importantly, the restriction led to an improvement in the adjusted R^2 , from .842 to .850. The same desirable results were obtained when all four industries were added together (not shown in Table E-1); in this case, the $\bar{P}^2$ showed a further improvement from .850 (in equation seven) to .854. At the time, these results seemed very encouraging, for by adding these industries together, we could save on degrees of freedom. However, as will be noted later, this "restriction" becomes less viable when non-industry variables are added to the regression and is probably even less desirable in the context of a full econometric model.

/...

- 10 -

In contrast to the primary industries, public administration and defense proved to be the weakest X variable; this runs counter to the fairly high simple correlation between PA and earnings indicated in Table C-1 of Appendix C. In Table E-1, its sign was always positive -- which probably was the expected result -- but, once further service industries were included, the magnitude of the coefficient declined fairly markedly and the t-statistic hovered around one. One possibility is that the weakness of this variable reflects its heterogenous nature. In the Census, public administration and defense groups together federal government employment (including military personnel, civilian personnel on military bases, and other federal employees) with provincial and local government employment. The former is probably a high-wage activity in the context of the Atlantic Region, but rates of pay in provincial and local government are probably at or below the all-Atlantic average. Accordingly, one method for improving this variable would be to split out federal -- or at least military -- employment from the PA category. Unfortunately, the necessary data was not available for 1961.

Beyond the above industries, the explanatory variables are perhaps best discussed in relation to individual equations. In equation (2), higher order business services and education and health were added together as a first

/...

approximation of export-oriented services: The coefficient on this variable was insignificant; in contrast, the remaining coefficients seemed to be reasonable, with t-statistics all greater than one; (thus, they are adding to the adjusted R^2). In equations (3) to (5), different service industry groupings were added cumulatively to the regression. Each entered the equation with a t-statistic greater than one, while the remaining industry variables, with the exception of PA, were relatively stable. Note should be made of the negative signs on RS and EH. These seemed defensible, particularly with respect to residential services. In the author's judgement, the negative sign on residential services can be explained by two factors: (1) the relatively small variability between counties in the demand for residential services; (2) higher productivity in residential services in the more developed sections of the Atlantic, the result in part of economies of scale (since the most developed parts also tend to be the most populous). The upshot would be higher employment shares for residential services in low wage relative to high wage counties in the Region, other things being equal. Probably the same conclusions hold for education and health, although with somewhat less force as indicated by its "larger" (i.e. less negative) coefficient in equations (4) to (12). It should be noted that these negative relationships would not have been predicted from simple correlation analysis where the coefficients between RS and W, and EH and W, were positive, although not statistically significant (Table C-1).

/...

- 12 -

After equation (5) was estimated, experiments were attempted with other industry variables. It was found that communications entered the regression with a fairly large positive co-efficient; however, the standard error was more than double the size of the coefficient suggesting collinearity between communications and another x variable, probably HOBS whose standard error moved upward fairly markedly with the former's inclusion in the regression. The indication was therefore that adding communications to HOBS would lead to an improvement in specification. Equation (6) substantiated this as, relative to (5), the standard error on HOBS declined and its t -statistic increased, while most important the adjusted R^2 moved up marginally.

One further experiment was attempted at this point. In equation (8), low and high value added manufacturing were replaced by total manufacturing to check if our original judgements concerning manufacturing - which initially were based on simple correlation results - would hold up under a more complete industry equation specification. Equation (8) shows that our original conclusions were correct. The co-efficient on total manufacturing - while negative - is not statistically significant and the overall fit has deteriorated greatly, with the adjusted R^2 falling from .842 in (6) to .767. In short, restricting the coefficient on LVAM and HVAM to the same value would represent an important specification error.

/...

At this stage, equation (6) seemed to be a fairly satisfactory specification; the signs on every coefficient were supportable, all t-statistics were above one, and five of the nine coefficients on the X variables were significant, using a one-tail test. Finally, the nine industry variables were explaining 84 per cent of the variation in the dependent variable (adjusted for degrees of freedom), a fairly satisfactory result for cross-section analysis. The residuals from equation (6) however, indicated that the regression was consistently underestimating earnings in counties where pulp and paper played a major role in the economic base. Accordingly, the decision was made to split up HVAM into pulp and paper and other high value added manufacturing, and add both to the regression (see equation (9)). As was anticipated, PP entered the regression with a positive coefficient which moreover was significant at the 95 per cent confidence level. In marked contrast, the coefficient on OHVAM was negative, although the t-statistic was below one. The suggestion is therefore that, as of 1961, secondary manufacturing (which accounts for most of the OHVAM category) was not a high-wage industry in the Atlantic Region. This is certainly contrary to previous analyses. This reflects the composition of this industry in the Atlantic which is dominated by shipbuilding and other medium wage activities. Finally, by splitting up HVAM, the $\bar{R}^2$ showed an important increase from .842 (in equation (6)) to .852, suggesting an improvement in specification.

But despite this probable improvement, equation (9) posed a number of difficulties. Perhaps most important, with the inclusion of PP and OHVAM separately, public administration ceased to have a t-statistic greater than one. Moreover, as indicated in equation (10), the t-statistic on OHVAM became greater than one (in absolute value) when PA was removed from the regression, while the student's t on PA went over one again when OHVAM was left out, as illustrated by equation (12).

Therefore, it came down to a decision between PA and OHVAM and thus between equations (10) and (12). (One other possibility would be to leave both out as in equation (11). However, this appeared to the author as an inferior specification relative to both (10) and (12), because of the lower $\bar{R}^2$ and the much higher - i.e. more negative -- coefficient on RS.) It was decided, somewhat reluctantly, to go with equation (12) in the subsequent analysis. (Therefore, this equation is indicated with an asterisk, as will other "key" equations in this discussion.) This decision was based on three factors: (1) the sign on PA is more in line with a priori expectations; (2) public administration is a much more important industry in the Atlantic than OHVAM, with its average employment share across the 46 counties coming to 10.23 per cent compared with 3.21 per cent for the latter; as a consequence, the industry "coverage" is much higher in the case of equation (12);

/...

- 15 -

(3) in equation (12, there is a better "balance" between negative and positive coefficients -- following from this, the constant term is smaller, and the coefficients on the three primary industries less negative in equation (12) relative to (10).

Therefore, it was concluded that equation (12) represented a marginally better specification than equation (10). Still, other high value added manufacturing was not forgotten in further work; for example, in 2 SLS estimation, it was included in one of the wage determination equation specifications, but with no success.

In addition to the industries discussed above, regressions were also estimated using mining, construction, and transportation and storage, as right hand variables. Construction entered the regression with a negative sign, while the sign for transportation and storage was positive, with both t-statistics having an absolute value of about 0.5. The possibility is therefore that with more degrees of freedom, these two industries might have a part to play in a wage determination equation. On the other hand, the coefficient on mining was of negligible size.

One of the initial approaches considered by the author was to estimate occupation and industry wage equations separately and then bring the two together in a manner that would best

summarize how demand and supply elements together determine earnings in the Atlantic Region. As was discussed in the review of the literature, industry composition is roughly indicative of the demand for skills in a given area, while occupational distribution is representative of their current supply. A few wage equations based on occupation alone will be outlined later but in general it was concluded that this approach would not be fruitful. This conclusion stems from the fact that most of the occupational categories are too closely identified with one or a few industries, for example, farmers and farm workers with agriculture, craftsmen and related workers with manufacturing and construction, and loggers and fishermen with their two respective industries. In these cases, the inclusion of both industry and occupational variables would likely lead to serious problems of multicollinearity.

The only occupational groups that would seem to be at least partially distinct are those in the broad white collar grouping, namely, the managerial, professional, clerk and sales categories. Only these groups would seem in general to cut across industry lines. It was decided therefore to estimate regressions using these groups together with the nine industries in equation (12). To save on degrees of freedom, the managerial and professional, and the clerical and sales categories, were added together; one equation was also estimated using the total white collar grouping. (These variables are defined in Appendix A.) The results are outlined in Table E-2.

Attention should first be focussed on equation (13) which uses the total white collar category as a right hand argument. In some respects, this represents an improved specification over equation (12). The adjusted R^2 moved up sharply from .854 to .913 while the standard error of the estimate fell by nearly a quarter. In equations 14 through 16, the white collar category was divided into the two groups defined above, with the two being included separately and then together in the regression. The clerical and sales group proved to be the stronger explanatory variable, but the managerial and professional group is also highly significant in equation (16). Finally, comparing equations (13) and (16) indicates that allowing the two groups to have differing co-efficients leads to a slight improvement in specification, as the marginal improvement in the adjusted R^2 illustrates.

But while including occupation variables leads to substantial improvement in an econometric sense, the economic interpretation of the equation becomes a great deal more difficult, particularly in a regional planning context. When based on industry alone, the interpretation of the regression co-efficients is relatively straight forward for, if the equation is properly specified^{2/} they should for the most part be a function of the average level of skills, plus the capital per worker and level of unionization, in a given industry.

The operative word in the above is "average" as related to skills for, in equations based on industry alone, we are restricting all types of labour within a given industry to having the same coefficient. Assuming two types of labour, white collar WC and other O, the industry equations implicitly take the following form:

$$W = \beta_0 + \beta_1 (WC_1 + O_1) + \beta_2 (WC_2 + O_2) + \dots + \beta_k (WC_k + O_k)$$

where WC_i is white collar employment in industry (i) as a percentage of total county employment, and similarly for O_i . Accordingly, WC_i plus O_i add to the employment share of industry (i).

However, when the white collar variable is included per equation (13), a different type of restriction is imposed. In this case, white collar workers across all industries are restricted to the same coefficient, while all that remains in the industry variables are the "other" workers. Therefore, the estimated regression takes the following form:

$$W = \beta_0 + \beta_1 (O_1) + \beta_2 (O_2) + \dots + \beta_k (O_k) + \beta_{k+1} \left(\sum_{i=1}^k WC_i \right)$$

Therefore, the impact of this restriction on the industry coefficients will vary with the importance of white collar employment within those industries. Comparing equation (13) with equation (12), it is found that the impact on the

five goods producing industries plus public administration and defense is relatively small and thus the coefficients in (13) are not significantly different from those in (12). This was the anticipated result for the white collar group likely constitutes only a small fraction of the work force in these activities. In sharp contrast, marked changes in coefficients are indicated for the three service industry categories. The change is most dramatic for higher order business service whose coefficient goes from significantly positive to significantly negative. The explanation for this lies in the above mathematics. By splitting out white collar workers, all that remains in HOBS are maintenance staff, labourers and other relatively low paying occupations. And since white collar workers likely make up the vast majority of the HOBS work force, this new restriction results in a marked shift in the sign and size of HOBS coefficient. The same arguments can be used to explain the more negative coefficients on education and health and particularly residentiary services. Looking at the two occupation groups separately (equations (14) and (15)), it is found that the major culprit is clerical and sales, for, in equation (14) where managerial and professional is the right hand argument, HOBS remains positive with a t-statistic greater than one, although the size of the coefficient has diminished relative to equation (12).

/...

Two conclusions emerge from the above analysis. First, the white collar groups, and especially clerical sales, are probably more industry specific than was originally thought to be the case. Second, the economic interpretation of a wage determination equation becomes clouded with the inclusion of white collar and other occupation variables. For example, it becomes much more difficult to discover the effects on average earnings of a given shift in industrial structure. It should be emphasized, however, that this problem only arises in the context of simple equation estimation. With the development of a full model, the white collar groups could be related in turn to the importance of certain industrial activities and thus earnings could be related indirectly to the industries in the white collar equations.^{3/} At this stage though, the decision was made to ignore the occupational dimension in further single equation estimation.

TABLE E-3

DEPENDENT VARIABLE: AVERAGE ANNUAL INCOME - \$

INDEPENDENT VARIABLES: INDUSTRY SHARES AND LABOUR FORCE UTILIZATION VARIABLES

EQ. NO.	CONSTANT	AG	FOP	FISH	LVAV	PP	HORS	EH	RS	PA	PR	UR	DER	EADP
(17)	3,008.3 (625.3) (4.81)	-29.19 (5.60) (-5.21)	-28.52 (8.95) (-3.19)	-27.45 (6.28) (-4.37)	-28.91 (7.94) (-3.64)	12.55 (8.95) (1.40)	23.98 (18.37) (1.31)	-21.68 (14.75) (-1.47)	-20.64 (27.94) (-1.65)	3.88 (6.00) (0.65)	8.06 (13.24) (0.61)			
(18)	3,392.4 (418.8) (8.10)	-30.14 (5.22) (-5.77)	-24.43 (9.35) (-2.51)	-26.82 (6.16) (-4.35)	-31.96 (8.15) (-3.92)	14.28 (7.32) (1.95)	24.80 (16.31) (1.34)	-17.68 (14.32) (-1.23)	-22.98 (18.40) (-1.25)	4.50 (5.04) (0.89)		-21.96 (16.05) (-1.37)		
(19)	3,488.5 (834.8) (4.18)	-27.45 (5.07) (-5.41)	-29.51 (8.91) (-3.31)	-27.07 (6.43) (-4.21)	-28.12 (7.86) (-3.58)	15.66 (7.44) (2.10)	28.86 (16.10) (1.79)	-19.50 (14.71) (-1.33)	-31.61 (17.71) (-1.79)	5.66 (5.12) (1.11)		-4.19 (15.22) (-0.28)		
(20)	2,889.1 (574.5) (5.03)	-30.50 (5.69) (-5.36)	-26.59 (9.16) (-2.91)	-27.47 (6.22) (-4.42)	-30.37 (8.06) (-3.77)	11.05 (8.61) (1.28)	19.34 (18.60) (1.04)	-21.87 (14.49) (-1.51)	-26.77 (18.12) (-1.48)	2.64 (5.92) (0.45)				12.15 (11.97) (1.01)
F-STAT (10,35) = 27.85 SEE = 196.12 R ² = .888 CV = 8.69%														
F-STAT (10,35) = 27.14 SEE = 198.39 R ² = .886 CV = 8.79%														

TABLE E-3 (CONT'D)

INDEPENDENT VARIABLE: AVERAGE ANNUAL EARNINGS (\$)

INDEPENDENT VARIABLES: INDUSTRY SHARES AND LABOUR FORCE UTILIZATION VARIABLES

EQ. NO.	CONSTANT	AG	FOF	FISH	LVAM	PP	HOBS	EH	PS	PA	SE	HO
(21)	3,777.6 (389.9) (9.69)	-20.04 (4.84) (-4.14)	-8.05 (9.72) (-0.83)	-18.81 (5.96) (-3.15)	-28.16 (6.67) (-4.22)	13.98 (6.43) (2.18)	30.65 (13.41) (2.29)	-24.60 (12.62) (-1.95)	-39.55 (15.43) (-2.56)	1.87 (4.52) (0.41)	-24.45 (6.99) (-3.50)	
	F-STAT (10,35) = 36.66 SEE = 173.28 $R^2 = .913$ $\bar{R}^2 = .888$ CV = 7.68%											
(22)	3,876.0 (384.8) (10.07)	-24.45 (4.28) (-5.71)	-38.21 (7.47) (-5.12)	-15.70 (6.08) (-2.58)	-26.67 (6.49) (-4.11)	18.91 (6.29) (3.01)	31.08 (13.02) (2.39)	-36.32 (12.90) (-2.81)	-21.81 (15.04) (-1.45)	2.26 (4.43) (5.06)		-3.21 (2.18) (-3.68)
	F-STAT (10,35) = 39.03 SEE = 168.38 $R^2 = .918$ $\bar{R}^2 = .894$ CV = 7.46%											
(23)	4,190.82 (353.25) (11.86)	-18.66 (14.17) (-4.47)	-18.82 (8.85) (-2.13)	-10.34 (5.62) (-1.84)	-27.17 (5.73) (-4.74)	17.09 (5.59) (3.06)	31.46 (11.51) (2.73)	-37.66 (11.41) (-3.30)	-29.76 (13.51) (-2.20)	-2.20 (4.23) (-0.55)	-20.12 (6.12) (-3.23)	-66.68 (18.18) (-3.67)
	F-STAT (11,34) = 46.42 SEE = 148.80 $R^2 = .938$ $\bar{R}^2 = .917$ CV = 6.59%											

- 24 -

Next, in line with the regression strategy discussed in the Methodology section, different labour force utilization measures were added separately to the "best" industry equation, equation (12). These results are outlined in Table E-3. Turning first to the original X variables, all but one of the industry coefficients performed fairly well during these experiments, the one exception being public administration and defense which went from positive in equation (12) to finally negative -- although with a t-statistic well below one -- in equation (23) in Table E-3.

Equations (17) through (20) focus on the more familiar labour force utilization measures. The participation and unemployment rates, and the dependency ratio $DER^{4/}$ all entered the regression with their anticipated signs, but none of the estimated coefficients were statistically significant and only one t-statistic -- that for $UR^{5/}$ -- has an absolute value greater than one.^{5/} In equation (20), employment as a proportion of the labour force source population E/WP -- which thus includes the effects of both the unemployment and participation rates -- was added to the regression; its sign was positive, as anticipated, but the student's (t) is barely one.

The critical labour force utilization variables turned out to be seasonality and hours (equations 21-23) as might be expected from the simple correlation results. As indicated there, the importance of these variables reflects the fact that our earnings measure is an annual figure. Accordingly, their inclusion in the regression "corrects" for differences between industries in the degree of seasonality and average hours worked per week. The industry coefficients thus become more clearly a function of skill levels, capitalization and the other variables discussed in the previous sections.

In analyzing equations (21) and (22), it was apparent that the two variables had differing impacts on a given industry, (for example, compare FOR in equation (21) with the same coefficient in (22)). The suggestion was therefore that a preferable specification would be to include both variables in an equation. This was done in equation (23) where SE and HO both enter with highly negative signs and with t-statistics well over three (in absolute value). Following from this, the adjusted R^2 moves up sharply from .854 in equation (12) to .917 in equation (23). One final point should be made regarding equation (23). The equation suggests that, once seasonality and hours are accounted for, the primary activities cease to be the lowest paying industries in the Atlantic. The suggestion is therefore that

TABLE E-4

DEPENDENT VARIABLE: AVERAGE ANNUAL EARNINGS \$

INDEPENDENT VARIABLES: INDUSTRY SHARES AND OTHER VARIABLES

EQ. NO.	CONSTANT	AG	FOR	FISH	LVAM	PP	HORS	EH	RS	PA	UPB	LED	NW VAFM 1/100	SE
(24)	3,180.2 (424.4) (7.49)	-24.64 (5.52) (-4.46)	-27.37 (8.72) (-3.14)	-26.14 (6.27) (-4.17)	-25.49 (7.82) (-3.26)	14.23 (7.39) (1.93)	25.88 (15.64) (1.66)	-20.93 (14.31) (-1.46)	-31.20 (17.37) (-1.80)	5.92 (4.17) (1.19)	2.48 (2.06) (1.21)			
	P-STAT (10,35) = 27.50	SEE = 197.24	R ² = .887	R ² = .855	CV = 8.74%									
(25)	3,302.1 (405.3) (8.15)	-28.28 (4.82) (-5.86)	-25.28 (8.67) (-2.92)	-25.13 (6.18) (-4.06)	-28.80 (7.45) (-3.87)	14.78 (7.15) (2.07)	21.05 (15.76) (1.34)	-28.27 (14.73) (-1.92)	-41.25 (17.82) (-2.31)	1.73 (5.39) (0.32)		15.01 (8.49) (1.77)		
	P-STAT (10,35) = 28.92	SEE = 192.86	P ² = .892	R ² = .861	CV = 8.55%									
(26)	3,473.4 (364.5) (9.53)	-32.10 (4.46) (-7.20)	-33.54 (7.38) (-4.54)	(-33.00) (5.62) (-5.89)	(-24.81) (6.66) (-3.72)	0.31 (7.64) (0.04)	13.41 (14.26) (0.95)	-34.11 (13.04) (-2.62)	-28.39 (15.15) (-2.31)	3.34 (4.33) (0.76)			7.44 (2.76) (3.61)	
	P-STAT (10,35) = 37.36	SEE = 171.99	R ² = .914	R ² = .890	CV = 7.61%									
(27)	3,223.5 (484.7) (6.65)	-27.49 (5.04) (-5.45)	-29.65 (8.77) (-3.38)	-27.34 (6.32) (-4.33)	-27.54 (7.77) (-3.54)	14.98 (7.77) (1.93)	29.14 (15.85) (1.84)	-20.14 (14.57) (-1.38)	-31.42 (18.21) (-1.67)	4.87 (6.15) (0.79)				3.60 (2.20) (0.26)

TABLE E-4 (CONT'D)

<u>DEPENDENT VARIABLE:</u>	<u>AVERAGE ANNUAL EARNINGS (\$)</u>
<u>INDEPENDENT VARIABLES: INDUSTRY SHARES AND OTHER VARIABLES</u>	

[illegible]

One final point should be made regarding equation (23). The equation suggests that, once seasonality and hours are accounted for, the primary activities cease to be the lowest paying industries in the Atlantic. The suggestion is therefore that annual earnings in these industries could be increased significantly if means could be found to increase the degree of manpower utilization in these primary activities. As will be detailed later, equation (23) becomes the final specification for the wage determination equation, once public administration was dropped from the regression.

Table E-4 outlines the results from adding other explanatory variables individually to equation (12). All seven variables possess their anticipated signs while five of the seven have t-statistics with an absolute value greater than one. It should be noted however that, in some cases, the coefficient may simply reflect the correlation of the variable with the two critical labour force utilization measures discussed earlier, seasonality and hours. This issue will be raised later when final specifications are outlined. Special note should be made of two of the new variables, the capital proxy variable or non-wage value

- 29 -

added per employee in manufacturing, ⁵a/ NWVAEM and the Newfoundland dummy variable which has a value of one for the ten Newfoundland Census Divisions and zero for all others.

NWVAEM entered the regression with a highly significant positive coefficient and brought about a fairly sharp upward movement in the $\bar{R}^2$, from .854 in (12) to .890 in (26). At the same time, it had a strong impact on the positive coefficients in the regression and particularly the pulp and paper variable whose coefficient became negligible in size. Similar to the inclusion of the occupation variables, the interpretation of the industry coefficients become less straight forward with the capital per worker proxy included in the regression.

Comparing equations (12) and (26) it can be argued that the high coefficient on pulp and paper was largely a reflection of the high capital per worker ratio in that industry and thus once the latter is included pulp and paper ceases to be significant. This suggests in turn that relative skill levels and the high level of unionization have little impact on earnings in this industry. Such conclusions, however, must be made with care for, among other things, this measure is probably too industry specific to be used as an overall capital proxy. Accordingly the impact of this

1.5.

variable on other goods producing industries is relatively slight. Because of these difficulties, the decision was made to ignore this capital proxy in further analysis.

It might be useful though to place the results using a capital proxy in the context of the earlier discussion on the "long term" competitive hypothesis. To the extent that longer term forces are predominant, the findings in equation (26) are inconsistent with the competitive hypothesis, for the equation suggests that higher wages in pulp and paper are a function only of the high capital per worker ratio in that industry (and/or in the areas where that industry is located). However, it cannot be said with certainty at any point in time whether longer or shorter term factors are the crucial determinants and thus the implications of equation (26) for the competitive hypothesis are far from clear. This points out the difficulty of deriving "Reder-like" conclusions from cross-section analysis.

Second, the analysis of residuals from equation (12) suggested that, among the provincial dummies that were available, the one for Newfoundland might make an important contribution to the regression. As was anticipated, NFD did enter the regression with a positive and significant coefficient (equation (30)) and resulted in some improvement in the overall fit, with R^2 going from .854 to .868. This variable was not included in further single equation estimation but was returned to later in experiments using 2SLS.^{6/}

/...

TABLE E-5
DEPENDENT VARIABLE: AVERAGE ANNUAL EARNINGS (\$) $R^2 = 0.958$
FINAL SPECIFICATIONS

EQ. NO.	CONSTANT	PRIM	LVAM	PP	HOBS	EH	RS	PA	SE	HO	LED	UR	URS
(31)	4,061.9 (396.6) (10.24)	-16.23 (4.86) (-3.34)	-23.91 (6.32) (-3.78)	15.25 (5.75) (2.65)	24.71 (14.61) (1.69)	-38.10 (12.85) (-2.96)	-38.00 (17.00) (-2.24)	-2.41 (4.52) (-0.53)	-19.92 (7.69) (-2.59)	-50.70 (16.20) (-3.13)	5.23 (7.89) (0.66)	7.74 (14.38) (0.54)	0.75 (1.74) (0.43)
	F-STAT (12, 33) = 38.10 SEE = 156.85 $R^2 = .933$ $R^2 = .908$ CV = 6.958												
(32)	3,988.8 (222.0) (17.97)	-16.84 (3.94) (-4.27)	-24.27 (5.12) (-4.74)	16.46 (5.15) (3.24)	24.82 (11.76) (2.11)	-32.90 (10.65) (-3.09)	-28.54 (11.65) (-2.45)	-18.41 (6.18) (-2.98)	-51.54 (14.78) (-3.49)	1.39 (6.54) (0.21)			
(33)	4,030.98 (164.11) (21.56)	-15.80 (4.38) (-2.66)	-23.43 (5.38) (-4.35)	16.47 (5.12) (3.22)	29.53 (12.78) (2.31)	-33.93 (10.81) (-3.14)	-34.46 (13.86) (-2.34)	-21.58 (6.81) (-3.17)	-50.81 (14.80) (-3.43)			6.64 (12.40) (0.53)	
(34)	4,019.0 (25.39)	-17.81 (3.85) (-4.63)	-17.57 (8.46) (-2.08)	9.68 (5.43) (-1.78)	-25.71 (5.02) (-5.12)	18.02 (5.26) (3.43)	32.26 (11.31) (2.85)	-35.12 (10.31) (-3.41)	-26.33 (11.84) (-2.22)	-19.50 (5.95) (-3.28)	-63.96 (17.30) (-3.70)		
	F-STAT (10, 35) = 52.08 SEE = 147.30 $R^2 = .937$ $R^2 = .919$ CV = 6.538												
	CONSTANT	AG	FOR	FISH	LVAM	PP	HOBS	EH	RS	SE	HO		

Table E-5 outlines the final specifications that were attempted. Among the variables added individually in Tables E-2 through E-4, five held particular promise, hours, seasonality, the level of education, the unemployment rate and urbanization. It was decided as an initial step to include all five in a single regression (see equation (31)). Further, to save on degrees of freedom, the three primary industries were added together and included as a single variable.^{7/} Among the five new variables, only two, SE and HQ, both possessed their anticipated signs and had t-values greater than one, while, in one case, UR, a switch in sign occurred. However, the author does find it noteworthy that LED and URB have their expected signs and possess coefficients whose magnitudes at least seem supportable. The possibility is then that if more degrees of freedom were available these variables could play a larger role in a wage determination equation. A further consideration is that URB, which is based on the Census definition, proves too broad a measure of urbanization for the present exercise; a better variable may be the proportion of the population living in centers of 10,000 or more, which would better capture the effects of larger urban centers on earnings. Such a variable should be investigated in further research in this area.^{8/}

In the end, it was decided to leave LED and URB out of the final specification, particularly after LED performed badly when added separately (equation (32)). Turning briefly to the industry variables, all but public administration held up very well. Accordingly, equation (34) became the final wage determination equation; its specification involved dropping PA from equation (12) and adding seasonality and hours. This became the pivotal equation in the later development of the model.

The author must admit to some disappointment with this specification. Seasonality and hours are for the most part only "compositional" or "structural" variables, while none of the behavioural hypotheses concerning the levels of education, labour force utilization, urbanization and so on added to the explanation of intercounty differences in earnings in the Atlantic Region. At the same time, this specification points out the crucial role played by industry in the wage determination process in the Atlantic Region. This of course is supportive of the author's main hypothesis outlined at the end of the literature section in the text. Moreover, the Farrar-Glauber equations suggest that multicollinearity is playing little or no role in equation (34). The F-statistics were highest for SE, FISH, HOBS and FOR, but the largest came to only 15.34. The main weakness of the

Farrar-Glauber equations is that they only point out the variable that may be causing the problem, and not whether the multicollinearity is harmful. Still, it would be surprising if F-Statistics of this magnitude were indicative of harmful multicollinearity. Finally, no analyst can be too disappointed with a cross-section equation that explains 92 per cent of the variability in the dependent variable. All in all then, equation (34) represented a useful -- although not totally satisfactory -- starting point for the development of the model.

This sub-section will be closed out with a few "odds and ends". One estimation was attempted using employment income per capita as the dependent variable and the nine industries in equation (12) as the right hand arguments. The results were:

(35) EIC =	534.2 (181.7) (2.94)	-5.84 AG (2.16) (-2.71)	-6.98 FOR (3.68) (-1.89)	-6.37 FISH (2.71) (-2.35)
		+0.03 LVAM (3.33) (0.01)	+7.37 PP (3.20) (2.30)	+30.89 HOBS (6.69) (4.62)
		-8.33 EH (6.27) (-1.33)	-3.07 RS (7.61) (-0.40)	+2.41 PA (2.18) (1.11)

F-STAT (9,36) = 19.75 SEE = 86.47 $R^2 = .832$

$\bar{R}^2 = .789$ CV = 12.968

The most prominent divergence between this equation and equation (12) is that here LVAM and RS cease to be significantly negative. This likely reflects the relatively high levels of labour force utilization associated with these two industries. In other respects though the results are remarkably similar, although the adjusted R^2 is somewhat lower. This is not surprising since certain key variables are left out, such as the unemployment and participation rates -- as perhaps summarized in the employment ratio -- and the degree of seasonality and the hours variable. These variables were included in a further regression; after some respecification, the results were as follows:

(36)	$\hat{E}IC =$	395.75 (100.33) (3.94)	-8.59 AG (1.44) (5.95)	-1.60 FISH (1.64) (-0.98)	-6.75 LVAM (1.94) (-3.48)
		+3.42 PP (1.81) (1.89)	+13.12 HOBS (3.77) (3.47)	-14.23 EH (3.55) (-4.01)	-4.83 PA (1.31) (-3.69)
		+25.68 ER (3.77) (6.82)	-3.97 SE (1.68) (-2.36)	-22.95 HO (5.68) (-4.04)	

F-STAT (10, 35) = 70.99 SEE = 46.32 $R^2 = .953$

$\bar{R}^2 = .940$ CV = 6.94%

- 36 -

As anticipated, the adjusted R^2 rose markedly from .79 to .94 while the coefficients on the three manpower utilization measures are all highly significant. Important changes also occurred to some of the β coefficients. Forestry dropped completely out of the equation and the coefficient on FISH dropped considerably, perhaps surprising results in light of the final wage determination equation. Also, LVAM became negative and highly significant, while the sign on PA switched from negative to positive. Consistent with equation (35), the coefficient on RS remained insignificant. Finally, comparing equation (36) with the final wage determination equation (equation (34)), the adjusted R^2 in (36) is even higher -- .94 compared with .92 -- while important differences exist with respect to some of the industry coefficients, particularly PA, RS, FOR and FISH.^{9/}

Clearly then, results can differ depending on the left hand variable used.^{10/} On balance, the expectation would be that average earnings might provide the more "reliable" results, since it is taken directly from the Census whereas employment income per capita had to be estimated. Still, these differences point out the need for caution in interpreting the regression results, and more generally the need for a sounder data base for this type of work. (Hopefully, the latter will be provided by the 1971 Census.)

/...

- 37 -

Equation (36) was then re-estimated using a different left hand variable, namely, average income per non-farm family.

$$\begin{array}{lll}
 (37) \quad \hat{INFF} = 2,933.2 & -10.67 \text{ AG} & +7.50 \text{ FISH} \\
 & (484.4) & (7.92) \\
 & (6.06) & (0.95) \\
 & -30.49 \text{ LVAM} & +34.34 \text{ PP} & +56.71 \text{ HOBS} \\
 & (9.37) & (8.72) & (18.22) \\
 & (-3.26) & (3.94) & (3.11) \\
 & -28.95 \text{ EH} & -11.09 \text{ PA} & +83.42 \text{ ER} \\
 & (17.16) & (6.33) & (18.19) \\
 & (-1.69) & (-1.75) & (4.59) \\
 & -14.60 \text{ SE} & -156.34 \text{ HQ} & \\
 & (8.13) & (27.45) & \\
 & (-1.80) & (-5.70) &
 \end{array}$$

$$F\text{-STAT} (10, 35) = 46.23 \quad SEE = 223.61 \quad R^2 = .930$$

$$\bar{R}^2 = .910 \quad CV = 6.07\%$$

Compared with equation (36), the only major change involves the change in sign from negative to positive for FISH, which was the weakest variable in the previous equation. Beyond that, the differences in the magnitudes of the coefficients are explained for the most part by the larger size of the INFF variable. Its mean comes to \$3,682 compared with \$667 for employment income per capita. The one possible exception to this is the coefficient on AG which is relatively much smaller in equation (37) than equation (36). This likely reflects the

/...

fact that the INFP series does not include farm incomes.^{11/}

This may also underlie the somewhat lower adjusted R^2 in equation (37).^{12/}

Next, attempts were made to estimate an earnings equation based on occupation alone and then occupation and labour force utilization measures. The final equations were as follows:

$$\begin{array}{rcll}
 (28) \quad \hat{W} = & 2,522.0 & -83.81 \text{ MA} & +56.03 \text{ CLE} \\
 & (530.8) & (31.30) & (15.21) \\
 & (4.75) & (-2.68) & (3.68) \\
 & +9.60 \text{ SR} & +38.88 \text{ TRO} & -20.46 \text{ FAFW} & -23.80 \text{ LF} \\
 & (6.98) & (30.84) & (6.65) & (6.64) \\
 & (1.37) & (1.26) & (3.08) & (-3.57)
 \end{array}$$

$$F\text{-STAT } (6,39) = 32.97 \quad SEE = 225.65 \quad R^2 = .835$$

$$\bar{R}^2 = .810 \quad CV = 10.00\%$$

$$\begin{array}{rcll}
 (39) \quad \hat{W} = & 3,724.51 & -81.17 \text{ MA} & +39.22 \text{ CLE} & -2.25 \text{ SR} \\
 & (591.44) & (31.21) & (14.14) & (7.16) \\
 & (6.30) & (-2.60) & (2.77) & (-0.31)
 \end{array}$$

$$\begin{array}{rcl}
 +11.18 \text{ TRO} & -17.49 \text{ FAFW} & -14.12 \text{ LF} \\
 (28.40) & (5.91) & (6.52) \\
 (0.39) & (-2.96) & (-2.17)
 \end{array}$$

$$\begin{array}{rcl}
 -20.40 \text{ SE} & & -53.94 \text{ HO} \\
 (7.71) & & (19.17) \\
 (-2.65) & & (-2.81)
 \end{array}$$

$$F\text{-STAT } (8,37) = 33.73 \quad SEE = 198.26 \quad R^2 = .879$$

$$\bar{R}^2 = .853 \quad CV = 8.78\%$$

- 39 -

Experiments were attempted with all available occupation categories before a reasonable specification was acquired. The six occupations included in equation (38) were the only ones that held up under repeated testing. Professional, sales, miners, and craftsmen and production workers -- all with positive expected signs based on the simple correlation results -- were among the categories dropped from the regression. The signs and magnitudes of the coefficients in equation (38) are in general accord with the corresponding industry results (particularly equation (12)). The one surprise is perhaps the negative coefficient on the managerial category. The factors underlying this are discussed in Appendix C.^{13/}

Similar to the industry analysis, the adjusted R^2 improved fairly considerably with the inclusion of the two manpower utilization variables, seasonality and hours. However, two occupation variables became insignificant. The demise of the service and related category, which includes government workers, corresponds to the dropping of public administration for the final industry equation once SE and HO were included. Lack of time prevented further specification of the occupation equation. Finally, it may be noted that the explanatory power of equations (38) and (39) is below that of the corresponding industry equations, equations (12) and (34).

This agrees with the standardization results in Appendix D, and adds further weight to our decision to work with the industrial dimension.

The final cross-sectional work involved estimating regressions for 1951 and 1966 using the more limited data available for those two years. In each case, comparisons are made with a similar equation estimated with 1961 data. The dependent variable for 1951 was median earnings, with the 1951 and 1961 regressions being as follows:

$$\begin{aligned}
 (40) \quad \hat{MEW} (51) = & 2,134.53 & -15.91 \text{ AG} (51) & -3.00 \text{ FISH} (51) \\
 & (149.59) & (2.92) & (3.57) \\
 & (14.27) & (-5.44) & (-3.64) \\
 & +7.37 \text{ MIN} (51) & -5.61 \text{ MAN} (51) & -8.71 \text{ TTRANS} (51) \\
 & (3.79) & (3.33) & (7.35) \\
 & (1.95) & (-1.69) & (-1.18) \\
 & +1.40 \text{ URB} (51) & -20.86 \text{ SE} (51) & \\
 & (1.36) & (3.40) & \\
 & (1.03) & (-6.15) & \\
 F\text{-STAT} (7,38) = & 40.70 & SEE = 133.69 & R^2 = .882 \\
 \bar{R}^2 = & .861 & CV = 11.22\%
 \end{aligned}$$

$$\begin{aligned}
 (41) \quad \hat{MEW} = & 2,923.54 & -22.77 \text{ AG} & -24.52 \text{ FISH} & +8.99 \text{ MIN} \\
 & (265.41) & (5.80) & (6.20) & (9.15) \\
 & (11.02) & (-3.92) & (-3.95) & (0.98) \\
 & -17.27 \text{ MAN} & -3.55 \text{ TTRANS} & +4.75 \text{ URB} & -29.09 \text{ SE} \\
 & (5.40) & (10.87) & (2.26) & (6.19) \\
 & (-3.20) & (-0.33) & (2.10) & (-4.70) \\
 F\text{-STAT} (7,38) = & 30.29 & SEE = 240.37 & R^2 = .848 \\
 \bar{R}^2 = & .820 & CV = 11.36\%
 \end{aligned}$$

The methodology applied here involved finding the best equation for 1951 (equation (40)) and then re-estimating the same specification for 1961 (equation (41)). Therefore, the latter represents a mis-specification^{14/} but it does provide a useful benchmark for comparison. Turning first to equation (40), one industry notable by its absence is forestry which entered the regression with a small positive coefficient but with a t-statistic below one. It seems then that as of 1951 forestry was not a low wage activity in the Atlantic Region. Also note should be made of the urbanization and seasonality variables. URB (51) -- with a positive coefficient -- is probably acting on behalf of higher order business services and other high wage urban oriented activities.^{15/} The coefficient on seasonality reflects in part the impact of hours; (no hours variable was offered by the 1951 Census).

A few points should be noted in comparing the 1951 and 1961 equations. First, the coefficient on mining, while remaining positive, became insignificant in 1961; this probably reflects the growing difficulties encountered by that industry, particularly coal mining, over the decade. Second, the t-statistic on TRANS fell from slightly above to well below one (in absolute value) between 1951 and 1961. Third, the positive coefficient on urbanization became significant in 1961, perhaps reflecting the growing role of HOBS and other services in the

/...

Atlantic Region over the decade. In sum, the comparison suggests that, between 1951 and 1961, there was both some change and some continuity in the association between earnings and other economic variables in the Atlantic Region. The relationships are far from static, but it appears that explanations can fairly readily be found for those changes that do occur.^{16/}

Turning to the 1966 experiments, only two explanatory variables are available for that year, per capita value added in manufacturing and retail sales per capita. The 1966 and 1961 regressions -- using employment income per capita as the right hand variable -- are as follows:

$$(42) \quad \begin{array}{lll} \text{EIC (66)} = & 304.93 & +0.019 \text{ VACM (66)} & +0.648 \text{ RSC (66)} \\ & (125.52) & (0.186) & (0.144) \\ & (2.42) & (0.10) & (4.50) \end{array}$$

$$F\text{-STAT (2,43)} = 13.70 \quad SEE = 243.92 \quad R^2 = .389$$

$$\bar{R}^2 = .361 \quad CV = 26.33\%$$

$$(43) \quad \begin{array}{lll} \text{EIC} = & 311.62 & +0.235 \text{ VACM} & +0.452 \text{ RSC} \\ & (61.87) & (0.130) & (0.097) \\ & (5.04) & (1.80) & (4.66) \end{array}$$

$$F\text{-STAT (2,43)} = 22.98 \quad SEE = 134.02 \quad R^2 = .517$$

$$\bar{R}^2 = .494 \quad CV = 20.09\%$$

- 43 -

The two equations indicate that per capita value added and retail sales per capita -- acting as proxies for a wide range of other variables -- explain relatively little of the intercounty variation in income per capita. Moreover, the equation's explanatory power declines fairly markedly between 1961 and 1966, with VACM ceasing to be significant in the latter year.

Finally, efforts were made to explain the growth in median earnings between 1951 and 1961, using shifts in industry shares and changes in other variables as the independent variables. The final specification was:

(44) GMEW = 68.42	-2.87	-2.34 SFOR
(4.92)	(0.49)	(0.63)
(13.90)	(-5.87)	(-3.70)
-2.71 SFISH	+1.80 SCON	-2.46 STTRANS
(0.56)	(1.11)	(0.84)
(-4.86)	(1.62)	(-2.91)
-2.22 STQTR	-1.22 STOSE	+0.306 CURB
(0.87)	(0.39)	(0.238)
(-2.54)	(-3.13)	(1.28)
-3.31 CSE	+0.89 CER	
(0.33)	(0.90)	
(-10.08)	(0.99)	
F-STAT (10, 35) = 37.00	SEE = 9.59	R ² = .914
R ² = .889	CV = 11.71%	

Given the data limitations, this is a fairly successful specification, with the equation explaining nearly 90 per cent of the intercounty variation in the growth in earnings.^{17/} Also, the signs on the coefficients seem defensible, and are generally in line with the cross-sectional results. The two exceptions to the latter statement are first the positive and significant coefficient on construction and second the fact that both service industries entered with a negative sign. The possibility exists though that the influence of higher order business services and other higher paying service activities is being picked up by the change in urbanization. Finally, note should be made of the positive coefficient on the change in the employment ratio. This is consistent with the hypothesis that the growth in earnings tends to be higher the greater the relative increase^{18/} in the demand for labour; for this reason, CER was retained in the regression despite a low t-value. To sum up, these results indicate a rough correspondence between our cross-sectional and time series findings. This agrees in turn with the conclusions of Chenery and Perloff noted in the literature chapter of the text.^{19/}

- 45 -

With the development of the model, these "odds and ends" came to play a much smaller role in the present exercise, and much of this work was left unfinished. Still, it is hoped that the above equations provide some further insights into the wage and income determination process in the Atlantic Region. Moreover, this work is suggestive of areas for further research. For example, the growth in earnings equation could be used as the base for a small model explaining changes over time in the Atlantic Region.

/...

FOOTNOTES

APPENDIX E:

1/ No F-Statistics were available from the first regressions that were run, since these used the Massager program.

2/ As will be outlined later, two non-industry variables had to be added to equation (12) before it was properly specified. This relates to the fact that wages here are an annual average.

3/ Unfortunately, when the two occupational categories were included in the wage determination equation estimated by Two Stage Least Squares the two categories failed to add to the adjusted R^2 . Still, the issue raised here is perhaps worthy of further investigation if a 1971 model is estimated.

4/ The dependency ratio is defined here as the percentage of the total population in the 0-14 and 65 and over age groups. It is used in this context to analyze the impact of the age structure on average earnings.

5/ In equations (17) through (34), the diagnostic statistics, i.e. F-Statistic, R^2 , etc., are included in the Table only if the additional variable(s) have a t-statistic greater than one and thus add to the adjusted R^2 .

6 a/ For the most part, this measure focusses on profits per employee.

6/ Although the Newfoundland dummy variable did not end up in the final 2SLS specification.

- ^{7/} For those familiar with I.P. Sharp's computer facilities you will know that this was partially dictated by the famous "work-space full" problem. The author found that 12 was the maximum number of X variables that could be included in an equation before running out of his allotment of computer work space. In the present instance, however, it probably also made econometric sense to reduce the number of X variables in the regression. Using 14 right hand variables with only 46 observations does seem a little extreme.
- ^{8/} As indicated in the text, some experiments were attempted with this alternative urbanization variable but with no success. Still, this question should be returned to in estimating a 1971 model.
- ^{9/} Despite the fact that equation (36) has been "standardized" in terms of equation (34) by the inclusion of the employment ratio in the latter. (See Appendix A, page 5.)
- ^{10/} This relates back to the statement made in the Introduction concerning possible limitations in the available earnings and income measures. The differences between the two equations also relate to the difficulties encountered in building the model where it was found that an "adjustment" equation -- rather than an identity -- had to be used in associating average earnings with employment income per capita.
- ^{11/} All of our analysis would suggest that the inclusion of farm incomes would have a downward effect on this series, with the impact being larger in the areas where agriculture is more important. Accordingly, equation (37) probably under-states the downward effect of agriculture on incomes. In this equation only the indirect effects of farming are picked up in the regression.
- ^{12/} An equation was also estimated using income per non-farm family as the dependent variable and the same industry variables as in equations (12) and (35). There were three major divergences from equation (35): (1) the coefficient

on LVAM was negative with a t-statistic slightly over one; (2) the coefficient on EH was negligible; (3) the coefficient on RS was negative and significant at the 95 per cent confidence level using a one-tail test. The adjusted R^2 was comparable, amounting to .785.

^{13/} See Appendix C, page 8.

^{14/} Therefore, it is not comparable with previous wage determination equations estimated for 1961. Among other considerations, equation (41) is based on the less detailed industry categories available for 1951. Beyond that, the equation excludes certain industries such as forestry which previous 1961 analyses indicated should be included.

^{15/} Accordingly, in deriving the final 1961 wage determination equation, URB had a positive coefficient but with a t-statistic well below one, reflecting probably the inclusion of HOBS in the same equation. See page 31 of the present Appendix.

^{16/} This provides further arguments for developing the model for 1971 and then comparing the results with the 1961 model. This should not only provide further information regarding development in the Atlantic Region over the previous decade but should also provide some insights regarding the future course of the Atlantic economy. These issues are noted in the concluding chapter of the text.

^{17/} Focussing on a single industry coefficient the coefficient on SFISH indicates that, on average, the increase in earnings rises by 2.7 percentage points for every percentage point decline in fishing's share of employment between 1951 and 1961.

^{18/} That is, relative to population.

^{19/} See page 15 of the literature section of the text.

APPENDIX F:

PRINCIPAL COMPONENTS

APPENDIX F:

PRINCIPAL COMPONENTS

(a) METHODOLOGY

The following outlines in greater detail how principal components are computed. Their derivation involves three steps: First, the eigenvalues -- or characteristic roots -- are extracted from the matrix of explanatory variables, (with the observations being expressed as deviations from their sample means or in some other 'normalized' way, since the concern is with the variation in the data.) The number of eigenvalues equals the number of explanatory variables (k). The eigenvalues have the useful property that, when ranked in terms of size, summed and then divided into the total, the percentage for the largest eigenvalue indicates the proportion of the variability in the matrix of explanatory variables (i.e. the X matrix) accounted for by the first principal component (i.e. the component associated with the largest eigenvector), and similarly for the second, third and so on.

Second, the eigenvectors, which represent a set of weights, are calculated for each eigenvalue, with the squares of the weights in each eigenvector being normalized so as to add to one. Finally, the principal component -- which for the present exercise is made up of 46 observations, one for

each county -- is derived for each eigenvector and eigenvalue by multiplying the X matrix by the set of weights (a). In matrix form, the first principal component would be:

$$z_1 = Xa_1$$

while a single observation for a given county (j) would be:

$$z_{1j} = a_{11} x_{1j} + a_{21} x_{2j} + \dots + a_{k1} x_{kj}$$

Accordingly, the observation for the principal component is simply the weighted sum of the observations for that county in the X matrix. The (k) principal components of the X matrix can then be given by the nxk matrix Z, i.e:

$$Z = XA$$

To quote from Johnston: "If the rank of X were ($r < k$), ($k-r$) latent roots would be zero and the variation in the X's could be completely expressed in terms of (r) independent variables. Even if X has full column rank, some of the λ 's (ed. eigenvalues) may be fairly close to zero so that a small number of principal components account for a substantial proportion of the variance of the X's."^{1/} This holds out the possibility that a relatively small number of components could be used in the regression analysis to follow.

When the X variables are for example industry shares, the weights (a) indicate the importance of a given industry (i) in the composition of a principal component. The make-up -- and thus the possible economic interpretation --

of a given principal component can be further discerned by calculating pair-wise correlations between each of the principal components and each explanatory variable. A table of correlation coefficients is included with the results below. As indicated in the text, two sets of principal components were computed, one from the industry variables, and a second from the occupation variables.

(1) INDUSTRY PRINCIPAL COMPONENTS

The industries used in the industry principal components computation are outlined below:

AG: Agriculture
FR: Forestry
FISH: Fishing
LVAM: Lower Value Added Manufacturing
PP: Pulp and Paper
HOBSS: Higher Order Business Services
EH: Education and Health
RS: Residential Services
PA: Public Administration
MIN: Mining
OHVAM: Other High Value Added Manufacturing
CON: Construction
TRANS: Transportation and Storage
UTIL: Utilities (i.e. electricity, gas and water)

Table F-1 indicates the proportionate contribution of each principal component - as represented by its eigenvalue - to the total variation in the matrix of explanatory variables.

Further, Table F-2 outlines the industry weights that are used in the computation of each principal component, while Table F-3 indicates the pair-wise correlations between industries and principal components. In short, Table F-1 outlines the relative importance of each component while the last two are suggestive of the possible economic interpretation to be attached to each.

TABLE F-1

EIGENVALUES FOR INDUSTRY PRINCIPAL COMPONENTS

<u>PC.</u>	<u>EIGENVALUES</u>	<u>ASSOCIATED Z^2 AS % OF $X'X$</u>	<u>CUMULATIVE PERCENTAGE</u>
1.	7,473.52	37.72	37.72
2.	4,208.99	21.24	58.96
3.	2,898.93	14.63	73.59
4.	1,721.81	8.69	82.28
5.	970.64	4.90	87.18
6.	729.90	3.68	90.86
7.	635.60	3.21	94.07
8.	403.55	2.04	96.10
9.	347.84	1.76	97.86
10.	230.33	1.16	99.02
11.	109.03	0.55	99.57
12.	70.43	0.36	99.93
13.	13.67	0.07	99.996
14.	0.80	0.00	100.00

/...

TABLE F-2

EIGENVECTOR (WEIGHTS) FOR INDUSTRY PRINCIPAL COMPONENTS

		PRINCIPAL COMPONENT									
		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	AG	-.13	-.59	.69	-.22	.02	.00	.11	.05	.12	.03
2.	FOR	-.10	.18	-.08	-.61	.07	-.05	-.55	-.38	.05	.14
3.	FISH	-.27	.70	.37	.01	.02	-.29	.36	.07	.05	.02
4.	LVAM	-.21	.13	.20	.47	-.33	-.59	-.28	-.23	-.06	-.11
5.	PP	-.04	-.04	-.36	-.43	-.49	.32	.48	.17	.03	-.03
6.	HOBS	.00	-.23	-.20	.29	-.11	-.38	-.03	-.09	.05	-.16
7.	EH	-.05	-.10	-.10	-.01	.02	-.29	.06	-.22	-.34	-.67
8.	RS	-.08	-.12	-.21	.15	-.09	-.14	.02	-.15	.06	.32
9.	PA	.91	.15	.21	.01	-.15	-.01	-.03	-.04	.06	.00
10.	MIN	.09	.02	-.12	-.02	.74	.45	.19	.01	-.08	-.16
11.	OHVAM	-.01	-.11	-.20	.26	.20	-.08	.23	-.26	.27	.42
12.	CON	-.01	-.03	-.02	.06	.02	-.08	-.17	.45	-.72	.38
13.	TRANS	-.07	.03	-.14	.07	.07	-.04	-.36	.65	.50	-.22
14.	UTIL	.00	-.02	-.02	.01	.00	.03	.01	-.02	-.02	.02

TABLE F-3

CORRELATION COEFFICIENTS FOR INDUSTRY PRINCIPAL COMPONENTS

		PRINCIPAL COMPONENT									
		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	AG	-.21	-.69	.67	-.16	.01	.00	.05	.02	.04	.01
2.	FOR	-.28	.34	-.12	-.75	.06	-.04	-.41	-.23	.03	.06
3.	FISH	-.41	.81	.35	.01	.01	-.14	.16	.02	.02	.00
4.	LVAM	-.50	.23	.29	.53	-.29	.44	-.19	-.12	-.03	-.04
5.	PP	-.11	-.08	-.56	-.52	-.45	.25	.35	.10	.01	-.01
6.	HOBS	.00	-.58	-.42	.48	-.14	-.41	-.03	-.07	.04	-.09
7.	EH	-.24	-.36	-.31	-.03	.03	-.43	.08	-.25	-.35	-.56
8.	RS	-.33	-.40	-.57	.32	-.14	-.18	.02	-.15	.06	.24
9.	PA	.98	.12	.14	.01	-.06	.00	-.01	-.01	.01	.00
10.	MIN	.26	.05	-.23	-.03	.81	.42	.17	.00	-.05	-.08
11.	OHVAM	-.04	-.32	-.48	.48	.28	-.10	.26	-.24	.23	.29
12.	CON	-.03	-.12	-.06	.13	.03	-.11	-.24	.49	-.74	.32
13.	TRANS	-.28	.10	-.35	.13	.10	-.05	-.42	.61	.43	-.15
14.	UTIL	-.03	-.28	-.21	.06	.01	.18	.05	-.11	-.07	.07

- 7 -

As indicated in Tables F-2 and F-3, further analysis was restricted to the first ten principal components, for these exhausted 99% of the variability in the X matrix. Several conclusions emerge from this analysis. Compared with other principal component studies, the size of the first few components seem relatively small. Accordingly, ten of the fourteen components are required before the vast majority of the variation in the X's is accounted for. This gives the impression that the variation in industry shares is not concentrated but rather spread among a large number of industrial activities. This is in accord with the multiple regression results which indicated that up to nine industries had to be included in a regression before a "correct" specification was secured. In short, the suggestion is that in the current context, principal components will solve any multicollinearity problems that are present, but will not greatly reduce the number of right hand variables required in a regression.

Second, as the weights and particularly the correlations indicate, the most important principal components can be fairly well identified with one or a few industries. This is particularly true for the first principal component which is almost an exact replication of the public administration variable, as indicated by the coefficient of .98. This reflects in turn the large intercounty variation in the public administration series. The second component is closely identified with three low-wage

/...

activities, forestry, fishing and low value added manufacturing, and similarly for the third where the key industry is agriculture. The fourth component is associated with urban-oriented activities, including low value added and other high value added manufacturing, plus higher order business services, residential services and to a lesser extent construction and transportation. The fifth is identified with mining and to a much lesser extent other high value added manufacturing, reflecting the linkage between mining and primary metals manufacturing mainly in Cape Breton. Similar interpretations can be provided for most of the other principal components, although interpretation becomes somewhat more difficult with the less important ones.

Accordingly, the principal components can be used to focus on the types of industrial activities that tend to "cluster" together in a given area. Of course, in the first component, public administration stands on its own: (its "association" with mining probably being coincidental). This perhaps is not surprising for, particularly in the case of a military base, the forward and backward linkages emanating from this type of activity are almost non-existent. The second component shows the tendency for fishing, forestry and primary manufacturing activities to cluster together, while the third indicates a similar tendency

/...

for agriculture, fishing and primary manufacturing. These results are supported by previous studies which have indicated that in the Atlantic, the same worker might be a farmer, a fisherman and a fish plant employee all in the same year. Components two and three are then clearly rural in orientation.

In contrast, the urban orientation of component four has already been noted, reflecting the linkages between manufacturing and particular tertiary activities. Therefore, to use the jargon of the day, principal component four offers a crude measure of the effects of "agglomeration economies". Finally, component five reflects the forward linkage from mining to primary metals and other higher value added manufacturing.

On the other hand, some "clustering" is notable by its absence. For example, Table F-3 indicates no tendency for pulp and paper to be associated spatially with the forestry industry. This reflects the fact that much of the pulwood for the larger pulp and paper mills is secured from neighbouring counties. This is a reflection in turn of the small size of most counties and the fact that they are not really economic entities in themselves. This difficulty came up many times in formulating the model.

As the above suggests, few problems arose in interpreting the larger components. Thus, in this instance, principal component analysis proved most useful in showing, among other things, the inter-relationships between industrial activities. A final question

/...

- 10 -

whether the principal components are to be preferred over the industry variables themselves in regression analysis. Two wage determination equations were estimated using the components as the right hand arguments. The first used all ten while the second left out those components with t-statistics below one. The regression results are outlined below:

$$(1) \quad W = 2,256.94 + 28.26 \text{ PC } 1 - 12.33 \text{ PC } 2$$

$$\quad \quad (30.28) \quad (2.38) \quad (1.17)$$

$$\quad \quad (74.53) \quad (11.89) \quad (-3.89)$$

$$\quad \quad -34.08 \text{ PC } 3 + 8.16 \text{ PC } 4 - 1.49 \text{ PC } 5$$

$$\quad \quad (3.81) \quad (4.95) \quad (6.59)$$

$$\quad \quad (-8.93) \quad (1.65) \quad (-0.23)$$

$$\quad \quad -1.43 \text{ PC } 6 + 15.77 \text{ PC } 7 + 25.54 \text{ PC } 8$$

$$\quad \quad (7.60) \quad (8.15) \quad (10.22)$$

$$\quad \quad (0.19) \quad (1.94) \quad (2.50)$$

$$\quad \quad + 6.41 \text{ PC } 9 - 15.48 \text{ PC } 10$$

$$\quad \quad (11.02) \quad (13.53)$$

$$\quad \quad (0.58) \quad (-1.14)$$

$$F = \text{STAT } (10, 35) = 25.09 \quad \text{SEE} = 205.37 \quad R^2 = 0.878$$

$$R^2 = 0.843 \quad \text{CV} = 9.10\%$$

$$(2) \quad W = 2,256.94 + 28.26 \text{ PC } 1 - 12.33 \text{ PC } 2 + 34.08 \text{ PC } 3$$

$$\quad \quad (29.24) \quad (2.29) \quad (3.06) \quad (3.68)$$

$$\quad \quad (77.20) \quad (12.32) \quad (4.03) \quad (9.25)$$

$$\quad \quad + 8.16 \text{ PC } 4 + 15.77 \text{ PC } 7 + 25.54 \text{ PC } 8 - 15.48 \text{ PC } 10$$

$$\quad \quad (4.78) \quad (7.87) \quad (9.87) \quad (13.06)$$

$$\quad \quad (1.71) \quad (2.00) \quad (2.59) \quad (-1.18)$$

$$F = \text{STAT } (7, 38) = 38.39 \quad \text{SEE} = 198.29 \quad R^2 = 0.876$$

$$R^2 = 0.853 \quad \text{CV} = 8.79\%$$

The first point to be noted is that the adjusted R^2 in equation 2 above is exactly equal to that in the "best" industry equation in the single equation analysis (equation 12 in Appendix E). The evidence suggests therefore that, as a group, the principal components capture no more of the intercounty variability in earnings than the industry variables themselves. This further suggests that equation 12 is a "correct" specification for an industry equation in the sense that, using industry alone, we can probably go no further in explaining the variation in earnings.

Second, it may be noted in passing that the removal of the three components from the regression has no impact on the coefficients of the remaining components and only a slight impact on their standard errors. This of course reflects the fact that principal components are "orthogonal", that is, completely non-collinear.

Beyond this, the coefficients on most of the components probably accord with our expectations based on the industry equations. For example, the negative coefficients on components two and three, and the positive coefficient on four, is consistent with the analysis in the previous paragraphs. Similarly, the positive coefficient on component seven reflects the importance of pulp and paper in its composition, while the importance of

/...

- 12 -

transportation in component eight probably underlies its positive coefficient. (Transportation entered the industry equation with a positive sign, although it was not significant.)

However, the very strong positive coefficient on principal component one marks a sharp divergence from this pattern, for public administration was by far the weakest variable in the industry equation. This divergence is very difficult to explain. One possibility is that, in computing the principal components, the influence of industries that are positively associated with earnings has been dissipated by being mixed with lower paying activities. For example, principal component four represents an amalgam of higher paying activities, such as higher order business services, with such low-wage industries as low value added manufacturing and residentiary services. As a consequence, the coefficient on component four, while positive, is not very large and is barely significant at the 95% confidence level. In contrast, public administration was concentrated into a single component which in fact is no more than a transformation of the original series. Because of this, component one may be capturing the positive influences which were associated with higher order business services in the industry equations.

The author considered one further possibility, that in some manner component one may be a more powerful explanatory variable than the original public administration series.

/...

- 13 -

To test this hypothesis, principal component one was inserted in place of public administration in equation 12 of the previous Appendix A:

$$(3) \hat{W} = 3,283.6 \quad -26.58 \quad AG \quad -29.11 \quad FOR \quad -25.37 \quad FISH$$

$$(415.4) \quad (5.49) \quad (8.87) \quad (7.25)$$

$$(7.90) \quad (-4.84) \quad (-3.28) \quad (-3.50)$$

$$-26.07 \quad LVAM + 16.12 \quad PP + 30.76 \quad HOBS$$

$$(8.39) \quad (7.48) \quad (15.44)$$

$$(-3.20) \quad (2.15) \quad (1.99)$$

$$-19.81 \quad EH \quad -30.84 \quad PS + 6.62 \quad PC 1$$

$$(14.41) \quad (17.74) \quad (5.54)$$

$$(-1.37) \quad (-1.74) \quad (1.19)$$

$$F - STAT (9,36) = 30.07 \quad SEE = 198.33 \quad P^2 = .883 \quad P^2 = .853 \quad CV = 8.79\%$$

Compared with PA in equation 12, the coefficient on PC 1 is slightly higher, 6.62 vis-a-vis 5.86, but the t - value has hardly improved at all. Also, the remaining coefficients remain essentially unchanged and the $\bar{R}^2$ is the same. In short, the specification does not improve significantly with the inclusion of PC 1 over public administration.

The author's conclusion then is that, while it is relatively easy to identify a principal component with a set of industries, it is much more difficult to provide an economic interpretation for the regression coefficient on a component. This greatly inhibits their utility in regression analysis, particularly in the development of a model which may be used in

/...

a regional planning context. Therefore, principal components were used in this exercise only for illustrative purposes and not in regression estimation.

(c) OCCUPATION PRINCIPAL COMPONENTS

A set of principal components were also computed from the matrix of occupational variables. The twelve occupation categories provided by the Census are outlined below:

MA:	Managerial
PROF:	Professional
CLE:	Clerical
SA:	Sales
SR:	Service and Recreation
TR:	Transport and Related
FA:	Farmers
FAW:	Farm Workers
LF:	Loggers and Fishermen
MI:	Miners
CR:	Craftsmen and Related
LA:	Labourers

The following tables are similar to those in the previous sub-section.

- 15 -

TABLE F-4

EIGENVALUES

<u>PC.</u>	<u>EIGENVALUES</u>	<u>ASSOCIATED $\frac{\lambda^2}{\sum \lambda^2}$ AS % OF $X'X$</u>	<u>CUMULATIVE PERCENTAGE</u>
1.	6,337.96	49.65	49.65
2.	3,168.95	24.82	74.47
3.	1,933.10	15.14	89.62
4.	535.30	4.19	93.81
5.	344.04	2.70	96.51
6.	150.88	1.18	97.69
7.	101.09	0.79	98.48
8.	77.14	0.60	99.08
9.	55.51	0.43	99.52
10.	37.30	0.29	99.81
11.	22.90	0.18	99.99
12.	1.25	0.01	100.00

TABLE F-5

EIGENVECTORS (WEIGHTS) FOR OCCUPATION PRINCIPAL COMPONENTSPRINCIPAL COMPONENT

	<u>1.</u>	<u>2.</u>	<u>3.</u>	<u>4.</u>	<u>5.</u>	<u>6.</u>	<u>7.</u>	<u>8.</u>
1. MA	.00	-.11	.07	-.12	.10	-.11	.23	.27
2. PROF	.05	-.10	.04	-.34	-.07	-.33	-.57	.42
3. CLE	.20	-.13	.24	-.61	.13	.33	-.01	-.38
4. SA	.05	-.08	.08	.21	-.04	.10	.30	.25
5. SP	.49	.79	-.01	.15	.15	.05	-.05	.04
6. TP	-.03	-.07	.11	.07	.01	-.18	.19	-.59
7. PA	.05	-.22	-.71	.13	.11	.37	-.38	-.16
8. FAW	.05	-.13	-.45	.06	.08	-.23	.55	.18
9. LF	-.84	.40	.00	-.07	.10	.15	-.03	.02
10. MI	.03	.01	.08	.19	-.90	.20	.01	.00
11. CR	.02	-.31	.46	.56	.33	.31	-.09	.21
12. LA	-.04	-.06	.08	.23	.03	-.62	-.22	-.32

/...

TABLE F-6

CORRELATION COEFFICIENTS FOR OCCUPATION PRINCIPAL COMPONENTSPRINCIPAL COMPONENT

	<u>1.</u>	<u>2.</u>	<u>3.</u>	<u>4.</u>	<u>5.</u>	<u>6.</u>	<u>7.</u>	<u>8.</u>
1. MA	.03	-.61	.29	-.27	.19	-.13	.23	.24
2. PROF	.29	-.41	.13	-.59	-.10	-.30	-.42	.27
3. CLE	.62	-.29	.41	-.55	.09	.16	.00	-.13
4. SA	.37	-.44	.33	-.48	-.08	.12	.29	.21
5. SR	.66	.75	.00	.06	.05	.01	.00	.01
6. TP	-.22	-.41	.48	.15	.01	-.21	.19	-.51
7. FA	.10	-.36	-.91	.09	.06	.13	-.11	-.04
8. FAW	.19	-.32	-.86	.06	.07	-.12	.24	.07
9. LF	-.95	.32	.00	-.02	.03	.03	.00	.00
10. MI	.14	.02	.20	.24	-.93	.13	.01	.00
11. CR	.06	-.57	.66	.42	.20	.13	-.03	.06
12. LA	-.26	-.29	.29	.44	.05	-.63	-.19	-.23

In this case, the first eight components are required before accounting for 99% of the variability in the X matrix. This suggests again that the variability is spread among a fairly large number of occupations, although the first eigenvalue is somewhat larger than with the industry components. Here also, a given principal component can be associated with a relatively few occupations, but this time, the economic interpretation or rationale is sometimes obscure. For example, component one is associated with such service-oriented occupations as service and related (which is linked in turn with public administration) plus clerical, which seems reasonable enough, but the second represents an amalgam of service and related occupations and loggers and fishermen, for which no economic rationale is immediately

/...

- 17 -

apparent. Further, the third component - and to a lesser extent the eighth - represent a "clustering" of urban-oriented occupations, and the fourth a clustering of manufacturing and construction related occupations, but the economic rationale underlying the fifth to seventh components is far from clear. In short, the occupational principal components would seem to be inferior to the industry components as an indicator of linkages between economic activities.

Similar to the industry components, the occupation components were used in regression analysis with earnings as the dependent variable. The results are as follows:

$$\begin{aligned}
 (4) \quad \hat{W} = & 2,256.93 & + 32.45 \text{ PC } 1 & + 4.64 \text{ PC } 2 \\
 & (35.60) & (3.03) & (4.29) \\
 & (63.39) & (10.70) & (1.08) \\
 & + 35.29 \text{ PC } 3 & -24.98 \text{ PC } 4 & -14.62 \text{ PC } 5 \\
 & (5.49) & (10.44) & (13.02) \\
 & (6.42) & (-2.39) & (-1.12) \\
 & -3.32 \text{ PC } 6 & + 9.16 \text{ PC } 7 & -65.99 \text{ PC } 8 \\
 & (19.65) & (24.02) & (27.50) \\
 & (-0.17) & (0.38) & (-2.40)
 \end{aligned}$$

$$F - \text{STAT } (8, 37) = 21.23 \text{ SEE} = 241.48 \text{ } P^2 = 0.821 \text{ } \bar{P}^2 = 0.782 \text{ CV} = 10.70\%$$

$$\begin{aligned}
 (5) \quad \hat{W} = & 2,256.93 & + 32.44 \text{ PC } 1 & + 4.64 \text{ PC } 2 \\
 & (34.76) & (2.96) & (4.19) \\
 & (64.93) & (10.96) & (1.11) \\
 & + 35.29 \text{ PC } 3 & -24.98 \text{ PC } 4 & -14.62 \text{ PC } 5 & -65.99 \text{ PC } 8 \\
 & (5.36) & (10.19) & (12.71) & (26.84) \\
 & (6.58) & (-2.45) & (-1.15) & (-2.46)
 \end{aligned}$$

$$F - \text{STAT } (6, 39) = 29.66 \text{ SEE} = 235.76 \text{ } R^2 = 0.820 \text{ } \bar{P}^2 = 0.793 \text{ CV} = 10.45\%$$

/...

- 18 -

The signs on the components seem to be consistent with the "economic interpretations" provided above. Of particular interest is the fact that, in equation 4, the components with the lowest t-values are the ones that were most difficult to interpret, i.e. components 2, 5, 6 and 7. Also noteworthy is the fact that the adjusted R^2 in equation 5 is somewhat below that for industry equation 3, .793 compared with .853. The suggestion is then that in the Atlantic Region occupational structure is a less basic determinant of earnings than industry composition, (although note should be made of the fact that the Census categories are perhaps more aggregated in the case of occupation.) This result diverges from the findings of Hanna who concluded that occupation was the more basic dimension. On the other hand, the above finding provides support for the author's "philosophical" position that on balance the direction of causation is more from industry (i.e. the demand for skills) to occupation (their supply) rather than the other way around.

The above outlines two of the three sets of principal components that the author intended to compute. The third would have included both industry and occupation variables plus the other explanatory variables noted previously, i.e. labour force utilization measures, levels of education and urbanization, and so on. It was thought that this third set, together with the first two, would provide some useful insights into economic

/...

- 19 -

relationships in the Atlantic Region, and in particular would offer some guidelines regarding specifications for the non-wage equations in the model.^{3/} Unfortunately, time did not allow the computation of this third set of principal components.^{4/} It is strongly recommended though that such a set be computed before attempting to formulate a 1971 model.

- 20 -

FOOTNOTESAPPENDIX F:

1/ See Johnston, page 324.

2/ Which is equivalent to a given eigenvalue as a percentage of the total variation in the X matrix. Therefore, each percentage represents the total contribution of a given principal component to the total variation of the X's, and since the components are orthogonal these contributions sum to unity. See Johnston, page 325.

3/ Particularly in light of the lack of theory underlying some of the equations.

4/ The first two sets were computed using an APL program of I.P. Sharp. This program uses the $X - \bar{X}$ normalization, which is the appropriate normalization for occupational and industrial structure since our concern is with the "absolute" percentage deviations in the series. The third set though would require normalization of the form $(X - \bar{X})/S_x$ (i.e. deviations from the mean divided by the standard deviation of the X variable) since this set would involve variables with very different characteristics. There was insufficient time to mount the data on a new system and compute the third set of principal components.

APPENDIX G:
COMPARISON BETWEEN
THE ORDINARY LEAST SQUARES
AND TWO STAGE LEAST SQUARES RESULTS

APPENDIX G:

COMPARISON BETWEEN THE ORDINARY LEAST SQUARES AND TWO STAGE LEAST SQUARES RESULTS

As outlined in the Methodology section of the text, the procedure adopted to estimate the equations in the model involved three steps: (1) finding an initial specification using Ordinary Least Squares; (2) re-estimating the initial equation using Two Stage Least Squares; and (3) changing the specification where required. This exercise held many surprises. In particular, it was found that considerable respecification was required after moving to Two Stage Least Squares estimation. Because of this, it was decided that a separate appendix comparing the OLS and 2SLS results was warranted. Before looking at each equation separately, a few general comments should be made regarding these comparisons.

First, previous authors -- in particular Munro and Holmes^{1/} have argued that the application of simultaneous equation estimation techniques offers a partial solution to the problem of multicollinearity. In the present exercise, the exact opposite seemed to be the case. Once the industry variables were properly defined, few problems with multicollinearity arose in single equation estimation. On the

/...

other hand, the problem of multicollinearity seemed to reappear in 2SLS estimation, particularly between the "fitted" endogenous variables (i.e. the variables from the first stage of 2SLS). For example, when discussing the wage equation in the text, it was noted that the collinearity between residantary services and higher order business services seemed to be much greater under 2SLS than with OLS. In retrospect, this result seems far from surprising. The actual variables may be a function of countless factors, (including unfortunately the error term), with the relevant factors varying from variable to variable. In contrast, in the first stage of 2SLS, the endogenous variables are all constrained to be a function of the same set of variables, namely, the exogenous variables in the model. Therefore, it is perhaps not surprising that some of the endogenous variables became more collinear.

The second point may be related at least partially to the first but other factors are probably involved as well. In developing time-series models, it has generally been found that the results from 2SLS estimation are not too different from the results from OLS; in other words, the extent of OLS bias is usually quite small, leading some analysts to conclude that 2SLS is not worth the extra work. As the following will emphasize, the most emphatically has not been the case in the present exercise. In most equations, the 2SLS regression

/...

- 3 -

diverged considerably from the OLS results, forcing considerable respecification, and even when a variable remained in the equation, the magnitude of its coefficient generally differed significantly between the two sets of results. The reason for this departure from previous econometric work is readily apparent. In time-series analysis, the R^2 's in the first stage have generally been very high, usually well over 0.9. The upshot is that the fitted (Y) values are very similar to the actual (Y) values and thus the former's inclusion in the regression has only a minimal impact on the estimated coefficients. However, in cross-section work, there is not the same tendency for all variables to "move together", the tendency that is generally responsible for the high R^2 's in time-series.^{2/} As a consequence, the R^2 's in the first stage tend to be much lower, leading to considerable divergence between the fitted and actual (Y) values and major differences between the OLS and 2SLS results. In the current exercise, the R^2 's from the first stage regressions of 2SLS were generally in the .65 to .85 range, indicating that considerable variation in the endogenous variables was left unexplained.

A further point related to the first two is that for all equations the adjusted R^2 's declined, sometimes fairly markedly, in moving from OLS to 2SLS estimation.

/...

It appears then, that some important variability was lost from the endogenous variables in the first stage of estimation. A final consideration is that, in general, the number of right hand variables required for an equation declined in progressing from OLS to Two Stage Least Squares estimation. While only conjecture, one possibility is that this is related in part to the fact that, by using Two Stage Least Squares, explicit account is taken of the simultaneity in the system. In some cases, this seems to manifest itself in a tendency for the variability that is associated with many variables in OLS to become concentrated in one variable in 2SLS. Whatever the reason, there is a number of instances in the following where, in the 2SLS equation, one variable takes over the job that was previously shared among two or more variables in the OLS regression.

In comparing the results, the OLS and 2SLS regressions for both the initial and final specifications are indicated. The original OLS regression is equation (a), and its counterpart 2SLS regression is equation (b). Where alterations were required, the ~~original regression~~ final specification is equation (c), while the 2SLS regression -- the equation used in the model -- is equation (d). The equation used in the model -- either (b) or (d) -- is marked with an asterisk.

Only one specification was required for employment income per capita. The results were as follows:

/...

TABLE G-4

DEPENDENT VARIABLE: AVERAGE EARNINGS PER WAGE EARNER (W)

EQ. NO.	CONSTANT	AG	FOR	FISH	LVM	PP	ROPE	EH	PS	SE	NO
4 (a)	4,018.9 (158.4) (25.4)	-17.85 (3.85) (-4.63)	-17.60 (8.46) (-2.08)	-9.70 (5.44) (-1.78)	-25.73 (5.02) (-5.32)	18.00 (5.26) (3.42)	32.18 (11.30) (2.85)	-35.09 (10.32) (-3.40)	-26.25 (11.82) (-2.22)	-19.50 (5.95) (-3.28)	-63.96 (17.31) (-3.69)
	$R^2 = .937$	$R^2 = .919$	P-STAT (10.35) = 52.04 SEE = 147.35 CV = 6.524								
4 (b)	3,923.9 (227.9) (17.21)	-19.92 (5.49) (-3.62)	-27.53 (13.09) (-2.10)	-11.16 (7.94) (-1.41)	-32.37 (11.59) (-2.79)	14.92 (7.78) (1.94)	15.31 (24.65) (0.62)	-22.76 (14.50) (-1.57)	-11.38 (25.96) (-0.44)	-14.39 (11.10) (-1.30)	-65.39 (31.64) (-2.07)
	$R^2 = .928$	$R^2 = .906$	P-STAT (10.35) = 44.62 SEE = 158.31 CV = 7.019								
4 (c)	4,001.16 (145.28) (27.54)	-18.76 (4.00) (-6.69)	-25.21 (8.79) (-2.87)	-12.78 (4.95) (-2.58)	-27.32 (5.42) (-5.05)	13.81 (5.51) (2.51)		-24.96 (9.20) (-2.71)		-18.36 (6.45) (-2.65)	-68.53 (16.52) (-4.15)
	$R^2 = .921$	$R^2 = .904$	P-STAT (8.37) = 53.88 SEE = 160.52 CV = 7.119								
-4 (d)	3,898.75 (195.78) (19.91)	-20.28 (5.42) (-3.74)	-28.92 (13.54) (-2.14)	-13.36 (6.45) (-2.07)	-34.19 (10.68) (-3.28)	12.13 (6.41) (1.88)		-17.99 (11.32) (-1.59)		-15.54 (11.48) (-1.35)	-58.41 (30.52) (-1.91)
	$R^2 = .915$	$R^2 = .897$	P-STAT (8.37) = 50.01 SEE = 166.10 CV = 7.738								

N.B.: (a) and (c) estimated by OLS, (b) and (d) by 2SLS.

/...

- 6 -

$$1. \quad (a) \quad \text{Ln } \hat{\text{EIC}} = 1.244 + 0.809 \text{ Ln W} + 0.767 \text{ Ln (ER } \div 100)$$

$$\quad \quad \quad (0.651) \quad (0.076) \quad (0.096)$$

$$\quad \quad \quad (1.91) \quad (10.61) \quad (8.00)$$

$$R^2 = .876 \quad \bar{R}^2 = .870 \quad F\text{-STAT (2,43)} = 151.98 \quad \text{SEE} = 0.104$$

$$1. \quad (b) \quad * \text{LN } \hat{\text{EIC}} = 1.087 + 0.827 \text{ LN W} + 0.751 \text{ Ln (ER } \div 100)$$

$$\quad \quad \quad (0.690) \quad (0.081) \quad (0.102)$$

$$\quad \quad \quad (1.58) \quad (10.24) \quad (7.37)$$

$$R^2 = .876 \quad \bar{R}^2 = .870 \quad F\text{-STAT (2,43)} = 151.74 \quad \text{SEE} = 0.104$$

Except for the constant term, the estimated coefficients are not significantly different between the two equations, while the decrease in the adjusted R^2 is only apparent at the fourth decimal place. All-in-all, this was one of the more stable equations in going from OLS to 2SLS estimation.

The wage equation (Table G-4) underwent same change in moving to 2SLS estimation. Perhaps most significantly, two variables -- HOBS and RS -- became insignificant. There is some indication -- particularly from the much higher standard errors -- that the collinearity between these two variables experienced a marked increase when the equation was estimated using 2SLS, but the lower absolute magnitudes of the coefficients would suggest that this is not the only consideration. Beyond this, the absolute values of a number of coefficients changed fairly markedly; particularly noteworthy is the absolute increases for the three primary industries plus LVAM.

/...

TABLE G-5
DEPENDENT VARIABLE: THE DEGREE OF INEQUALITY IN INCOME (PII)

EQ. NO.	CONSTANT	ER	SE	HO	LED	HOB	OHVAM	FOR	FISH	EH
5 (a)	1.113 (0.164) (6.81)	-.0195 (.0053) (-3.68)	.0066 (.0031) (2.12)	.0216 (.0087) (2.47)	.0061 (.0039) (1.55)	-.0064 (.0060) (-1.07)	-.0096 (.0052) (-1.86)	-.0114 (.0041) (-2.75)	-.0046 (.0024) (-1.89)	.0052 (.0057) (0.90)
	$R^2 = .569$	$\bar{R}^2 = .461$	F-STAT (9,36) = 5.28	SEE = 0.0823	CV = 9.43%					
5 (b)	1.183 (0.217) (5.45)	-.0226 (.0088) (-2.55)	.0018 (.0061) (0.30)	.0388 (.0174) (2.23)	.0056 (.0072) (0.77)	-.0029 (.0089) (0.33)	-.0194 (.0098) (-1.99)	-.0088 (.0056) (-1.57)	-.0067 (.0035) (-1.92)	.0033 (.0081) (0.41)
	$R^2 = .491$	$\bar{R}^2 = .364$	F-STAT (9,36) = 3.86	SEE = 0.0894	CV = 10.24%					
5 (c)	1.220 (0.149) (8.19)	-.0150 (.0043) (-3.53)	.0154 (.0063) (2.45)				-.0086 (.0044) (-1.94)	-.0061 (.0039) (-1.57)		
	$R^2 = .432$	$\bar{R}^2 = .377$	F-STAT (4,41) = 7.80	SEE = 0.0885	CV = 10.14%					
5 (d)*	1.103 (0.171) (6.46)	-.0108 (.0051) (-2.11)	.0158 (.0072) (2.18)				-.0122 (.0058) (-2.10)	-.0044 (.0042) (-1.03)		
	$R^2 = .410$	$\bar{R}^2 = .353$	F-STAT (4,41) = 7.13	SEE = 0.0902	CV = 10.33%					

N.B.: (a) and (c) estimated by OLS, (b) and (d) by 2SLS.

TABLE G-6
DEPENDENT VARIABLE: UNEMPLOYMENT RATE UR

EQ. NO.	CONSTANT	AC	LVAR	HORS	RS	PA	TRANC	GLFD	GF	SE
5 (a)	-4.40 (4.49) (-0.98)	-0.123 (0.042) (-2.94)	-0.151 (0.060) (-2.50)	-0.184 (0.108) (-1.70)	0.387 (0.148) (2.62)	-0.078 (0.044) (-1.77)	-0.105 (0.087) (-1.20)	0.149 (0.056) (2.65)	-0.097 (0.044) (-2.18)	0.173 (0.060) (2.88)
	$R^2 = .798$	$\bar{R}^2 = .748$	F-STAT (9,36) = 15.81		SEE = 1.633		CV = 28.89%			
5 (b)	-3.14 (6.53) (-0.48)	-0.133 (0.054) (-2.48)	-0.203 (0.102) (-1.98)	-0.030 (0.246) (-0.12)	0.187 (0.374) (0.50)	-0.133 (0.088) (-1.51)	-0.146 (0.102) (-1.43)	0.171 (0.090) (1.91)	-0.103 (0.065) (-1.59)	0.160 (0.094) (1.69)
	$R^2 = .767$	$\bar{R}^2 = .709$	F-STAT (9,36) = 13.19		SEE = 1.752		CV = 31.00%			
6 (c)	-5.33 (4.57) (-1.17)	-0.125 (0.043) (-2.92)	-0.127 (0.060) (-2.11)		0.252 (0.128) (1.98)	-0.077 (0.045) (-1.69)	-0.123 (0.089) (-1.38)	0.178 (0.055) (3.24)	-0.122 (0.043) (-2.84)	0.183 (0.062) (2.98)
	$R^2 = .782$	$\bar{R}^2 = .735$	F-STAT (8,37) = 16.57		SEE = 1.674		CV = 29.62%			
6 (d)	-2.96 (6.36) (-0.47)	-0.134 (0.053) (-2.55)	-0.200 (0.100) (-2.00)		0.147 (0.186) (0.79)	-0.137 (0.080) (-1.71)	-0.151 (0.095) (-1.59)	0.176 (0.081) (2.18)	-0.106 (0.059) (-1.81)	0.158 (0.094) (1.69)
	$R^2 = .761$	$\bar{R}^2 = .709$	F-STAT (8,37) = 14.69		SEE = 1.754		CV = 31.03%			

N.B.: (a) and (c) estimated by OLS, (b) and (d) by 2SLS.

TABLE G-7

DEPENDENT VARIABLE: PARTICIPATION RATE PR

EQ. NO.	CONSTANT	CP	SE	GLPP	GE	PFELF	PA	AG	PP	HOBS
7 (a)	34.99 (3.82) (9.17)	-0.204 (0.174) (-1.17)	-0.053 (0.068) (-0.78)	-0.135 (0.072) (-1.89)	0.165 (0.056) (2.96)	0.343 (0.105) (3.27)	0.133 (0.053) (2.91)	0.136 (0.047) (2.91)	0.193 (0.066) (2.92)	0.118 (0.150) (0.79)
	$R^2 = .869$ $\bar{R}^2 = .836$ F-STAT (9,36) = 26.52 SEE = 1.953 CV = 4.23%									
7 (b)	31.67 (7.48) (4.23)	-0.133 (0.454) (-0.29)	0.012 (0.105) (0.12)	-0.139 (0.144) (-0.96)	0.194 (0.100) (1.94)	0.236 (0.203) (1.03)	0.076 (0.111) (0.68)	0.138 (0.059) (2.36)	0.189 (0.081) (2.34)	0.335 (0.259) (1.29)
	$R^2 = .851$ $\bar{R}^2 = .814$ F-STAT (9,36) = 22.83 SEE = 2.083 CV = 4.51%									
7 (c)	34.87 (3.41) (10.22)	-0.219 (0.049) (-4.48)	0.230 (0.039) (5.88)	0.310 (0.103) (3.01)	0.136 (0.049) (2.77)	0.131 (0.041) (3.16)	0.188 (0.065) (2.90)	0.197 (0.140) (1.41)		
	$R^2 = .859$ $\bar{R}^2 = .833$ F-STAT (7,38) = 33.16 SEE = 1.969 CV = 4.26%									
7 (d)*	33.20 (4.96) (6.69)	-0.176 (0.058) (-3.01)	0.216 (0.054) (3.93)	0.171 (0.155) (1.10)	0.093 (0.090) (1.04)	0.147 (0.044) (3.31)	0.186 (0.078) (2.39)	0.361 (0.207) (1.74)		
	$R^2 = .846$ $\bar{R}^2 = .818$ F-STAT (7,38) = 29.90 SEE = 2.058 CV = 4.46%									

N.B.: (a) and (c) estimated by OLS, (b) and (d) by 2SLS.

/...

- 10 -

The specification for the degree of inequality in income almost disintegrated in moving to 2SLS estimation (Table G-5). As 5 (b) indicates, four of the nine explanatory variables ceased adding to the adjusted R^2 under 2SLS; moreover, the coefficient on FISH failed to stand up under repeated 2SLS testing. As a consequence, only four of the nine variables in the original OLS specification were included in the final equation (5 (d)), while the magnitudes of the remaining coefficients were altered considerably. The final point that should be noted is the considerable drop in explanatory power in moving to 2SLS estimation.

For the unemployment rate equation (Table G-6), only one variable, HOBS, was removed from the equation after moving to 2SLS, (although the t-value on one other, RS, also moved below one -- this was kept in because of probable multicollinearity between RS, GLFP and GE and because the size of the coefficient seemed reasonable.) Otherwise, the specification held up fairly well.

Similarly, the participation rate equation did not do too badly under 2SLS (Table G-7). One disappointment though was that UR and SE -- reflecting "the discouraged worker withdrawal" hypothesis -- had to be removed from the equation. Other significant changes included the substantial

/...

TABLE G-8

DEPENDENT VARIABLE: SEASONALITY (SE)

EQ. NO.	CONSTANT	AG	FOR	FISH	PA	PS	EH	CON	UF	PFR
8 (a)	14.25 (6.49) (2.20)	0.393 (0.079) (4.96)	0.572 (0.160) (3.57)	0.380 (0.094) (4.05)	-0.115 (0.072) (-1.60)	-0.352 (0.280) (-1.26)	-0.287 (0.215) (-1.33)	0.82 (0.193) (1.98)	0.696 (0.244) (2.85)	0.041 (0.023) (1.74)
	$R^2 = .864$		$\bar{R}^2 = .830$	F-STAT (9,36) = 25.41		SEE = 3.346		CV = 15.26%		
8 (b)	6.35 (9.47) (0.67)	0.434 (0.101) (4.30)	0.725 (0.214) (3.38)	0.472 (0.120) (3.94)	-0.048 (0.095) (-0.50)	0.163 (0.440) (0.37)	-0.465 (0.298) (-1.56)	0.429 (0.207) (2.07)	0.470 (0.407) (1.15)	0.052 (0.025) (2.03)
	$R^2 = .848$		$\bar{R}^2 = .810$	F-STAT (9,36) = 22.26		SEE = 3.541		CV = 16.15%		
8 (c)	5.62 (3.71) (1.52)	0.461 (0.069) (6.72)	0.659 (0.154) (4.28)	0.476 (0.073) (6.50)	-0.044 (0.209) (-0.21)	-0.188 (0.210) (-0.90)	0.442 (0.194) (2.28)	0.676 (0.249) (2.71)	0.043 (0.024) (1.79)	
	$R^2 = .854$		$\bar{R}^2 = .823$	F-STAT (8,37) = 27.12		SEE = 3.416		CV = 15.58%		
8 (d)*	2.10 (4.30) (0.49)	0.466 (0.079) (5.90)	0.759 (0.206) (3.69)	0.516 (0.081) (6.37)	0.338 (0.272) (1.24)	-0.444 (0.300) (-1.48)	0.452 (0.205) (2.20)	0.472 (0.413) (1.14)	0.053 (0.026) (2.06)	
	$R^2 = .838$		$\bar{R}^2 = .804$	F-STAT (8,37) = 24.00		SEE = 3.597		CV = 16.40%		

N.B.: (a) and (c) estimated by OLS; (b), and (d) by 2SLS

TABLE G-9

DEPENDENT VARIABLE: HOURS (HO)

EQ. NO.	CONSTANT	FOP	FISH	EH	PA	TRANS	MIN
9 (a)	12.24 (1.23) (9.96)	-0.123 (0.041) (2.99)	0.097 (0.027) (3.57)	-0.256 (0.084) (-3.06)	-0.214 (0.020) (-5.63)	-0.082 (0.066) (-1.25)	-0.073 (0.047) (-1.56)
	$R^2 = .671$	$\bar{R}^2 = .621$	F-STAT (6,39) = 13.28	SEE = 1.309	CV = 16.34%		
9 (b)*	13.03 (1.45) (9.01)	-0.123 (0.041) (-2.97)	0.089 (0.029) (3.12)	-0.324 (0.105) (-3.07)	-0.121 (0.021) (-5.64)	-0.094 (0.067) (-1.40)	-0.077 (0.047) (-1.62)
	$R^2 = .666$	$\bar{R}^2 = .615$	F-STAT (6,39) = 12.96	SEE = 1.320	CV = 16.47%		

N.B.: (a) estimated by OLS, (b) by 2SLS

TABLE G-10

DEPENDENT VARIABLE: PER CENT FEMALE IN THE EMPLOYED LABOUR FORCE (PFELF)

EQ.
NO.

	CONSTANT	LVAN	HOBS	EH	RS	FISH	TRANS	PFP
10 (a)	6.54 (2.26) (2.89)	0.170 (0.075) (2.27)	0.573 (0.155) (3.69)	0.772 (0.157) (4.91)	0.225 (0.164) (1.37)	-0.059 (0.058) (-1.02)	-0.182 (0.116) (-1.57)	
	$R^2 = .787$	$\bar{R}^2 = .754$	F-STAT (6,39) = 24.02		SEE = 2.391	CV = 10.78%		
10 (b)	6.06 (2.67) (2.27)	0.147 (0.124) (1.18)	0.458 (0.243) (1.89)	0.596 (0.203) (2.94)	0.473 (0.276) (1.71)	-0.064 (0.079) (-0.81)	-0.211 (0.124) (-1.70)	
	$R^2 = .767$	$\bar{R}^2 = .732$	F-STAT (6,39) = 21.45		SEE = 2.499	CV = 11.27%		
10 (c)	6.50 (1.96) (3.32)		0.693 (0.132) (5.27)	0.535; (0.153) (3.51)	0.277 (0.152) (1.82)	-0.173 (0.108) (-1.60)	0.046 (0.015) (3.17)	
	$R^2 = .807$	$\bar{R}^2 = .783$	F-STAT (5,40) = 31.52		SEE = 2.246	CV = 10.13%		
10 (d)*	5.29 (2.46) (2.15)		0.656 (0.196) (3.34)	0.258 (0.205) (1.26)	0.562 (0.251) (2.23)	-0.225 (0.118) (-1.90)	0.054 (0.016) (3.36)	
	$R^2 = .775$	$\bar{R}^2 = .747$	F-STAT (5,40) = 27.63		SEE = 2.424	CV = 10.93%		

N.B.: (a) and (c) by OLS, (b) and (d) by 2SLS.

increase in the HOBS variable, and the corresponding decline in PFELF, suggesting a reversal in their relative roles in moving to 2SLS.

Two significant changes occurred to the seasonality equation in estimating by 2SLS (Table G-8). First, the t-value on PA dropped below one in absolute terms. Second, the sign on RS switched to become more in accord with prior expectations. In general then, the specification improved in going to 2SLS estimation. The equation for HO (Table G-9) altered very little under 2SLS, largely reflecting the fact that five of the six variables in the equation are exogenous. The equation for PFELF (Table G-10) fared almost as well, with the t-value on only coefficient, FISH, falling below one. However, with the addition of PFR, a second variable, LVAM ceased adding to the adjusted R^2 and was removed from the equation.

The results for the GLFP and NMR equations are as follows:

$$\begin{array}{lll}
 11. & (a) & \hat{GLFP} = 95.54 + 0.854 \text{ NMR} + 0.828 \text{ NRI} \\
 & & (0.603) \quad (0.012) \quad (0.024) \\
 & & (158.38) \quad (72.12) \quad (35.13) \\
 & & R^2 = .995 \quad \bar{R}^2 = .995 \quad F\text{-STAT} (2,43) = 4,408.9 \quad SEE = 1.506 \\
 11. & (b) & * \hat{GLFP} = 95.48 + 0.850 \text{ NMR} + 0.829 \text{ NRI} \\
 & & (0.622) \quad (0.014) \quad (0.024) \\
 & & (153.46) \quad (60.61) \quad (34.53) \\
 & & R^2 = .995 \quad \bar{R}^2 = .995 \quad F\text{-STAT} (2,43) = 4,402.13 \quad SEE = 1.507
 \end{array}$$

TABLE G-13

DEPENDENT VARIABLE: GROWTH IN EMPLOYMENT (GE)

EQ. NO.	CONSTANT	AG	FOR	FISH	MIN	LVAM	RS	PA	URB
13 (a)	140.55 (30.61) (4.60)	-0.455 (0.411) (-1.11)	-0.966 (0.592) (-1.63)	-0.601 (0.416) (-1.44)	-1.142 (0.589) (-1.94)	-0.622 (0.537) (-1.16)	-2.699 (1.255) (-2.15)	1.112 (0.323) (3.44)	0.284 (0.141) (2.02)
	$R^2 = .776$	$\bar{R}^2 = .728$	F-STAT (8,37) = 16.04			SEE = 13.972	CV = 13.848		
13 (b)	169.26 (46.24) (3.66)	-0.719 (0.466) (-1.54)	-1.261 (0.673) (-1.87)	-0.884 (0.611) (-1.45)	-1.388 (0.707) (-1.96)	-0.750 (0.966) (-3.78)	-3.875 (2.650) (-1.46)	0.876 (0.457) (1.92)	0.244 (0.239) (1.02)
	$R^2 = .770$	$\bar{R}^2 = .721$	F-STAT (8,37) = 15.50			SEE = 14.157	CV = 14.038		
13 (c)	123.55 (26.94) (4.59)	-0.303 (0.390) (-0.77)	-0.613 (0.510) (-1.20)	-0.610 (0.418) (-1.46)	-0.954 (0.569) (-1.68)		-2.400 (1.233) (-1.96)	1.280 (0.290) (4.42)	0.338 (0.133) (2.54)
	$R^2 = .778$	$\bar{R}^2 = .725$	F-STAT (7,38) = 17.98			SEE = 14.034	CV = 13.918		
13 (d)*	170.68 (47.83) (3.57)	-0.679 (0.479) (-1.42)	-0.988 (0.593) (-1.66)	-1.102 (0.562) (-1.96)	-1.387 (0.732) (-1.90)		-4.678 (2.526) (-1.85)	0.880 (0.473) (1.86)	0.347 (0.205) (1.69)
	$R^2 = .747$	$\bar{R}^2 = .700$	F-STAT (7,38) = 16.04			SEE = 14.655	CV = 14.528		

N.B.: (a) and (c) by OLS, (b) and (d) by 2SLS

TABLE G-14

DEPENDENT VARIABLE: LEVEL OF URBANIZATION (URR)

EQ. NO.	CONSTANT	MIN	PP	EH	PA	HOBS	OHVAM	LVAM	NFD
14 (a)	-49.15 (13.24) (-3.71)	1.01 (0.55) (1.83)	1.49 (0.40) (3.68)	2.15 (0.96) (2.25)	0.92 (0.21) (4.35)	2.67 (0.81) (3.31)	2.70 (0.87) (3.09)	1.01 (0.48) (2.08)	24.89 (5.32) (4.66)
	$R^2 = .740$	$\bar{R}^2 = .684$	F-STAT (8,37) = 13.18		SEE = 12.512		CV = 34.67%		
14 (b)	-33.03 (23.04) (-1.43)	0.38 (1.08) (0.35)	1.26 (0.51) (2.45)	1.79 (1.26) (1.42)	0.82 (0.26) (3.21)	1.89 (1.77) (1.07)	4.05 (2.29) (1.77)	0.31 (0.93) (0.34)	24.79 (5.83) (4.25)
	$R^2 = .704$	$\bar{R}^2 = .640$	F-STAT (8,37) = 10.98		SEE = 13.365		CV = 37.04%		
14 (c)	-25.46 (8.71) (-2.92)	1.14 (0.40) (2.86)	1.65 (0.91) (1.82)	0.74 (0.18) (4.18)	2.04 (0.75) (2.72)	3.16 (0.77) (4.08)			22.75 (5.17) (4.40)
	$R^2 = .695$	$\bar{R}^2 = .648$	F-STAT (6,39) = 14.83		SEE = 13.201		CV = 36.58%		
14 (d)*	-25.50 (10.54) (-2.42)	1.14 (0.42) (2.72)	1.78 (1.17) (1.52)	0.79 (0.19) (4.09)	1.39 (1.03) (1.35)	4.79 (1.13) (4.02)			24.83 (5.56) (4.46)
	$R^2 = .661$	$\bar{R}^2 = .609$	F-STAT (6,39) = 12.67		SEE = 13.924		CV = 38.59%		

N.B.: (a) and (c) by OLS, (b) and (d) by 2SLS

5

5

$$12. \quad (a) \quad \hat{NMR} = -75.08 + 0.729 \text{ GE} - 0.145 \text{ NRI} - 0.069 \text{ PFR}$$

(3.32)	(0.033)	(0.090)	(0.034)
(-22.61)	(21.79)	(-1.60)	(-2.03)

$$R^2 = .931 \quad \bar{R}^2 = .926 \quad F\text{-STAT} (3,42) = 187.84 \quad SEE = 5.432$$

$$12. \quad (b) \quad *NMR = -72.84 + 0.700 \text{ GE} - 0.112 \text{ NRI} - 0.074 \text{ PFR}$$

(3.68)	(0.039)	(0.094)	(0.034)
(-19.79)	(17.97)	(-1.20)	(-2.14)

$$R^2 = .929 \quad \bar{R}^2 = .924 \quad F\text{-STAT} = 184.38 \quad SEE = 5.479$$

The equation for GLFP showed almost no change, as the equations with the highest adjusted R^2 's also tended to be the most stable. The alteration in the NMR equation was somewhat larger -- note in particular the drop in absolute magnitude of the NRI coefficient -- but it also stood up very well relative to most equations.

Although only one variable dropped out, LVAM, the GE equation (Table G-13) changed considerably in moving to 2SLS estimation. Particularly noteworthy are the sharp absolute increases experienced by RS (an increase that perhaps proved troublesome in solving the model) and FISH, and the increase in URB when LVAM was removed.

Two variables, LVAM and MIN, dropped out of the URB equation (Table G-14) when it was estimated using 2SLS. This perhaps correspond to the strong increase in the OHVAM coefficient, as the variability spread among the three variables

/...

TABLE G-15

DEPENDENT VARIABLE: LEVEL OF EDUCATION (LEVS)

EQ. NO.	CONSTANT	FOR	FISH	HOBBS	RS	EH	FA	TRANS
15 (a)	4.88 (5.86) (0.89)	-0.340 (0.136) (-2.50)	-0.181 (0.089) (-2.03)	0.638 (0.277) (2.30)	0.556 (0.291) (1.91)	0.411 (0.266) (1.55)	0.211 (0.067) (3.17)	-0.219 (0.190) (-1.15)
	$R^2 = .790$	$\bar{R}^2 = .751$	F-STAT	(7,38) = 20.42	SEE = 3.643	CV = 17.50%		
15 (b)	3.98 (7.33) (0.54)	-0.326 (0.159) (-2.05)	-0.171 (0.095) (-1.81)	0.654 (0.535) (1.22)	0.629 (0.584) (1.08)	0.358 (0.359) (1.00)	0.218 (0.082) (2.65)	-0.233 (0.206) (-1.13)
	$R^2 = .789$	$\bar{R}^2 = .750$	F-STAT	(7,38) = 20.32	SEE = 3.650	CV = 17.53%		

N.B.: (a) estimated by OLS, (b) by 2SLS

TABLE C-16

DEPENDENT VARIABLE: HIGHER ORDER BUSINESS SERVICES HOURS

EQ. NO.	CONSTANT	LED	EIC - .100	OHVAM	EH	TRANS	POP - 1,000	RS	AG	EVAM
16 (a)	-6.23 (1.87) (-3.32)	0.099 (0.077) (1.28)	0.718 (0.315) (2.28)	0.111 (0.110) (1.01)	0.306 (0.116) (2.64)	0.199 (0.093) (2.13)	0.211 (0.208) (1.31)	0.306 (0.120) (2.54)	0.092 (0.038) (2.54)	
	$R^2 = .818$	$\bar{R}^2 = .779$		F-STAT (8,37) = 20.80	SEE = 1.793			CV = 16.32%		
16 (b)	-7.38 (2.68) (-2.75)	0.221 (0.135) (1.64)	0.431 (0.588) (0.22)	-0.042 (0.206) (-0.21)	0.071 (0.176) (0.41)	0.191 (0.116) (1.64)	0.017 (0.010) (1.76)	0.635 (0.203) (3.13)	0.076 (0.045) (1.69)	
	$R^2 = .756$	$\bar{R}^2 = .703$		F-STAT (8,37) = 14.29	SEE = 2.348			CV = 18.94%		
16 (c)	-5.71 (1.81) (-3.17)	0.266 (0.049) (5.46)		0.230 (0.121) (1.90)	0.251 (0.097) (2.58)	0.014 (0.008) (2.46)	0.365 (0.120) (3.05)	0.066 (0.038) (1.75)	0.0370 (0.036) (0.65)	
	$R^2 = .783$	$\bar{R}^2 = .743$		F-STAT (7,38) = 19.57	SEE = 1.922			CV = 17.61%		
16 (d)*	-7.27 (1.98) (-3.67)	0.273 (0.061) (4.48)		0.100 (0.165) (0.61)	0.215 (0.107) (2.01)	0.021 (0.009) (2.36)	0.510 (0.181) (2.81)	0.073 (0.040) (1.74)	0.024 (0.035) (1.23)	
	$R^2 = .756$	$\bar{R}^2 = .711$		F-STAT (7,38) = 16.82	SEE = 2.038			CV = 18.66%		

N.B.: (a) and (c) estimated by OLS, (b) and (d) by 2SLS

TABLE G-17

DEPENDENT VARIABLE: OTHER HIGH VALUE ADDED MANUFACTURING (OHVAM)

EQ. NO.	CONSTANT	MIN	HOB\$	RS	URB	NFD	UD
17 (a)	-6.09 (1.77) (-3.45)	0.256 (0.085) (3.01)	0.145 (0.150) (0.97)	0.424 (0.146) (2.91)	0.037 (0.021) (1.79)	-2.38 (0.92) (-2.58)	
	$R^2 = .623$	$\bar{R}^2 = .575$	F-STAT (5,40) = 13.20	SEE = 2.144	CV = 66.77%		
17 (b)	-7.56 (2.19) (-3.46)	0.281 (0.093) (3.04)	0.125 (0.235) (0.53)	0.563 (0.231) (2.44)	0.024 (0.028) (0.87)	-2.22 (1.07) (-2.08)	
	$R^2 = .610$	$\bar{R}^2 = .561$	F-STAT (5,40) = 12.52	SEE = 2.179	CV = 67.86%		20
17 (c)	-6.19 (1.80) (-3.45)	0.297 (0.080) (3.74)	0.285 (0.124) (2.31)	0.416 (0.148) (2.81)		-1.76 (0.84) (-2.09)	-1.28 (0.92) (-1.39)
	$R^2 = .611$	$\bar{R}^2 = .563$	F-STAT (5,40) = 12.58	SEE = 2.176	CV = 67.77%		
17 (d)*	-7.55 (2.16) (-3.49)	0.303 (0.081) (3.76)	0.217 (0.190) (1.14)	0.556 (0.233) (2.39)		-1.83 (0.89) (-2.05)	-1.24 (0.93) (-1.33)
	$R^2 = .602$	$\bar{R}^2 = .553$	F-STAT (5,40) = 12.12	SEE = 2.201	CV = 68.54%		

N.B.: (a) and (c) estimated by OLS, (b) and (d) by 2SLS

- 21 -

under OLS seemed to be concentrated in OHVAM under simultaneous equation estimation. This would appear to be the preferable result. The LED equation (Table G-15) stood up very well under 2SLS, as all variables continued adding to the adjusted R^2 and the alterations in the sizes of the coefficients were relatively minor.

The HOBS equation (Table G-16) changed considerably when estimated using Two Stage Least Squares. The t-values on two coefficients, EIC and OHVAM, fell well below one as the variability associated with these variables seemed to become concentrated in LED under 2SLS. Moreover, LVAM, which had a t-value below one using OLS, added to the adjusted R^2 when simultaneous equation estimation techniques were applied. Beyond the "demise" of URB, the OHVAM equation (Table G-17) did not fare too badly under 2SLS, particularly for an equation with a relatively low $\bar{R}^2$.

The equations for RS, EH and LVAM are as follows:

$$18. \quad (a) \quad \hat{RS} = 14.81 + 0.907 (EIC \div 100) + 0.033 URB - 0.072 GE$$

(1.47)	(0.232)	(0.020)	(0.013)
(10.09)	(3.92)	(1.71)	(-5.42)

$$R^2 = .541 \quad \bar{R}^2 = .508 \quad F\text{-STAT} (3,42) = 16.48 \quad SEE = 2.090$$

$$18. \quad (b) \quad \hat{*RS} = 14.98 + 1.013 (EIC \div 100) + 0.043 URB - 0.084 GE$$

(1.66)	(0.287)	(0.024)	(0.015)
(9.04)	(3.53)	(1.74)	(-5.56)

$$R^2 = .526 \quad \bar{R}^2 = .492 \quad F\text{-STAT} (3,42) = 15.51 \quad SEE = 2.124$$

/...

- 22 -

$$19. \quad (a) \quad \hat{EH} = 3.26 + 0.403 \text{ HOBS} + 5.04 \text{ UD}$$

$$\quad \quad \quad (0.72) \quad (0.061) \quad (0.63)$$

$$\quad \quad \quad (4.55) \quad (6.65) \quad (7.95)$$

$$R^2 = .694 \quad \bar{R}^2 = .680 \quad F\text{-STAT} (2,43) = 48.83 \quad SEE = 1.536$$

$$19. \quad (b) \quad *EH = 3.15 + 0.414 \text{ HOBS} + 5.05 \text{ UD}$$

$$\quad \quad \quad (0.77) \quad (0.066) \quad (0.63)$$

$$\quad \quad \quad (4.07) \quad (6.24) \quad (7.96)$$

$$R^2 = .694 \quad \bar{R}^2 = .680 \quad F\text{-STAT} (2,43) = 48.78 \quad SEE = 1.536$$

$$20. \quad (a) \quad LVAM = 14.43 + 0.109 \text{ FISH} - 0.488 (W \div 100) + 0.290 \text{ RS}$$

$$\quad \quad \quad (6.26) \quad (0.114) \quad (0.174) \quad (0.249)$$

$$\quad \quad \quad (2.30) \quad (0.95) \quad (-2.80) \quad (1.16)$$

$$R^2 = .326 \quad \bar{R}^2 = .278 \quad F\text{-STAT} (3,42) = 6.76 \quad SEE = 4.629$$

$$20. \quad (b) \quad *LVAM = 6.09 + 0.188 \text{ FISH} - 0.431 (W \div 100) + 0.729 \text{ RS}$$

$$\quad \quad \quad (7.40) \quad (0.127) \quad (0.202) \quad (0.298)$$

$$\quad \quad \quad (0.82) \quad (1.48) \quad (-2.13) \quad (2.45)$$

$$R^2 = .276 \quad \bar{R}^2 = .224 \quad F\text{-STAT} (3,42) = 5.33 \quad SEE = 4.80$$

The RS equation underwent only minor changes in moving to 2SLS, as all variables remained significant and the changes in the coefficients were relatively minor. Also, the EH equation remained essentially unchanged. Finally, the equation for LVAM if anything improved, despite the drop in the $\bar{R}^2$. Important here is the increased size of the FISH and RS coefficients.

/...

FOOTNOTES

APPENDIX G:

^{1/}See page 67 of their article.

^{2/}Time series variables tend to "move together" because of the common time trend in each variable. The latter results in higher R^2 's, but also leads to problems with auto-correlation. The latter tends to be minimal in cross-sectional work; this becomes the "good news" part of this "good news, bad news" story in the context of the present analysis.

APPENDIX H:

COMPARISON OF PRELIMINARY REPORT

RESULTS WITH FINAL RESULTS

APPENDIX H:

COMPARISON OF PRELIMINARY REPORT RESULTS WITH FINAL RESULTS

This appendix is for the benefit of those people who received the author's preliminary report (including the supplementary paper). The differences between the preliminary and final estimation and solution results mainly reflect the fact that, at the time of the preliminary report, a computer software package for Two Stage Least Squares and for solving non-linear systems was not available to the author. More specifically, the preliminary and final results can differ for the following reasons:

- (1) Because of "work space full" problems with the computer programs used during the preliminary work, the number of explanatory variables included in the first stage of Two Stage Least Squares had to be reduced from (14) to (11). This was accomplished by excluding the working population ratio - since it only appeared in an identity - and adding the three primary industries together. In the final analysis, this raised more difficulties than was originally anticipated and, in particular, doubtless led to some mis-specification.

- 2) The 2SLS algorithm used in preliminary work provided "correct" estimates of the coefficients - abstracting from the difficulties noted in (1) above - but incorrect estimates of the "diagnostic" statistics (R^2 , $\bar{R}^2$, standard errors, etc.). This probably also resulted in some specification error.
- 3) Since a non-linear solution algorithm was not available when the preliminary model was solved (see the "supplementary paper"), linear approximations had to be derived for the non-linear equations in the model; the model was then solved using linear algebra.
- 4) Considerable "rethinking" regarding specification was done before the final estimation work was completed.

No attempt will be made here to compare all of the preliminary and final estimation and solution results. Rather, a few examples will be provided. Below, comparisons are presented between the preliminary (P) and final (F) regressions for the wage and participation rate equations in the model. (Only the t-statistics are indicated below the estimated coefficients).

4 (P) $\hat{W} = 3790.51 - 8.59 \text{ AG} - 18.80 \text{ FOR} - 23.05 \text{ FISH}$
 (14.29) (-1.08) (1.73) (-3.01)

- 21.88 LYAM +10.81 PP -14.64 EH
 (-1.59) (1.38) (-1.00)

- 21.23 SE -62.76 HO +224.1 NFD
 (-1.72) (-1.58) (2.07)

$$R^2 = .872 \quad \bar{R}^2 = .848$$

4 (F) $\hat{W} = 3898.75 - 20.28 \text{ AG} - 28.92 \text{ FOR} - 13.36 \text{ FISH}$
 (19.91) (-3.74) (-2.14) (-2.07)

-34.19 LVAM +12.13 PP -17.99 EH
 (-3.20) (1.89) (-1.59)

-15.54 SE -58.41 HO
 (-1.35) (-1.91)

$$R^2 = .915 \quad \bar{R}^2 = .897$$

- 4 -

$$6 \text{ (P) } \hat{PR} = 23.50 - 0.200 \text{ GLFP} + 0.312 \text{ GE} \\ (3.98) \quad (-2.79) \quad (5.16)$$

$$+0.455 \text{ PFELE} + 0.242 \text{ AG} \\ (3.62) \quad (3.91)$$

$$+0.163 \text{ PP} + 0.076 \text{ FISH} \\ (1.90) \quad (0.98)$$

$$R^2 = .737 \quad \bar{R}^2 = .696$$

$$6 \text{ (F) } \hat{PR} = 33.20 - 0.176 \text{ GLFP} + 0.216 \text{ GE} \\ (6.69) \quad (-3.01) \quad (3.93)$$

$$+ 0.171 \text{ PFELF} + 0.147 \text{ AG} \\ (1.10) \quad (3.31)$$

$$+ 0.186 \text{ PP} + 0.093 \text{ PA} + 0.361 \text{ HOBS} \\ (2.39) \quad (1.04) \quad (1.74)$$

$$R^2 = .846 \quad \bar{R}^2 = .818$$

These comparisons indicate both similarities and important differences between the preliminary and final estimation results.

For the Wage equation, there is only one difference in specification - the Newfoundland dummy variable (NFD) was not significant in the final equation - but the magnitudes of the coefficients altered greatly between the preliminary and final regressions. For the participation rate equation, two additional variables were added to the final equation, PA and HOBS, while one was dropped, FISH. This led in turn to alterations to other variables, particularly P FELF. Finally, it may be noted that the adjusted R^2 's are higher in the final equations. This was generally the case.

Despite the differences between the estimation results and the solution algorithms, the solution results for the preliminary and final reports are quite similar. The following compares the results for Sensitivity Test A - which focussed on employment shifts out of agriculture and into transportation and pulp and paper - and Scenario H. The latter involved the 1971 mean values for the X variables, plus adjustments to the intercepts in the wage, level of education, higher order business services, other high value added manufacturing, residentiary services and lower value added manufacturing equations. The preliminary and final solution results are noted by P and F respectively.

COMPARISON OF SOLUTION RESULTS

	<u>Sensitivity test A</u>		<u>Scenario H</u>	
	P	F	P	F
EIC	680.32	688.17	880.20	908.03
ER	27.09	27.13	30.74	31.29
ELFB	94.38	94.37	94.52	94.46
W	2336.24	2377.66	2884.63	2920.84
DII	0.870	0.870	0.809	0.802
UP	5.62	5.63	5.48	5.54
PR	45.93	46.19	48.09	48.70
SE	20.64	20.53	17.20	17.51
HO	7.83	7.80	8.13	7.84
PFEF	21.94	21.93	23.14	26.26
GLFP	110.37	110.34	110.91	107.31
NMR	-4.63	-4.64	1.69	-2.34
OE	103.05	103.10	111.06	105.35
URB	38.46	38.06	46.78	57.04
LED	20.51	20.56	28.20	30.78
HOBS	11.13	11.09	12.86	14.88
OHVAM	3.36	3.31	4.97	5.98
RS	15.01	14.96	15.16	16.82
EH	8.51	8.50	9.24	10.06
LVAM	8.24	8.00	6.38	7.38
IR	2.74	3.14	9.41	3.18

- 7 -

The changes implied by the solution results tend to be larger for the final model relative to the preliminary results. This is in general the desired outcome. Beyond this, the results are not too dissimilar, suggesting that, at the very least, the preliminary results were not misleading.

APPENDIX I:

DATA SOURCES

APPENDIX I:

DATA SOURCES

Almost all data used in this exercise was taken from the Population Census of Statistics Canada for the years 1951, 1961 and 1966. Of particular significance is the Labour Force section (i.e. Volume 3 in the 1961 Census) of the Census which provided the industrial and occupational distributions, average and median earnings, and the labour force utilization data, among other series. The published data on the industrial distribution of the labour force was augmented by unpublished information provided to the author by the Census Division. This provided us with twenty-six industrial groupings, rather than the twelve in the published material.

For the most part, the data could be used directly from the Census, once the required calculations had been made (percentage distributions, percentage changes and so on). However, two income series, employment and personal income per capita, were exceptions to this. In both cases, Census materials were used as a basis for the development of estimates.

In estimating personal income for 1961, the major source was the Population Sample Census of 1961 which provided estimates of total money income in the non-farm sector by

/...

by county and Census Division. The major difficulty therefore was to estimate income in the rural farm sector. This was done in three steps: (1) total farm earnings were calculated on a county basis from data found in the Labour Force Census; (2) the percentage distribution of farm earnings by county in a given province was applied to the provincial total of net income of farm operators taken from the National Accounts in order to derive county estimates of the latter; (3) government transfer payments in the rural farm sector were estimated for each county by:

(i) calculating the rural farm portion of government transfer payments for the province as a whole by applying the rural farm share of population to the provincial total of government transfer payments taken from the National Accounts and

(ii) distributing the rural farm portion according to the percentage distribution by county of the rural farm population -- this assumes that, within the rural farm sector, per capita government transfer payments are equal across counties within a given province, irrespective of relative income levels.

The estimates of farm income were added to non-farm income to obtain "total allocated" income by county. Personal income was then obtained by "marking up" the allocated income figure by the ratio of total personal income to total allocated income at the provincial level. Allocated income generally accounted for about four-fifths of total personal income.

/...

Finally, employment income for 1961 was calculated by adding non-farm employment income -- also included in the Population Sample Census -- to the estimates of farm earnings and net income of farm operators. The Census of Manufacturers Survey provided the basic data for the calculation of value added and non-wage value added per employee in manufacturing. These were the only series for which a source other than the Population Census was used.

END

26

02

75

FIN