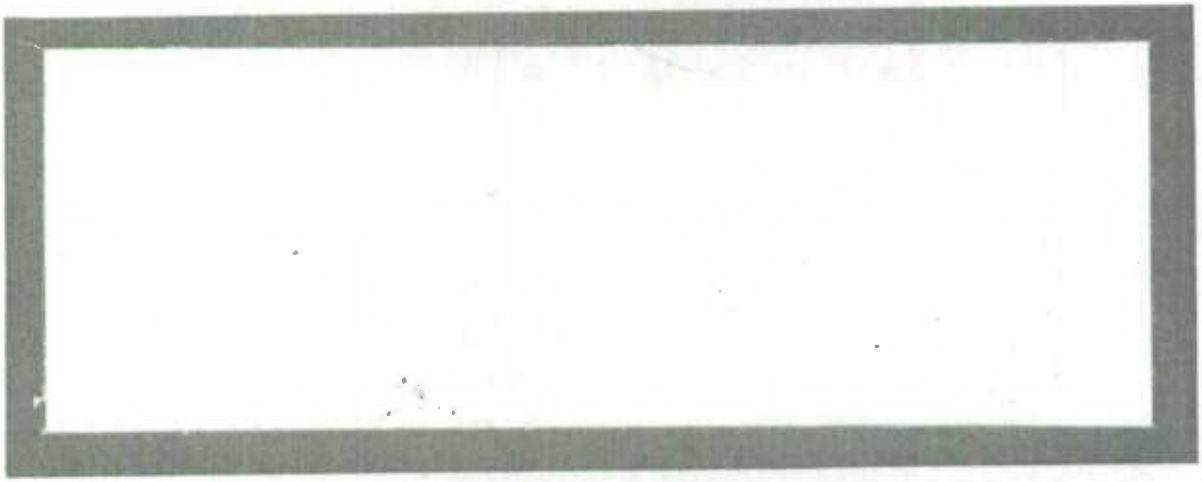
Statistics Canada Statistique Canada



Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

Canada

EXPLORATORY ANALYSES PERFORMED ON THE COMBINED.MASTER FILE

M.A. Hidiroglou

Working Paper No. BSMD 85-003E

Exploratory Analyses Performed

on the Combined.Master File

(Tax Year 1981)

M. Hidiroglou
BSMD
April 1984

EXPLORATORY ANALYSES PERFORMED ON THE COMBINED.MASTER FILE

(TAX YEAR 1981)

Introduction

The four digit SIC codes on the Combined.Master were regrouped into 18 major divisions and 76 major groups. Records which had been included in the prespecified sample and non-birth cross-sectional sample of the T1 returns were analyzed at different levels of aggregation controlling for province, major division and major group. Records within the $25,000 to $500,000 Gross Business Income range were analyzed with the purpose of modelling non-zero wages and salaries to other variables common to the COMSCREEN and Combined.Master files.

a. Transformation of Data

Previous analyses indicated that Gross Business Income was the most strongly correlated variable with non-zero wages and salaries across the majority of major divisions and major groups.

Scatter plots of non-zero wages and salaries versus Gross Business Income indicated that the spread of non-zero wages and salaries increased as Gross Business Income increased. This indicated that a transformation of the data would be required. Since Gross Business Income was the most

correlated variable with non-zero wages and salaries, five models were tried using this auxiliary variable. The models were fitted at the Canada level by Major Division. The models were:

(i) $SALWAG = INT1 + SLOPE1 * GBI + E1$

(ii) $SALWAG = SLOPE2 * GBI + E2$

(iii) $SALWAG/\sqrt{GBI} = INT3/\sqrt{GBI} + SLOPE3 * \sqrt{GBI} + E3$

(iv) $SALWAB/\sqrt{GBI} = SLOPE4 * \sqrt{GBI} + E4$

(v) $LOG(SALWAG) = INT5 + SLOPE5 * LOG(GBI) + E5$

where SALWAG = wages and salaries,

GBI = Gross Business Income,

INT = intercept,

SLOPE = slope of the regression.

Examination of the standardized residuals and the adjusted $\bar{R}^2_p$ term, indicated that the square root transformation was the best. Furthermore, the intercept term was not sufficiently significant to include in the model. (See Appendix I).

b. Searching for fits

Scatter plots of the ratio of the mean of non-zero wages and salaries to the mean of the gross business income within selected intervals of the gross business income were obtained to determine whether these ratios were constant over the gross business income intervals or whether they

increased or decreased over the intervals or whether there existed breaks between which these ratios were constant. These scatter plots which were done at the major division by province cross-classification indicated that a mixture of these conditions could exist depending upon the major division and provincial cross-classification. (See Appendix 2)

As a result of the above scatter plots, eight regression models were fitted to reflect if the conclusions drawn from the scatter plots held. These models were

i) $SALWAG/\sqrt{GBI}$ vs. $\sqrt{GBI}$ = Linear

ii) $SALWAG/\sqrt{GBI}$ vs. $\sqrt{GBI}$, $(GBI)^{3/2}$ = Linear & Quadratic

iii) $SALWAG/\sqrt{GBI}$ vs. $(GBI)^{3/2}$ = Quadratic

iv) $SALWAG/\sqrt{GBI}$ vs. $\sqrt{GBI}$ = Linear, break at \$100K Gross Business Income

v)-viii) Models i) through iv) with provinces added as dummy variables to test if the fits differed by province.

Dummy variables defined as zero, one with the zero indicating that the observation does not belong to the group and one if it belongs to the group, were used to construct models v) - viii). (See Appendix 3).

ie. $Y(i) = \sum_j b(j)\, d(i,j).\, x(i) + \text{Error}$ where

$d(i,j) = 1$ if the i^{th} observation belongs to the j^{th} dummy group

0 otherwise

The results of the above fits are summarized in the following table. Note that major division was chosen as the classification variable for evaluation of the above models.

TABLE 1: Summary of the Fits by Major Division

Major Division	Best Fit at the Canada Level	Provinces Significantly Different	R^2	Number of Observations
1. Logging & Forestry	Linear*	Yes*	0.74	980
2. Mining	Linear*	No	0.62	88
3. Manufacturing	Linear*	Yes*	0.69	2,987
4. Construction	Linear*	Yes*	0.64	12,585
5. Transportation	Linear*	Yes*	0.57	3,992
6. Communication	Linear*	No	0.72	248
7. Wholesale	Linear & Quadratic*	Yes*	0.46	1,396
8. Retail	Linear & Quadratic*	Yes*	0.51	18,545
9. Finance & Insurance	Break at $100K*	No	0.83	15
10. Real Estate	Linear*	Yes*	0.52	303
11. Business Service	Linear*	Yes*	0.64	534
12. Educational Service	Linear & Quadratic*	Yes**	0.74	91
13. Health and Social	Linear*	Yes**	0.70	378
14. Accommodation	Linear & Quadratic	Yes*	0.76	6,005
15. Other Services	Linear*	Yes*	0.62	5,127

*: Significant at the 1% level

**: Significant at the 5% level

From the above table, one concludes that for the most part, the fits are linear within each major division and that they differ in slope between provinces in the majority of the cases. For those major divisions which have a combination of linear and quadratic terms, although the addition of the quadratic term is statistically significant, the adjusted coefficient of determination (R^2_D) is increased only slightly.

The linear model $SALWAG/\sqrt{GBI}$ versus $\sqrt{GBI}$ was fitted by major group at the Canada level with and without the provinces used as dummy variables. The disaggregation of major division into major groups does not significantly improve the fits as can be observed from Table 2. An average coefficient of determination for the major groups within the major divisions weighted by the number of observations within each group was computed and compared to the corresponding coefficient of determination obtained within the major divisions to assess the overall effect of disaggregation. (See appendix 4).

TABLE 2: COMPARISON OF WEIGHTED COEFFICIENT OF DETERMINATION FOR MAJOR GROUPS WITHIN MAJOR DIVISIONS TO THE COEFFICIENT OF DETERMINATION WITHIN MAJOR DIVISIONS
- Provinces used as dummy variables -

Major Division	Weighted R^2_p for Groups within Divisions	R^2_p for Divisions
1. Logging and Forestry**	0.75	0.74
2. Mining	0.69	0.60
3. Manufacturing	0.72	0.69
4. Construction	0.65	0.65
5. Transportation **	0.57	0.57
6. Communication **	0.75	0.73
7. Wholesale	0.48	0.43
8. Retail	0.53	0.50
9. Finance and Insurance**	0.44	0.42
10. Real Estate**	0.55	0.52
11. Business Service*	0.64	0.64
12. Educational Service*	0.72	0.72
13. Health and Social*	0.70	0.70
14. Accommodation	0.74	0.74
15. Other Services	0.64	0.62

Note: * indicates one major group/major division

** indicates two major groups/major division

For major divisions such as Wholesale and Retail, where the R^2_p is not very high, 0.43 and 0.50 respectively, the corresponding rate of the R^2_p for major groups within these major divisions is (0.26-0.51) and (0.40-0.62).

The auxiliary variable Gross Business Income (GBI) was the most highly correlated variable with Wages and Salaries for most major divisions and major groups and their cross-classification with provinces. A number of other variables common to COMSCREEN and to the Combined.Master files were entered into the regression provided that they were sufficiently correlated with Wages and Salaries. The best fits using these other variables and Gross Business Income were found using a stepwise regression procedure. These fits were done for major divisions, and major groups both at the Canada and provincial level. The selected auxiliary variables were

DEPR - DEPRECIATION

BSNT - NET PROFIT

GROSPRO - GROSS PROFIT

PARTNER - A ZERO ONE VARIABLE TO INDICATE ABSENCE OR PRESENCE OF PARTNERSHIP

GBI - GROSS BUSINESS INCOME

PROFGRS - GROSS PROFESSIONAL INCOME

QGBI - THE SQUARE OF GROSS BUSINESS INCOME.

Table 3, summarizes these fits for major divisions at the CANADA level. Note the square root transformation was applied to all those fits and that the form of the models fitted was:

SALWAG/$\sqrt{\text{GBI}}$ VS. ABOVE VARIABLES DIVIDED BY $\sqrt{\text{GBI}}$

TABLE 3: Best Model at the Major Division Level using several Auxiliary variables

Major Division	Number of Observations	Best one variable Model	R^2_p	Associated Best Model Form	R^2_p
1. Logging & Forestry	981	GBI	0.71	GBI,DEPR,BSNT, QGBI	0.76
2. Mining	88	GBI	0.62	GBI,BSNT	0.72
3. Manufacturing	2,990	GBI	0.68	GBI,BSNT	0.70
4. Construction	12,587	GBI	0.64	GBI,DEPR,BSNT	0.66
5. Transportation	3,995	GBI	0.57	GBI,DEPR,BSNT	0.62
6. Communication	248	GBI	0.72	GBI,DEPR,BSNT	0.77
7. Wholesale	1,399	GBI	0.43	GBI,DEPR,GROSPRO	0.49
8. Retail	18,551	GROSPRO	0.58	GBI,BSNT,GROSPRO	0.60
9. Finance & Insurance	11	GBI	0.44	GBI	0.44
10. Real EState	303	GBI	0.48	GBI	0.48
11. Business Service	534	GBI	0.62	GBI,BSNT	0.68
12. Educational Service	91	GBI	0.71	GBI,DEPR,BSNT	0.78
13. Health and Social	378	GBI	0.69	GBI,DEPR,BSNT	0.81
14. Other Services	5,129	GBI	0.62	GBI,DEPR,BSNT, QGBI	0.64
15. Accommodation	6,009	GBI	0.73	GBI,DEPR,BSNT, QGBI	0.80

From Table 3, one observes that the addition of more auxiliary variables does not significantly improve the fit once the most important auxiliary variable has been taken into account. There are a few exceptions to this

rule such as the Health and Social major division where the R^2_p increases from 0.69 to 0.81. Note, however, that this is a major division that does contain many observations. Generally speaking, the fit does not improve significantly when there are a large number of observations associated with the major divisions. Similar conclusions to those reached above hold when the fits are carried out at the major group level. (For more detailed information see Appendix 5).

A summary of the best one-variable and several variable fit for provinces within each major division is provided in Table 4. The best one-variable fit contains the most highly correlated auxiliary variable with Wages and Salaries while the best several variable fit contains several variables which have improved the fit. Since too many tables would be provided if all the provinces were included individually, a range of the R^2_p is given for provinces within each major division, which summarizes the information contained in Appendix 6.

TABLE 4: Range of R^2_p from the best one-variable fit and several variable fit over provinces within major divisions

Major Division	One Variable Fit	Several Variables Fit
1. Logging and Forestry	(0.69, 0.87)	(0.72, 0.89)
2. Mining	(0.54, 0.73)	(0.67, 0.85)
3. Manufacturing	(0.63, 0.78)	(0.63, 0.79)
4. Construction	(0.57, 0.71)	(0.61, 0.73)
5. Transportation	(0.46, 0.63)	(0.46, 0.72)
6. Communication	(0.63, 0.87)	(0.63, 0.87)
7. Wholesale	(0.42, 0.72)	(0.46, 0.72)
8. Retail	(0.34, 0.64)	(0.36, 0.66)
9. Finance & Insurance	(0.44, 0.44)	(0.44, 0.44)
10. Real Estate	(0.36, 0.73)	(0.46, 0.73)
11. Business Service	(0.54, 0.77)	(0.54, 0.77)
12. Educational Service	(0.72, 0.79)	(0.77, 0.78)
13. Health and Social	(0.67, 0.81)	(0.79, 0.89)
14. Accommodation	(0.58, 0.83)	(0.63, 0.87)
15. Other Services	(0.55, 0.67)	(0.58, 0.73)

Note: Cells with less than 10 observations excluded from the range.

From Table 4, one observes that the disaggregation of the fits for major divisions from the Canada level to the provincial level improves R^2_p for some provinces in some instances and worsens R^2_p in others. The range of the R^2_p for both the one-variable and several variable fit behaves like the R^2_p given in Table 3 for the corresponding major divisions. The lower and upper limits of the range are slightly increased so one moved from the one-variable to the several variable fit, indicating that the addition of auxiliary variable does not significantly improve the one-variable fit.

Some of the major divisions such as Mining, Finance and Insurance, and Educational Service do not contain enough data to be disaggregated from the Canada to the provincial level. For these major divisions, the fits should be done at the Canada level.

Similar conclusions to the above are reached for major groups. The disaggregation of the fits for major groups from the Canada level to the provincial level does not substantially improve the fits. For major groups, the fits should be done at the provincial level whenever there are enough observations and at the Canada level otherwise.

Correlations between the sum of salaries and wages, depreciation and net profit before taxes were in some instances higher than the same correlation between salaries and wages to gross business income. This seemed to indicate that a model using the sum of these variables as the dependent variable versus gross business income as the independent variable could be superior to a model using salaries and wages as the dependent variable versus gross business income, depreciation and net profit as the independent variables. These two models were compared using the residual sum of squares and the conclusion was that the better model was the one which allowed the additional variables to enter into the regression as independent variables. (See Appendix 7).

Conclusions

The four digit SIC codes on the Combined.Master were regrouped into 18 major divisions and 76 major groups. Records which had been included in the prespecified sample and non-birth cross-sectional sample of the T1 returns were analyzed at different levels of aggregation controlling for province, major division and major group. Records within the $25,000 to $500,000 Gross Business Income range were analyzed with the purpose of modelling non-zero Wages and Salaries to other variables common to the COMSCREEN and Combined.Master files.

The following conclusions can be drawn concerning the modelling of Wages and Salaries:

a) For most major divisions (or major groups), the proportion of filers showing non-zero Wages and Salaries increases as the Gross Business Income increases;

b) Scatter plots of Wages and Salaries versus Gross Business Income indicated that the data must be transformed: a square root transformation of the Gross Business Income is sufficient.

c) Gross Business Income is the most highly corelated variables with Wages and Salaries (non-zero) for most major divisions, with the exception of Wholesale and Retail where gross profit is the most highly correlated

variable. The addition of variables other than Gross Business Income or Gross Profit does not significantly improve the regression fit.

d) The modelling should be done at the provincial level by major group level if there are enough observations. Otherwise, collapsing over similar major groups and provinces should occur to provide the necessary number of observations.

APPENDIX 1

Regression for T1 Sample

(Total Salaries and Wages versus

by Major Division (Tax Year 1981)

Across Canada

GBI: $25,000 - $500,000

SMC = 0 or 1 (Old Records)

116,223 records

Model	Equation	
1	SALWAG vs. Intercept,	GBI
2	SALWAG vs	GBI
3	SALWAG/$\sqrt{GBI}$ vs. 1/$\sqrt{GBI}$,	$\sqrt{GBI}$
4	SALWAG/$\sqrt{GBI}$ VS.	$\sqrt{GBI}$

SALWAG - TOTAL WAGES AND SALARIES

GBI - Gross Business Income

Agriculture

Model	N	F	R^2	Int.	Slope	t(Int)	t(Slope)
1	266	177.4	0.40	1.87	0.17	1.04***	13.32*
2	267	480.72	0.64		0.18		21.93*
3	267	213.97	0.62	-0.31	0.19	-0.26***	13.31*
4	267	429.37	0.62		0.19		20.72*
				Fishing			
1	13	6.23	0.30	-10.04	0.55	-9.22***	2.47**
2	14	19.25	0.62		0.36		4.37*
3	14	10.09	0.62	-11.53	0.59	- 1.30***	2.79**
4	14	17.51	0.59		0.33		4.18*
				Logging and Forestry			
1	981	1458.49	0.60	-7.07	0.31	-5.66*	38.19*
2	982	2809.49	0.75		0.28		53.0*
3	982	1314.53	0.73	-5.84	0.30	-8.11*	38.48*
4	982	2404.52	0.71		0.25		49.03*
				Mining			
1	90	54.47	0.38	-0.39	0.20	-0.12***	7.38*
2	89	184.55	0.68		0.20		13.59
3	89	71.26	0.62	-1.56	0.22	-0.59***	6.86*
4	89	143.25	0.62		0.20		11.97*

Manufacturing

Model	N	F	R^2	Int.	Slope	t(Int)	t(Slope)
1	2990	2350.30	0.44	-0.9	0.21	-1.44***	48.48*
2	2991	6409.8	0.69		0.21		80.06*
3	2991	3266.02	0.69	-1.56	0.22	-3.94*	50.99*
4	2991	6485.06	0.68		0.20		80.53*

Construction

1	12587	10060.85	0.44	-0.78	0.19	-3.52*	100.30*
2	12588	25230.77	0.67		0.19		158.84*
3	12588	11523.03	0.65	-1.27	0.20	-8.1*	94.20*
4	12588	22861.32	0.64		0.18		151.2*

Transportation

1	3995	2920.42	0.42	-5.03	0.21	-11.26*	54.04*
2	3996	6054.35	0.60		0.17		77.81*
3	3996	2709.37	0.58	-2.40	0.18	-7.73*	46.51*
4	3996	5281.20	0.57		0.16		72.67*

Communication

1	248	327.44	0.57	0.20	0.22	0.14***	18.1*
2	249	821.19	0.77		0.22		26.66*
3	249	314.47	0.72	-0.77	0.23	-0.72***	15.56*
4	249	629.66	0.72		0.22		25.09*

Model	N	F	R^2	Int.	Slope	t(Int)	t(Slope)
Wholesale							
1	1399	188.04	0.12	4.74	0.04	7.34*	13.71*
2	1400	1055.33	0.43		0.06		32.49*
3	1400	581.02	0.45	3.30	0.05	8.01*	15.67*
4	1400	1050.54	0.43		0.07		32.41*
Retail							
1	18551	3992.37	0.18	3.75	0.06	21.17*	63.19*
2	18552	19623.79	0.51		0.07		140.09*
3	18552	9727.62	0.51	2.76	0.06	24.26	68.93*
4	18552	18287.69	0.50		0.08		135.23*
Finance and Insurance							
1	15	0.91	-0.006	4.87	0.01	2.27**	0.95***
2	16	12.83	0.48		0.04		3.58*
3	16	16.42	0.69	5.52	0.007	3.54*	0.43***
4	16	11.16	0.44		0.05		3.34*
Real Estate							
1	303	75.92	0.20	2.37	0.14	1.10***	8.71*
2	304	221.21	0.42		0.15		14.87*
3	304	138.63	0.48	1.71	0.14	1.30***	9.08*
4	304	274.89	0.48		0.16		16.58*

Business Service

Model	N	F	R^2	Int.	Slope	t(Int)	t(Slope)
1	534	508.48	0.49	-2.54	0.30	-1.43***	22.55*
2	535	977.21	0.65		0.28		31.26*
3	535	459.83	0.63	-3.72	0.31	-3.59*	22.29*
4	535	887.04	0.62		0.27		29.78*

Educational Service

1	91	214.91	0.70	-10.58	0.43	-5.4	14.66*
2	92	339.38	0.79		0.35		18.42*
3	92	125.03	0.73	-6.92	0.38	-2.90*	10.95*
4	92	223.22	0.71		0.30		14.94*

Health and Social

1	378	390.26	0.58	-5.99	0.22	-15.08*	90.94*
2	379	830.52	0.69		0.32		28.82*
3	379	445.22	0.70	-4.88	0.35	-3.39*	19.91*
4	379	885.05	0.69		0.30		29.24*

Accommodation

1	6009	8270.04	0.58	-5.99	0.22	-15.08*	90.94*
2	6010	17626.99	0.75		0.19		132.77*
3	6010	8669.78	0.74	-3.90	0.20	-16.59*	92.34*
4	6010	16319.87	0.73		0.18		127.75*

Other Services

Model	N	F	R^2	Int.	Slope	t(Int)	t(Slope)
1	5129	2209.13	0.30	3.37	0.19	8.21	47.0*
2	5130	6745.57	0.57		0.21		82.13*
3	5130	4151.95	0.62	0.48	0.22	1.63***	49.89*
4	5130	8298.53	0.62		0.23		91.10*
Unproperly Classified							
1	2076	599.20	0.22	2.28	0.13	3.02*	24.48*
2	2077	1829.88	0.47		0.15		42.77*
3	2077	1038.46	0.50	1.07	0.14	2.32**	26.50*
4	2077	2067.17	0.50		0.15		45.47*

*: Significant at the 1% level

**: Significant at the 1 - 5% level

*** Significant to at least the 5% level

Appendix 2

Summary of Plots of Wages and Salaries

Versus Gross Business Income (T1 Sample)

SMC = 0 or 1 (Old Records)

by Major Division and Province

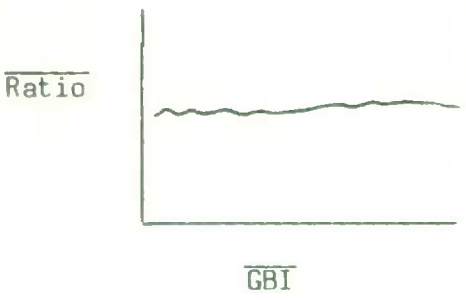
GBI: $25,000 - $500,000

116,623 records

Intervals of GBI

1. $ 25K - $ 50K
2. $ 50K - $100K
3. $100K - $150K
4. $150K - $200K
5. $200K - $250K
6. $250K - $300K
7. $300K - $350K
8. $350K - $400K
9. $400K - $450K
10. $450K - $500K

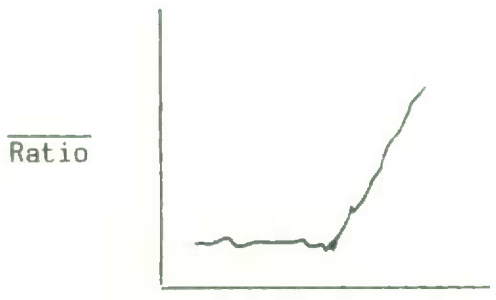
Shapes of Plots



Linear

Quadratic Upward

Quadratic Downward



Break

$\overline{GBI}$ - means of GBI within above intervals

$\overline{Ratio}$ - means of SALWAG/GBI

Graph of Average SALWAG/GBI for selected intervals of GBI
Relationship between SALWAG and GBI

Major Division	Province			
	Alberta	British Columbia	Manitoba	New Brunswick
Agriculture	Linear	Linear	Linear	Linear
Fishing		**Not Enough Points**		
Logging & Forestry	Quadratic Upward	Quadratic Upward	Linear	Linear
Mining		**Not Enough Points**		
Manufacturing	Quadratic Upward	Linear	Quadratic Upward	Linear
Construction	Linear	Linear	Linear	Linear
Transportation	Linear	Linear	Linear	Linear
Communication		**Not Enough Points**		
Wholesale		**Not Enough Points**		
Retail	Quadratic Downward	Quadratic Downward	Quadratic Downward	Quadratic Downward
Finance & Insurance		**Not Enough Points**		
Real Estate	Linear	Linear	Linear	-

Graph of Average SALWAG/GBI for selected intervals of GBI
Relationship between SALWAG and GBI

Major Division	Province			
	Newfoundland	Nova Scotia	N. W. Territories	Ontario
Agriculture	-	-	-	
Fishing		**Not Enough Points**		
Logging & Forestry	Linear	Quadratic Upward	-	Linear
Mining		**Not Enough Points**		
Manufacturing	Linear	Linear	-	Linear
Construction	Quadratic Downward	Quadratic Downward	-	Linear
Transportation	Linear	Linear	-	Linear
Communication		**Not Enough Points**		
Wholesale		**Not Enough Points**		
Retail	Linear	Quadratic Downward	Linear	Quadratic Downward
Finance & Insurance		**Not Enough Points**		
Real Estate	-	-	-	Linear

Graph of Average SALWAG/GBI for selected intervals of GBI
Relationship between SALWAG and GBI

Major Division	Province			
	Prince Edward Island	Quebec	Saskatchewan	Yukon
Agriculture	-	Linear	Quadratic Downward	-
Fishing		**Not Enough Points**		
Logging & Forestry	-	Linear	Linear	-
Mining		**Not Enough Points**		
Manufacturing	Linear	Linear	Linear	-
Construction	Linear	Linear	Quadratic Downward	Linear
Transportation	Linear	Linear	Linear	Quadratic Upward
Communication		**Not Enough Points**		
Wholesale		**Not Enough Points**		
Retail	Quadratic Downward	Quadratic Downward	Quadratic Downward	Linear
Finance & Insurance		**Not Enough Points**		
Real Estate	-	Linear	Linear	-

Graph of Average SALWAG/GBI for selected intervals of GBI
Relationship between SALWAG and GBI

Major Division	Province			
	Alberta	British Columbia	Manitoba	New Brunswick
Business Survice	Linear	Quadratic Upward	Linear	Linear
Educational Service		**Not Enough Points**		
Health & Social		**Not Enough Points**		
Accommodation	Quadratic Upward	Quadratic Upward	Linear	Quadratic Upward
Other Services	Linear	Linear	Linear	Linear

Graph of Average SALWAG/GBI for selected intervals of GBI
Relationship between SALWAG and GBI

Major Division	Province			
	Newfoundland	Nova-Scotia	N.W. Territories	Ontario
Business Survice	Linear	Linear	-	Linear
Educational Service		**Not Enough Points**		
Health & Social		**Not Enough Points**		
Accommodation	Linear	Quadratic Upward	Linear	Quadratic Upward
Other Services	Quadratic Downward	Quadratic Downward	-	Quadratic Downward

Graph of Average SALWAF/GBI for selected intervals of GBI
Relationship between SALWAG and GBI

Major Division	Province			
	Prince Edward Island	Quebec	Saskatchewan	Yukon
Business Survice	-	Linear	Linear	-
Educational Service		**Not Enough Points**		
Health & Social		**Not Enough Points**		
Accommodation	Quadratic Upward	Quadratic Upward	Linear	Linear
Other Services	Linear	Linear	Quadratic Downward	Linear

Appendix 3

Resume of regression

of 8 models

With COMBINED.MASTER

(Tax Year 1981)

SMC = 0 or 1 (Old Records)

GBI: $25,000 - $500,000

116,223 records

Model

1	LINEAR	: SALWAG/$\sqrt{GBI}$ vs $\sqrt{GBI}$
2	LINEAR & QUADRATIC	: SALWAG/$\sqrt{GBI}$ vs $\sqrt{GBI}$ $(GBI)^{3/2}$
2A	QUADRATIC	: SALWAG/$\sqrt{GBI}$ vs $(GBI)^{3/2}$
3	LINEAR WITH BREAK	: SALWAG/$\sqrt{GBI}$ vs $\sqrt{GBI}$ break at $100K
4	LINEAR WITH PROVINCE	: MODEL 1, Provinces as dummy variables
5	LINEAR & QUADRATIC PROV.	: MODEL 2, Provinces as dummy variables
5A	QUADRATIC PER PROVINCE	: MODEL 2A, Provinces as dummy variables
6	LINEAR PER PROV. WITH BREAK:	MODEL 3, Provinces as dummy variables

Tolerance: = 2 standard deviations to exclude oubliers from model.

DF : degrees of freedom from residual sums of squares

SSQ : residual sums of squares

$\bar{R}^2_p$: adjusted coefficient of determination

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(3) Logging and Forestry	EQ1: LINEAR	WITH	979	3045.6	0.7103
		EXCLUD.	937	1756.7	0.758
	EQ2: LIN. + QUAD.	WITH	978	2885.4	0.725
		EXCLUD.	927	1608.1	0.779
	EQ2A: QUAD.	WITH	979	4144.1	0.606
		EXCLUD.	924	2359.8	0.690
	EQ3: LIN. BREAK	WITH	978	2874.4	0.726
		EXCLUD.	931	1654.5	0.776
	EQ4: LIN. PROV.	WITH	970	2703.7	0.741
		EXCLUD.	933	1532.7	0.808
	EQ5: PROV. LIN. QUAD	WITH	960	2525.6	0.755
		EXCLUD.	910	1302.3	0.830
	EQ5A: PROV. QUAD.	WITH	970	3788.5	0.636
		EXCLUD.	910	2084.3	0.724
	EQ6: PROV. LIN. BRK.	WITH	961	2550.9	0.753
		EXCLUD.	923	1416.0	0.821

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(4) Mining	EQ1: LINEAR	WITH	87	212.5	0.622
		EXCLUD.	83	108.3	0.710
	EQ2: LIN. + QUAD.	WITH	86	212.5	0.618
		EXCLUD.	82	106.9	0.710
	EQ2A: QUAD.	WITH	87	293.8	0.478
		EXCLUD.	82	139.6	0.624
	EQ3: LIN. BREAK	WITH	86	211.6	0.629
		EXCLUD.	82	104.8	0.716
	EQ4: LIN. PROV.	WITH	78	201.4	0.601
		EXCLUD.	76	121.1	0.686
	EQ5: PROV. LIN. QUAD	WITH	69	184.4	0.587
		EXCLUD.	67	105.9	0.688
	EQ5A: PROV. QUAD.	WITH	78	257.8	0.489
		EXCLUD.	74	124.03	0.661
	EQ6: PROV. LIN. BRK.	WITH	69	180.0	0.597
		EXCLUD.	66	95.1	0.696

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(6) Construction	EQ1: LINEAR	WITH	12,584	21,339.2	0.645
		EXCLUD.	12,008	12,377.7	0.698
	EQ2: LIN. + QUAD.	WITH	12,583	21,314.9	0.645
		EXCLUD.	11,993	12,301.7	0.700
	EQ2A: QUAD.	WITH	12,584	33,226.7	0.447
		EXCLUD.	11,849	18,347.9	0.555
	EQ3: LIN. BREAK	WITH	12,583	21,264.5	0.646
		EXCLUD.	11,983	12.253.9	0.701
	EQ4: LIN. PROV.	WITH	12,573	21,144.8	0.648
		EXCLUD.	11,992	12,143.9	0.702
	EQ5: PROV. LIN. QUAD	WITH	12,561	21,066.0	0.649
		EXCLUD.	11,968	12,016.7	0.706
	EQ5A: PROV. QUAD.	WITH	12,573	32,745.5	0.455
		EXCLUD.	11,835	18,016.1	0.455
	EQ6: PROV. LIN. BRK.	WITH	12,561	21,034.2	0.650
		EXCLUD.	11,955	11,980.3.	0.707

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	R^2_p
(7) Transportation	EQ1: LINEAR	WITH	3,991	7,005.2	0.57
		EXCLUD.	3,836	3,733.8	0.63
	EQ2: LIN. + QUAD.	WITH	3,990	6,762.9	0.584
		EXCLUD.	3,799	3,453.9	0.655
	EQ2A: QUAD.	WITH	3,991	8,424.3	0.482
		EXCLUD.	3,751	3,929.0	0.595
	EQ3: LIN. BREAK	WITH	3,990	6,951.0	0.573
		EXCLUD.	3,827	3,672.3	0.634
	EQ4: LIN. PROV.	WITH	3,980	6,893.9	0.575
		EXCLUD.	3,824	3,693.0	0.633
	EQ5: PROV. LIN. QUAD	WITH	3,969	6,487.9	0.599
		EXCLUD.	3,779	8,143.0	0.660
	EQ5A: PROV. QUAD.	WITH	3,980	8,143.0	0.627
		EXCLUD.	3,742	3,860.9	0.627
	EQ6: PROV. LIN. BRK.	WITH	3,969	6,801.9	0.580
		EXCLUD.	3,807	3,632.7.	0.638

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\overline{R^2}_p$
(8) Communication	EQ1: LINEAR	WITH	247	432.4	0.718
		EXCLUD.	237	259.6	0.776
	EQ2: LIN. + QUAD.	WITH	246	432.4	0.717
		EXCLUD.	236	259.3	0.775
	EQ2A: QUAD.	WITH	247	814.0	0.470
		EXCLUD.	231	450.6	0.590
	EQ3: LIN. BREAK	WITH	246	431.6	0.718
		EXCLUD.	234	248.1	0.784
	EQ4: LIN. PROV.	WITH	238	404.2	0.727
		EXCLUD.	223	205.5	0.805
	EQ5: PROV. LIN. QUAD	WITH	228	373.0	0.737
		EXCLUD.	215	211.3	0.806
	EQ5A: PROV. QUAD.	WITH	238	717.3	0.515
		EXCLUD.	221	395.5	0.636
	EQ6: PROV. LIN. BRK.	WITH	229	380.7	0.732
		EXCLUD.	215	207.0	0.809

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(9) Wholesale	EQ1: LINEAR	WITH	1,395	1,630.5	0.43
		EXCLUD.	1,317	717.1	0.511
	EQ2: LIN. + QUAD.	WITH	1,394	1,544.1	0.459
		EXCLUD.	1,330	730.9	0.537
	EQ2A: QUAD.	WITH	1,395	2,146.9	0.249
		EXCLUD.	1,308	964.1	0.315
	EQ3: LIN. BREAK	WITH	1,394	1,585.9	0.445
		EXCLUD.	1,323	701.7	0.536
	EQ4: LIN. PROV.	WITH	1,386	1,607.8	0.434
		EXCLUD.	1,309	715.9	0.514
	EQ5: PROV. LIN. QUAD	WITH	1,376	1,517.0	0.461
		EXCLUD.	1,316	736.5	0.540
	EQ5A: PROV. QUAD.	WITH	1,386	2,129.1	0.250
		EXCLUD.	1,296	937.2.5	0.313
	EQ6: PROV. LIN. BRK.	WITH	1,376	1,554.9	0.448
		EXCLUD.	1,310	712.6	0.539

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(10) Retail	EQ1: LINEAR	WITH	18,544	18,877.7	0.496
		EXCLUD.	17,651	9,107.8	0.585
	EQ2: LIN. + QUAD.	WITH	18,543	18,254.8	0.513
		EXCLUD.	17,669	8,757.1	0.601
	EQ2A: QUAD.	WITH	18,544	25,369.6	0.323
		EXCLUD.	17,520	12.610.9	0.407
	EQ3: LIN. BREAK	WITH	18,543	18,469.6	0.507
		EXCLUD.	17,663	8,811.7	0.60
	EQ4: LIN. PROV.	WITH	18,533	18,741.8	0.50
		EXCLUD.	17,654	9,093.7	0.589
	EQ5: PROV. LIN. QUAD	WITH	18,521	18,090.0	0.517
		EXCLUD.	17,656	8,704.1	0.605
	EQ5A: PROV. QUAD.	WITH	18,533	25,274.2	0.325
		EXCLUD.	17,504	12,541.0	0.412
	EQ6: PROV. LIN. BRK.	WITH	18,521	18,311.0	0.511
		EXCLUD.	17,647	8,738.5	0.603

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(11) Finance and Insurance	EQ1: LINEAR	WITH	14	6.44	0.443
		EXCLUD.	13	4.6	0.436
	EQ2: LIN. + QUAD.	WITH	13	4.9	0.543
		EXCLUD.	12	2.8	0.723
	EQ2A: QUAD.	WITH	14	9.9	0.149
		EXCLUD.	13	6.8	0.160
	EQ3: LIN. BREAK	WITH	13	1.8	0.830
		EXCLUD.	12	0.97	0.909
	EQ4: LIN. PROV.	WITH	10	4.8	0.417
		EXCLUD.	10	4.8	0.441
	EQ5: PROV. LIN. QUAD	WITH	9	2.3	0.685
		EXCLUD.	9	2.3	0.700
	EQ5A: PROV. QUAD.	WITH	10	7.6	0.084
		EXCLUD.	10	7.6	-0.017
	EQ6: PROV. LIN. BRK.	WITH	9	1.4	0.814
		EXCLUD.	8	0.5	0.931

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(12) Real Estate	EQ1: LINEAR	WITH	302	894.3	0.477
		EXCLUD.	286	419.8	0.524
	EQ2: LIN. + QUAD.	WITH	301	888.6	0.478
		EXCLUD.	286	388.3	0.561
	EQ2A: QUAD.	WITH	302	1,222.5	0.284
		EXCLUD.	278	565.5	0.303
	EQ3: LIN. BREAK	WITH	301	890.5	0.477
		EXCLUD.	286	410.8	0.541
	EQ4: LIN. PROV.	WITH	293	790.9	0.523
		EXCLUD.	275	342.4	0.598
	EQ5: PROV. LIN. QUAD	WITH	285	746.9	0.537
		EXCLUD.	265	315.9	0.611
	EQ5A: PROV. QUAD.	WITH	293	1063.2	0.359
		EXCLUD.	268	456.4	0.407
	EQ6: PROV. LIN. BRK.	WITH	286	750.1	0.536
		EXCLUD.	268	337.6	0.611

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	R^2_p
(13) Business Service	EQ1: LINEAR	WITH	533	2,315.0	0.625
		EXCLUD.	510	1,257.2	0.682
	EQ2: LIN. + QUAD.	WITH	532	2,299.9	0.626
		EXCLUD.	506	1,201.0	0.695
	EQ2A: QUAD.	WITH	533	3,397.0	0.449
		EXCLUD.	501	1,738.8	0.554
	EQ3: LIN. BREAK	WITH	532	2,267.9	0.632
		EXCLUD.	505	1,194.0	0.694
	EQ4: LIN. PROV.	WITH	524	2,182.5	0.640
		EXCLUD.	502	1,200.3	0.696
	EQ5: PROV. LIN. QUAD	WITH	515	2,110.4	0.646
		EXCLUD.	489	1,078.3	0.715
	EQ5A: PROV. QUAD.	WITH	524	3,086.5	0.491
		EXCLUD.	491	1,533.4	0.599
	EQ6: PROV. LIN. BRK.	WITH	516	2,097.1	0.649
		EXCLUD.	487	1,057.6	0.719

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(15) Educational Service	EQ1: LINEAR	WITH	90	265.6	0.713
		EXCLUD.	86	196.0	0.697
	EQ2: LIN. + QUAD.	WITH	89	236.8	0.741
		EXCLUD.	86	199.4	0.742
	EQ2A: QUAD.	WITH	90	364.7	0.605
		EXCLUD.	86	269.0	0.646
	EQ3: LIN. BREAK	WITH	89	269.0	0.625
		EXCLUD.	89	251.9	0.625
	EQ4: LIN. PROV.	WITH	83	241.4	0.717
		EXCLUD.	79	170.6	0.739
	EQ5: PROV. LIN. QUAD	WITH	76	190.2	0.756
		EXCLUD.	72	124.7	0.812
	EQ5A: PROV. QUAD.	WITH	83	255.9	0.70
		EXCLUD.	77	151.3	0.770
	EQ6: PROV. LIN. BRK.	WITH	80	213.5	0.740
		EXCLUD.	76	148.8	0.788

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(16) Health and Social	EQ1: LINEAR	WITH	377	1,515.4	0.694
		EXCLUD.	362	960.8	0.729
	EQ2: LIN. + QUAD.	WITH	376	1,498.4	0.697
		EXCLUD.	359	933.8	0.732
	EQ2A: QUAD.	WITH	377	2,521.9	0.491
		EXCLUD.	357	1,487.5	0.623
	EQ3: LIN. BREAK	WITH	376	1,477.9	0.701
		EXCLUD.	359	914.3	0.738
	EQ4: LIN. PROV.	WITH	368	1,429.2	0.704
		EXCLUD.	352	874.4	0.743
	EQ5: PROV. LIN. QUAD	WITH	358	1,344.3	0.714
		EXCLUD.	342	841.3	0.756
	EQ5A: PROV. QUAD.	WITH	368	2,289.5	0.526
		EXCLUD.	348	1,266.0	0.668
	EQ6: PROV. LIN. BRK.	WITH	359	1,366.4	0.717
		EXCLUD.	344	845.9	0.761

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(17) Accommodation	EQ1: LINEAR	WITH	6,004	9,228.1	0.731
		EXCLUD.	5,835	5,005.6	0.806
	EQ2: LIN. + QUAD.	WITH	6,003	8,677.4	0.747
		EXCLUD.	5,793	4,443.3	0.831
	EQ2A: QUAD.	WITH	6,004	12,240.3	0.643
		EXCLUD.	5,770	6,728.5	0.750
	EQ3: LIN. BREAK	WITH	6,003	8,901.8	0.740
		EXCLUD.	5,825	4,738.4	0.818
	EQ4: LIN. PROV.	WITH	5,993	8,953.8	0.738
		EXCLUD.	5,818	4,823.7	0.813
	EQ5: PROV. LIN. QUAD	WITH	5,981	4,100.2	0.843
		EXCLUD.	5,759	4,100.2	0.843
	EQ5A: PROV. QUAD.	WITH	5,993	11,736.7	0.657
		EXCLUD.	5,758	6,447.3	0.763
	EQ6: PROV. LIN. BRK.	WITH	5,981	8,514.3	0.751
		EXCLUD.	5,774	4,383.9	0.831

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(18) Other Services	EQ1: LINEAR	WITH	5,126	13,205.7	0.618
		EXCLUD.	4,884	8,127.3	0.668
	EQ2: LIN. + QUAD.	WITH	5,125	13,014.1	0.624
		EXCLUD.	4,926	8,225.6	0.664
	EQ2A: QUAD.	WITH	5,126	23,652.6	0.316
		EXCLUD.	4,855	13,842.0	0.440
	EQ3: LIN. BREAK	WITH	5,125	13,189.3	0.618
		EXCLUD.	4,886	8,069.54	0.670
	EQ4: LIN. PROV.	WITH	5,115	13,050.5	0.622
		EXCLUD.	4,866	7,885.8	0.675
	EQ5: PROV. LIN. QUAD	WITH	5,103	12,808.1	0.628
		EXCLUD.	4,888	7,937.1	0.672
	EQ5A: PROV. QUAD.	WITH	5,115	23,512.5	0.318
		EXCLUD.	4,852	13,754.2	0.446
	EQ6: PROV. LIN. BRK.	WITH	5,104	12,967.4	0.623
		EXCLUD.	4,859	7,805.09	0.677

TOLERANCE
$\pm$ 2.0

DIVI	MODEL	OUTLIERS	DF	SSQ	$\bar{R}^2_p$
(99) Unproperly Classified	EQ1: LINEAR	WITH	2,072	5,370.5	0.499
		EXCLUD.	1,979	2,515.1	0.580
	EQ2: LIN. + QUAD.	WITH	2,071	5,334.5	0.502
		EXCLUD.	1,984	2,471.3	0.592
	EQ2A: QUAD.	WITH	2,072	7,351.2	0.314
		EXCLUD.	1,950	3,540.6	0.40
	EQ3: LIN. BREAK	WITH	2,071	5,346.4	0.501
		EXCLUD.	1,985	2,536.8	0.583
	EQ4: LIN. PROV.	WITH	2,062	5,274.9	0.505
		EXCLUD.	1,971	2,509.9	0.586
	EQ5: PROV. LIN. QUAD	WITH	2,051	5,141.2	0.604
		EXCLUD.	1,960	2,379.7	0.604
	EQ5A: PROV. QUAD.	WITH	2,062	7,235.9	0.322
		EXCLUD.	1,940	3,474.0	0.420
	EQ6: PROV. LIN. BRK.	WITH	2,051	5,196.5	0.510
		EXCLUD.	1,963	2,448.6	0.595

TOLERANCE
$\pm$ 2.0

Appendix 4

Regressions for Old Records

(Total Salaries and Wages)

SMC = 0 or 1

by Major Group (Tax Year 1981)

Across Canada and by Province (Dummy)

GBI: $25,000 - $500,000

116,223 records

$$\frac{SALWAG}{\sqrt{GBI}} \quad Y\phi \sqrt{GBI}$$

		N	CANADA R^{-2}_{can}	PROVINCE R^{-2}_{prov}
A	Agricultural Ind. 01	143	0.65	0.70
	Service to Agric. 02	124	0.59	0.62
B	Fishing and Trapping 03	14	0.59	0.54
C	Logging 04	964	0.72	0.75
	Forestry Service 05	18	0.62	0.61
D	Mining 06	6	0.44	0.30
	Crude Petroleum 07	7	0.55	0.72
	Quarry and Sand 08	34	0.75	0.71
	Service to Mineral 09	45	0.73	0.72
E.	Food Ind. 10	616	0.70	0.71
	Beverage Ind. 11	7	0.89	0.82
	Plastic Prod. 16	20	0.83	0.78
	Leather and Prod. 17	27	0.62	0.68
	Prim. Textile 18	27	0.62	0.68
	Textile Prod. 19	66	0.62	0.68
	Clothing Ind. 24	130	0.63	0.62
	Wood Ind.	377	0.75	0.75

See footnote at end of table

		N	CANADA R^{-2}_{can}	PROVINCE R^{-2}_{prov}
E.	Furniture and Fixt. 26	336	0.77	0.78
	PAPER AND Prod. 27	4	0.84	0.90
	Print and Publ. 28	466	0.74	0.74
	Primary Metal 29	14	0.78	0.81
	Fab Metal Prod. 30	299	0.71	0.71
	Machinery Ind. 31	16	0.47	0.45
	Transp. Equip. Ind. 32	77	0.69	0.69
	Electrical and Electronic 33	17	0.75	0.77
	Non-Metallic Prod. 35	80	0.74	0.72
	Chemical and Chemical Product 37	5	0.92	0.97
	Other Manuf. Inf. 39	426	0.68	0.69
F.	Building, Developing 40	803	0.60	0.61
	Ind'l and Heavy Constn. 41	110	0.73	0.76
	Trade Contracting 42	10,288	0.65	0.65
	Service Ind. 44	1,385	0.66	0.66
G.	Transportation Ind. 45	3,969	0.57	0.57

See footnote at end of table

		N	CANADA R^{-2}_{can}	PROVINCE R^{-2}_{prov}
	Storage and Warehousing 47	11	0.58	0.87
H	Communication Ind. 48	114	0.72	0.73
	Other Utility Ind. 49	134	0.74	0.77
I	Farm Prod. Ind. Wholesale 50	58	0.31	0.26
	Petroleum Prod. Ind. Wholesale 51	189	0.44	0.51
	Food, Bev., Drugs, Tobacco Wholesale 52	220	0.34	0.33
	Apparel and Dry Goods Wholesale 53	41	0.62	0.61
	Household Goods Wholesale 54	12	0.59	0.58
	Motor Vehicle, Parts Wholesale 55	39	0.51	0.51
	Metals, Hardware, Plumbing Wholesale 56	171	0.50	0.48
	Machinery Equip. Wholesale 57	248	0.54	0.61
	Other Prod. Ind. Wholesale 59	419	0.43	0.47
J.	Food, Bev., Drugs Retail 60	4,644	0.54	0.55

See footnote at end of table

		N	CANADA R^{-2}_{can}	PROVINCE R^{-2}_{prov}
	Shoe Apparel, Fab. Retail 61	1,570	0.53	0.53
	Household Furniture Retail 62	1,241	0.53	0.54
	Automotive Vehicles Retail 63	4,529	0.50	0.50
	General Merch. Retail 64	980	0.52	0.52
	Other Stores Retail 65	3,517	0.62	0.62
	Non-Stores Retail 69	2,066	0.39	0.40
K.	Invest. Inter. 72	5	0.65	0.82
	Other Financial Inter. 74	10	0.35	0.25
L.	Real estate Ind. 75	127	0.40	0.39
	Ins. and Real Estate 76	176	0.63	0.67
M.	Business Service Ind. 77	534	0.62	0.64
O.	Educational Service Ind. 85	91	0.71	0.72
P.	Health and Social Ind. 86	378	0.69	0.70
Q.	Accommodation Service 91	1,357	0.52	0.55

See footnote at end of table.

		N	CANADA $\bar{R}^2_{can}$	PROVINCE $\bar{R}^2_{prov}$
	Food and Beverage 92	4,652	0.79	0.80
R.	Amusement and Recr. 96	1,203	0.56	0.56
	Personal and Household 97	2,108	0.78	0.78
	Membership Org. 98	2	-	-
	Other Service Ind. 99	1,815	0.53	0.53

$\bar{R}^2_{can}$: Adjusted coefficient of determination at the Canada level.

$\bar{R}^2_{prov}$: Adjusted coefficient of determination for pooled provinces (dummy variables).

Appendix 5

Best Model Form

Stepwise Regression

by Major Group

Model with division by GBI

SMC = 0 or 1 (Old Records)

GBI $25,000 - $500,000

116,223 records

Variables	Meaning
LGB	Gross Business Income
BSNT	Net Profit
GROSPRO	Gross Profit
QBI	Square of Gross Business Income
DEPR	Depreciation
RENTGRS	Gross Rental Income
PARTNER	Partnership indicator - 0 no partnership - 1 partnership

Model 1: Best one-variable fit.

	Major Group	N	Best Model Form	R^2	Model 1	R^2
A.	Agriculture					
	Agriculture	142	LGB BSNT GROSPRO	0.68	LGB	0.65
	Service to Agric.	124	LGB GROSPRO	0.61	LGB	0.59
B.	Fishing					
	Fishing & Trapping	13	DEPR GROSPRO QBI	0.85	QBI	0.63
C.	Logging and Forestry					
	Logging	964	LGB DEPR BSNT QBI	0.77	LGB	0.72
	Forestry Service	17	LGB DEPR	0.72	LGB	0.62
D.	Mining					
	Mining Ind.	5	BSNT	0.94	BSNT	0.94
	Crude Petroleum	6	BSNT	0.60	BSNT	0.60
	Quarry and Sand Pit	33	LGB RENTGRS	0.78	LGB	
	Serv. Ind. Inc. to Mineral	44	LGB GROSPRO	0.75	LGB	0.73
E.	Manufacturing					
	Food Ind.	615	LGB BSNT GROSPRO PARTNER	0.78	GROSPRO	0.72
	Beverage Ind.	6	GROSPRO	0.95	GROSPRO	0.95
	Rubber Prod. Ind.	1	-	-	-	-
	Plastic Prod. Ind.	19	LGB DEPR	0.89	LGB	0.82
	Leather and Prod.	27	LGB DEPR BSNT	0.76	LGB	0.62
	Prim Textile	7	GROSPRO Q&I	0.86	LGB	0.72
	Textile Prod.	66	LGB BSNT	0.77	LGB	0.75
	Clothing Ind.	130	LGB BSNT	0.64	LGB	0.63
	Wood Inc.	377	LGB BSNT PARTNER	0.77	LGB	0.75
	Furniture and Fixt	336	LGB DEOR BSNT	0.82	LGB	0.77
	Paper and Allied Prod.	4	GROSPRO	0.89	LGB	0.89
	Print, Publ.	467	LGB BSNT PARTNER	0.78	LGB	0.74
	Primary Metal Ind.	14	LGB GROSPRO	0.84	LGB	0.78
	Fab. Metal Prod.	299	LGB BSNT QBI	0.76	LGB	0.70
	Machinery Ind.	16	LGB QBI	0.57	LGB	0.47
	Transp. Equip. Ind.	77	LGB BSNT	0.74	LGB	0.69
	Electrical & Electronic	17	LGB QBI	0.84	LGB	0.75
	Non-Metallic	80	LGB	0.74	LGB	0.74
	Chemical	5	LGB	0.92	LGB	0.92
	Other Manuf. Ind.	426	LGB BSNT	0.70	LGB	0.68
F.	Construction					
	Building, Developing	803	LGB QBI	0.62	LGB	0.60
	Ind. and Heavy Const.	110	LGB DEPR BSNT	0.77	LGB	0.73
	Trade Contracting	10,289	LGB DEPR BSNT	0.67	LGB	0.65
	Service Ind.	1,385	LGB BSNT	0.67	LGB	0.66

Major Group	N	Best Model Form	R^2	Model 1	R^2
G. Transportation					
Transportation Ind.	3,984	LGB DEPR BSNT	0.62	LGB	0.57
Storage & Warehousing	11	LGB QBI	0.68	LGB	0.58
H. Communication					
Communication Ind.	114	LGB BSNT QBI	0.79	LGB	0.72
Other Utility Ind.	134	LGB DEPR BSNT	0.78	LGB	0.74
I. Wholesale					
Farm Prod. Ind.	59	LGB RENT GAS BSNT	0.45	LGB	0.31
Pet. Prod. Ind.	190	LGB DEPR QBI	0.59	LGB	0.43
Food, Bev., Drugs, Tobacco	220	DEPR GROSPRO PROFGRS	0.45	GROSPRO	0.34
Apparel and Dry Goods	41	LGB	0.62	LGB	0.62
Household Goods	12	GROSPRO	0.85	GROSPRO	0.85
Motor Vehicle, Parts	39	BSNT GROSPRO	0.72	GROSPRO	0.63
Metals, Hardware, Plumbing	171	LGB BSNT GROSPRO	0.65	LGB	0.49
Machinery, Equipment	248	LGB DEPR GROSPRO	0.62	LGB	0.56
Other Prod. Ind.	419	LGB QBI	0.46	LGB QBI	0.43
J. Retail					
Food Bev. Drug	4,646	BSNT GROSPRO PARTNER	0.74	GROSPRO	0.70
Shoe Apparel Fab Yarn	1,570	LGB BSNT GROSPRO	0.58	GROSPRO	0.54
Household Furniture	1,241	LGB GROSPRO PARTNER	0.58	GROSPRO	0.55
Automotive Vehicles	4,529	LGB GROSPRO QBI	0.60	GROSPRO	0.55
General Retail	980	DEPR BSNT GROSPRO	0.66	GROSPRO	0.58
Other Retail	3,518	LGB BSNT GROSPRO	0.71	GROSPRO	0.65
Non-store Retail	2,067	LGB BSNT GROSPRO	0.48	GROSPRO	0.44
K. Finance and Insurance					
Invest. Inter.	5	LGB	0.65	LGB	0.65
Other Financial Inter	10	LGB	0.35	LGB	0.35
L Real Estate					
Real Estate & Ins.	127	LGB BSNT	0.44	LGB	0.40
Insurance & Real Estate	176	LGB BSNT	0.70	LGB	0.63
M. Business Service					
Business Service Ind.	534	LGB BSNT GROSPRO	0.69	LGB	0.62
O. Education Service					
Educational Service Ind.	91	LGB DEPR BSNT	0.78	LGB	0.71

	Major Group	N	Best Model Form	R^2	Model 1	R^2
P.	Health and Social					
	Health and Social Ind.	378	LGB DEPR BSNT	0.81	LGB	0.69
Q.	Accommodation					
	Accomodation Serv. Ind.	1,357	LGB BSNT QBI	0.67	LGB	0.52
	Food & Beverage Ind.	4,652	LGB BSNT GRSPRO Q61	0.85	LGB	0.79
R.	Other Services					
	Amusement and Recr. Ind.	1,204	LGB BSNT QBI	0.60	LGB	0.56
	Personal and Household Goods	2,108	LGB BSNT DEOR	0.85	LGB	0.78
	Membership Org. Ind.	2	LGB	0.99	LGB	0.99
	Other Service Ind.	1,815	LGB BSNT GROSPRO	0.54	LGB	0.53

Appendix 6

Best Model Form

Stepwise Regression

DY DIVI and DPROV

SMC = 0 or 1 (Old Records)

Model with division by $\sqrt{GBI}$ throughout the regression equation.

Variables	Meaning
LGB	Gross Business Income
BSNT	Net Profit
GROSPRO	Gross Profit
QBI	Square of Gross Business Income
DEPR	Depreciation
RENTGRS	Gross Rental Income
PARTNER	Partnership indicator
	- 0 no partnership
	- 1 partnership

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Logging and Forestry	AB 41	LGB	0.75	0.75 LGB
	BC 295	LGB, DEPR, BSNT, QBI	0.76	0.69 LGB
	MB 22	LGB	0.80	0.80 LGB
	NB 103	LGB, DEPR	0.77	0.75 LGB
	NF 14	LGB	0.72	0.72 LGB
	NS 95	LGB, OBI	0.89	0.87 LGB
	ON 185	LGB, OBI	0.70	0.69 LGB
	NT			
	*PE 2	LGB, GROS PRO	1.0	0.98 LGB
	PQ 195	LGB, BSNT, QBI	0.74	0.69 LGB
	SK 28	LGB, RENT GRS	0.82	0.74 LGB
	YT			

* - Not enough observations.

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Mining	AB 24	LGB	0.67	0.67 LGB
	BC 15	LGB, BSNT	0.85	0.54 LGB
	*MB 3	LGB, BSNT, RENT GRS	1.0	0.87 LGB
	*NB 2	LGB, BSNT	1.0	0.98 LGB
	*NF 1	LGB	1.0	1.0 LGB
	NS 3	LGB, DEPR, PARTNER	1.0	0.83 DEPR
	NT			
	ON 17	LGB	0.72	0.72 LGB
	PE,			
	PQ 8	LGB	0.65	0.65 LGB
	SK 13	LGB	0.73	0.73 LGB
	*YT 2	DEPR, BSNT	1.0	0.85 BSNT

* - Not enough observations.

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Manufacturing	AB 114	LGB, BSNT	0.77	0.71 LGB
	BC 279	LGB, BSNT	0.78	0.76 LGB
	MB 115	LGB, BSNT	0.72	0.68 LGB
	NB 63	LGB, BSNT	0.79	0.77 LGB
	NF 36	LGB	0.73	0.73 LGB
	NS 79	LGB, BSNT, GROSPRO	0.75	0.68 LGB
	*NT 1	LGB	1.0	1.0 LGB
	ON 1199	LGB, BSNT, FARMGAS	0.70	0.68 LGB
	PE 15	LGB	0.78	0.78 LGB
	PQ 1012	LGB, BSNT	0.68	0.67 LGB
	SK 74	LGB	0.63	0.63 LGB
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Construction	AB 1125	LGB, DEPR, BSNT, GROSPRO	0.64	0.60 LGB
	BC 1567	LGB, BSNT	0.58	0.57 LGB
	MB 549	LGB, DEPR, PARTNER	0.63	0.62 LGB
	NB 345	LGB, BSNT	0.68	0.67 LGB
	NF 148	LGB	0.67	0.67 LGB
	NS 524	LGB, DEPR, PARTNER	0.73	0.71 LGB
	*NT 7	PARTNER	0.79	0.79 LGB
	ON 5175	LGB, DEPR, BSNT, GROSPRO RENTGRS, PARTNER	0.66	0.65 LGB
	PE 85	LGB	0.70	0.70 LGB
	PQ 2426	LGB, DEPR, BSNT, PARTNER	0.71	0.70 LGB
	SK 622	LGB, DEPR, BSNT	0.61	0.58 LGB
	*YT 12	LGB	0.64	0.64 LGB

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Transportation	AB 588	LGB, DEPR, BSNT	0.62	0.54 LGB
	BC 572	LGB, DEPR, BSNT	0.69	0.58 LGB
	MB 253	LGB, DEPR, BSNT	0.62	0.57 LGB
	NB 185	LGB, BSNT	0.63	0.61 LGB
	NF 70	LGB	0.62	0.62 LGB
	NS 149	LGB, BSNT, PARTNER, RENTGRS	0.72	0.63 LGB
	*NT 1	LGB	1.0	1.0 LGB
	ON 1074	LGB, DEPR, BSNT	0.60	0.54 LGB
	*PE 12	LGB	0.46	0.46 LGB
	PQ 840	LGB, DEPR, BSNT	0.65	0.62 LGB
	SK 241	LGB	0.62	0.62 LGB
	*YT 7	LGB	0.83	0.83 LGB

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Communication	AB 19	LGB	0.62	0.62 LGB
	BC 22	LGB	0.74	0.74 LGB
	*MB 8	LGB	0.56	0.56 LGB
	*NB 9	LGB	0.90	0.90 LGB
	*NF 4	LGB	0.87	0.87 LGB
	NS 10	LGB	0.87	0.87 LGB
	*NT 2	LGB, BSNT	1.0	0.99 LGB
	ON 81	LGB, BSNT	0.74	0.70 LGB
	PE			
	PQ 77	LGB, DEPR	0.79	0.77 LGB
	SK 16	LGB	0.63	0.63 LGB
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Wholesale	AB 81	LGB, BSNT	0.49	0.42 LGB
	BC 105	LGB	0.46	0.46 LGB
	MB 59	LGB	0.52	0.52 LGB
	NB 33	LGB, RENTGRS	0.69	0.51 RENTGRS
	NF 20	DEPR	0.72	0.72 DEPR
	NS 52	DEPR	0.47	0.47 DEPR
	NT			
	ON 572	LGB, DEPR, GROSPRO, QBI	0.53	0.47 LGB
	*PE 9	LGB	0.44	0.44 LGB
	PQ 399	LGB, DEPR, BSNT, GROSPRO RENT GRS	0.53	0.43 LGB
	SK 66	DEPR, GROSPRO	0.67	0.51 GROSPRO
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Retail	AB 81	LGB, BSNT, GROSPRO, QBI, PARTNER	0.53	0.49 GROSPRO
	BC 1431	LGB, BSNT, GROSPRO, QBI, PARTNER	0.55	0.51 LGB
	MB 758	LGB, BSNT, GROSPRO, QBI, PARTNER	0.60	0.55 GROSPRO
	NB 453	LGB, GROSPRO	0.64	0.62 GROSPRO
	NF 388	LGB, GROSPRO	0.36	0.34 LGB
	NS 585	LGB, BSNT, GROSPRO, QBI	0.60	0.57 GROSPRO
	*NT 6	LGB	0.68	0.68 LGB
	ON 6975	LGB, DEPR, BSNT, GROSPRO QBI, PARTNER	0.60	0.57 GROSPRO
	PE 75	LGB, DEPR	0.61	0.54 LGB
	PQ 6267	LGB, BSNT, GROSPRO, PARTNER	0.66	0.64 GROSPRO
	SK 712	LGB' BSNT, GROSPRO	0.51	0.47 GROSPRO
	*YT 8	LGB, RENTGRS	0.98	0.90 RENTGRS

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Finance and Insurance	*AB 1	LGB	1.0	1.0 LGB
	BC 1	LGB	1.0	1.0 LGB
	MB			
	NB			
	NF			
	NS			
	NT			
	ON 11	LGB	0.44	0.44 LGB
	PE			
	*PQ 1	LGB	1.0	1.0 LGB
	SK			
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Real Estate	AB 21	LGB	0.47	0.47 LGB
	BC 47	LGB, BSNT	0.46	0.36 LGB
	MB 15	LGB	0.47	0.47 LGB
	*NB 5	RENTGRS	0.70	0.70 RENTGRS
	*NF 1	LGB	1.0	1.0 LGB
	*NS 7	LGB	0.78	0.78 LGB
	NT			
	ON 141	LGB	0.59	0.59 LGB
	PE			
	PQ 44	LGB	0.47	0.47 LGB
	SK 21	LGB	0.73	0.73 LGB
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Business Service	AB 51	LGB	0.62	0.62 LGB
	BC 61	LGB, BSNT	0.67	0.62 LGB
	MB 11	LGB	0.54	0.54 LGB
	*NB 6	LGB	0.62	0.62 LGB
	*NF 5	BSNT	0.77	0.77 BSNT
	*NS 9	LGB	0.63	0.63 LGB
	*NT 1	LGB	1.0	1.0 LGB
	ON 265	LGB, DEPR, BSNT	0.69	0.60 LGB
	PE			
	PQ 111	LGB, BSNT, GROSPRO	0.77	0.73 LGB
	SK 14	LGB	0.77	0.77 LGB
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Educational Service	*AB 8	LGB	0.69	0.69 LGB
	*BC 7	LGB	0.78	0.78 LGB
	MB 4	LGB, DEPR, BSNT, PARTNER	1.0	0.71 PARTNER
	NB			
	NF			
	*NS 3	LGB, DEPR, BSNT	1.0	0.67 LGB
	NT			
	ON 43	LGB, DEPR	0.77	0.72 LGB
	*PE 1	LGB	1.0	1.0
	PQ 22	LGB	0.79	0.79 LGB
	*SK 3	DEPR, BSNT, PARTNER	1.0	0.95 DEPR
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Health and Social	AB 32	LGB, BSNT, RENTGRS	0.89	0.81 LGB
	BC 51	LGB, BSNT	0.83	0.73
	*MB 7	LGB	0.55	0.55 LGB
	*NB 8	LGB	0.94	0.94 LGB
	*NF 10	LGB	0.82	0.82 LGB
	*NS 14	LGB	0.81	0.81 LGB
	NT			
	ON 142	LGB, DEPR, BSNT, RENTGRS	0.85	0.69 LGB
	*PE 3	LGB	0.98	0.98 LGB
	PQ 102	LGB, DEPR, BSNT	0.79	0.67 LGB
	*SK 9	BSNT	0.91	0.91 BSNT
	YT			

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Accommodation	AB 254	LGB, DEPR, BSNT, QBI	0.68	0.63 QBI
	BC 458	LGB, DEPR, BSNT, GROSPRO, PARTNER	0.86	0.58 QBI
	MB 236	LGB, DEPR, BSNT, GROSPRO	0.79	0.73 LGB
	NB 111	LGB, GROSPRO	0.84	0.83 LGB
	NF 76	LGB	0.79	0.79 LGB
	NS 131	LGB, BSNT, GROSPRO	0.87	0.83 LGB
	*NT 4	LGB	0.98	0.98 LGB
	ON 2717	LGB, DEPR, BSNT, GROSPRO, QBI, RENTGRS, PARTNER	0.83	0.77 LGB
	GPE 24	GROSPRO, QBI	0.87	0.81 QBI
	PQ 1763	LGB, DEPR, BSNT, GROSPRO, QBI	0.86	0.82 LGB
	SK 222	LGB, DEPR, BSNT, GROSPRO, QBI	0.82	0.67 GROSPRO
	* YT 7	QBI	0.84	0.84 QBI

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Other Services	AB 423	LGB, BSNT, QBI	0.64	0.60 QBI
	BC 546	LGB, DEPR, BSNT, GROSPRO, QBI	0.66	0.60
	MB 189	LGB, RENTGRS	0.58	0.55 LGB
	NB 70	LGB	0.59	0.59 LGB
	NF 44	LGB, QBI	0.73	0.65 LGB
	NS 135	LGB, BSNT, QBI	0.67	0.57 LGB
	*NT 4	DEPR, BSNT, GROSPRO, QBI	1.0	0.68 GROSPRO
	ON 2370	LGB, DEPR, BSNT, GROSPRO, QBI, PARTNER	0.64	0.62 LGB
	*PE 20	LGB	0.41	0.41 LGB
	PQ 1158	LGB, DEPR, BSNT, QBI	0.69	0.67 LGB
	SK 162	LGB, QBI	0.52	0.55 LGB
	*YT 6	DEPR, BSNT, GROSPRO, QBI, RENTGRS, PARTNER	1.0	0.77 DEPR

DIVI	D PROV Number	Best Model Form	R2 Several Variables	R2 Best Variable
Unproperly Clasified	AB 157	LGB	0.53	0.53 LGB
	BC 187	LGB	0.50	0.50 LGB
	MB 78	LGB	0.42	0.42 LGB
	NB 81	LGB	0.47	0.47 LGB
	NF 37	LGB	0.43	0.43 LGB
	NS 68	LGB, PROFGRS	0.64	0.57 LGB
	NT			
	ON 695	LGB, GROSPRO	0.51	0.50 LGB
	PE 22	LGB	0.67	0.67 LGB
	PQ 667	LGB, PROFGRS	0.51	0.50 LGB
	SK 79	LGB	0.52	0.52 LGB
	*YT 2	DEPR, BSNT	1.0	0.99 DEPR

Comparison of Two Methods of Regression for Estimating Wages and Salaries

M. Hidiroglou

March 1984

Two methods for regressing total salaries and wages (TOTSAW) as a function of gross business income (GBI), net profit before taxes (BSNT) and depreciation (DEPR) are considered. In the first method, total salaries and wages, net profit and depreciation are added up and regressed on gross business income using the following regression equation:

$$\frac{WBD}{\sqrt{GBI}} = a_1\sqrt{GBI} + \text{error} \qquad (1)$$

where WBD = TOTSAW + DEPR + BSNT.

In the second method, total salaries and wages are regressed on gross business income net profit and depreciation using the following equation:

$$\frac{TOTSAW}{\sqrt{GBI}} = a_2\sqrt{GBI} + b_2\frac{DEPR}{\sqrt{GBI}} + c_2\frac{BSNT}{\sqrt{GBI}} + \text{error} \qquad (2)$$

Dividing the variables by the square root of GBI stabilizes the error term in that the standardized residuals become more homogeneous. The use of model (1) may be preferred to model (2) because there may be higher correlations between WBD and GBI as opposed to TOTSAW and GBI. In order to compare these two methods, residual sums of squares for both models were computed. The residual sum of squares is defined as:

$$SSE = \sum (SAWHAT-TOTSAW)^2$$

where SAWHAT = WBDHAT - BSNT - DEPR and WBDHAT is the predicted value for

WBD under model (1). For model (2), SAWHAT is simply the predicted value for TOTSAW. These two methods of regression were computed using a test file of the Combined.Master. The results are provided in the following table:

Table: Comparison of Two Regression Methods

Major Division	Size	SSE for		Correlation between GBI and	
		Method 1	Method 2	WBD	TOTSAW
1. Logging and Forestry	22	9554	8331	0.93	0.88
2. Manufacturing	13	2115	868	0.97	0.95
3. Construction	87	25241	8482	0.78	0.80
4. Transportation	30	5361	2521	0.75	0.57
5. Communication	6	535	545	0.99	0.85
6. Wholesale	8	22496	315	0.51	0.61
7. Retail	104	37358	10365	0.63	0.57
8. Accommodation	21	3716	2070	0.91	0.88
9. Other Services	20	10656	2651	0.34	0.31
10.Unprop. Classified	16	5046	3515	0.68	0.54

From the above table, it is clear that the residual sum of squares is in general smaller for model (2). This observation holds even when the correlations between WBD and GBI are superior to those between TOTSAW and GBI. For instance, in the case of Transportation, the correlation between WBD and GBI is 0.75 which is higher than 0.57 (the correlation between TOTSAW and GBI). However, the residual sum of squares under model (1) is

larger than that obtained under model (2).

Note that model (1) is equivalent to

$$\frac{\text{GBI-TOTSAW-BSNT-DEPR}}{\sqrt{\text{GBI}}} = a_3 \sqrt{\text{GBI}} + \text{ERROR}$$

The conclusion is that models relating total salaries and wages to other variables on the Combined.Master should be arrived at using standard regression model building techniques and that fixed linear combinations of variables should not be considered as dependent even if they display higher correlations with the auxilliary variables than some single dependent variable.