THE UNIVERSITY OF CALGARY

A Monte Carlo Approach To Investigate Sampling Variability In A Semi-log Labour Supply Function

by

Kian Soon Yeo

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF ARTS

DEPARTMENT OF ECONOMICS

CALGARY, ALBERTA

MARCH, 1998

0-612-31313-1

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "A Monte Carlo Approach To Investigate Sampling Variability In A Semi-log Labour Supply Function" submitted by Kian Soon Yeo in partial fulfillment of the requirements for the degree of Master of Arts.

Supervisor, Dr. D. V. Gordon, Department of Economics

Dr. F. J. Atkins, Department of Economics

Dr. T. S. Fung, University Computing Services

March 18/98
Date

ABSTRACT

In the literature of labour economics, the discrepancies in the estimation of wage elasticities are large. Many applied economists have tried to look for both economic and statistical explanations in order to account for these variations. The hypothesis of this thesis is to examine if these variations are attributed to the size of the sample used by a researcher. Monte Carlo procedures are used to investigate the effect of sample size on the stability of the estimated coefficients in a semi-log labour supply function. An experiment is performed by drawing independent small samples from a large database, and estimating the labour supply function using an ordinary least squares estimator. A number of statistical procedure are used to evaluate the stability of the estimated coefficients for different sample sizes. The results show that, at small sample sizes, the estimated coefficients vary more than predicted from statistical theory. Large sample characteristics of the data appear to be reached at about 6,000 observation points.

Acknowledgments

I would like to express my immense gratitude to my supervisor Dr. Daniel V. Gordon for his guidance, patience, and encouragement in overseeing this thesis. I would also like to thank the rest of my defence committee for their suggestions and advice.

I am also very grateful to Dr. Tak S. Fung for his aid in writing the computer programs for this thesis, and Dr. William A. Kerr for his helpful comments on this thesis.

Finally, I would like to thank my family for their support and encouragement.

Dedications

TO

Mum, Dad,

and Brothers

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 The Issue

Is sampling variability a problem? If using different sample sizes produces estimates that are not statistically different from the true values, then sampling variability is not a problem. However, if the estimated coefficients are statistically different from the true values when samples of different sizes are used, then sampling variability is a problem. If using a smaller sample size produces estimates where the expected values are not equal to the true values, these estimated coefficients are bias. While bias can be expected to exist, question remain regarding the exact form which the bias will take. How biased are these estimates? Are these estimates biased upwards or downwards? Are these estimates consistent as the sample size changes?

Formulas to ascertain the minimum sample size based on the tests of hypotheses and to determine the accuracy of predictions arising from multiple regression equation have been developed by statisticians. Medical statisticians like Lachin (1981) and Donner (1984), have derived formulas to determine the minimum number of patients required in clinical trials. In economics, however, very little has been done on the ovation of the minimum sample size for any empirical estimation of the parameters. Many coefficient estimates, such as the estimations of price elasticity and labour supply elasticity, vary from one study to another. In order to account for these variations, applied economists have tried to look for both economic and statistical explanations. In labour economics, the discrepancies in the estimation of wage elasticity are large. For example, the results for men suggest that the range for uncompensated wage elasticity is -0.38 and +0.14. For women, it is between -0.89 and +15.24 (Killingsworth, 1983). However, not all researchers have a criteria as to the size of the sample used. If they cannot choose the size of the sample used, they must accept a given sample size as a constraint on their empirical work. Nakamura and Nakamura (1983) used census data with samples up to 1,994, Mroz (1987) used samples up to 753, Ham (1982) had 835 observations, and Nakamura, Nakamura and Cullen (1979) used samples up to 4,762.

1.2 The Technique

A Monte Carlo experiment has been selected to investigate the problem of sampling variability. The use of Monte Carlo techniques provides the freedom to choose the true values of the coefficients and to generate the error terms from a normal distribution. Of course, the true values of the coefficients are never actually known. As a result, true values are usually arbitrarily selected by the researcher (Gujarati, 1995: 85-86). In this thesis, a set of true values are assumed to arise when the total sample is used for estimation. Various sample sizes are then employed to generate coefficients for the same equation. These coefficients are then tested to determine if they are statistically different from the "assumed" true values.

Sampling variability may be the cause of the wide range of parameter estimates often observed in applied economics, especially when labour supply function have been estimated. Therefore, in this thesis, attention is focused on the elasticity of labour supply. Using the Monte Carlo technique, coefficients in a labour supply function are estimated using a large sample size. These parameters are assumed to be the "true" values. Using the values of the "true" coefficients, and the actual data for the explanatory variables, error terms are generated from a normal distribution using a random number generating process.[1] Having all the required data on the right-hand side of the model, predicted values of the dependent variable are obtained for each observation. Using different sample sizes, various labour supply estimates are obtained and tested to determine if they are statistically different from their "true" value.

1.3 Objectives

The size of a sample could be a contributing factor to the wide range of labour supply elasticity estimates observed in labour economics. The Monte Carlo technique can be used to investigate the problem of sampling variability. Hence, the objectives of this thesis are:

(1) to review the existing literature on sample size determination;

(2) to test the effect of sample size on the parameters of the semi-log labour supply function using a Monte Carlo experiment;

(3) to determine empirically if changes in sample size will result in the labour supply elasticity estimates which are significantly different from the "true" value; and

[1] Random numbers can be generated from a random number table, or from the computer software which is used to

(4) to investigate if small sample sizes will affect the stability of the estimated coefficient of labour supply elasticity given a pre-specified labour supply function.

In short, the objective of this thesis is to provide an econometric investigation of the effect of sample size on empirical studies in applied labour economics using the Monte Carlo technique.

1. 4 Outline

The organization of this thesis is as follows. A review of the literature on sample size determination and a detailed discussion of the important statistical models relating to appropriate sample size is presented in Chapter 2. This review is provided to explain the importance of sample size as a determinant of the accuracy of the estimates of coefficients of regression equations. The importance of sample size evaluation in clinical trials and the general methodologies which are used to derive specific equations for sample size determination is discussed in detail. The analytical power of a variety of statistical procedure are compared.

Chapter 3 describes the Monte Carlo methodology. A literature survey on the applications of Monte Carlo experiments is presented to provide an insight to the ways in which the Monte Carlo technique has been used by researchers. A simplified Monte Carlo experiment is presented as an example and the validity of statistical testing is

---

perform the regression.

explained. The questions chosen to be answered in this thesis are presented. The theoretical model of a labour supply function is presented and its specification is discussed. The labour supply behaviour of male respondents in 1986 is investigated in more detail. Chapter 3 also defines and explains the problem of sample selectivity bias and the use of the inverse Mills ratio to correct for this bias. Furthermore, the incorporation of taxes into labour supply functions is discussed. This is followed by a discussion of the application of Monte Carlo experiments to the labour supply function.

Data from the Labour Market Activity Survey (LMAS) were used in the experiment conducted in this thesis. Chapter 4 discusses the methodology which was used to obtain the data. The variables which are incorporated into the labour supply function are explained in detail. The issues of total and partial non-response as sources of non-sampling error are also discussed in Chapter 4.

Chapter 5 presents the results from the econometric estimations and discusses the implications of the findings. A conclusion and suggestion for future extension to this work are provided in Chapter 6.

CHAPTER 2: A LITERATURE REVIEW

This chapter reviews the literature concerning sample size determination. Section 2.1 presents the formulas which have been derived for sample size determination based on the needs of hypotheses testing and establishing the accuracy of predictions made on the basis of multiple regression equations.

Sample size has also been important in the design of clinical trials. In Section 2.2, a review of literature in clinical trials is presented. Considerable interest has focused on the design of randomized controlled trials (RCTs). If the number of patients exceeds the minimum required number, clinical trials will be more expensive than necessary and possibly prolonged. The investigator must strike a balance between enrolling sufficient patients to detect important differences, but not so many patients such that important resources would be wasted. Formal sample size planning in the design of clinical trials usually depends on relatively simple and well-known formulae presented in introductory statistics texts.

2.1 Sample Size Determination

The purpose of this section is to provide a brief explanation of the formulas pertaining to appropriate sample size derived by statisticians. These formulas are based on needs of hypotheses testing and determining the accuracy of predictions arising from multiple regression equations.

Mace (1974) derived formulas for sample-size determination when the experimental objective is formulated as a test hypothesis. The purpose of determining the minimum sample size is to assist a researcher in limiting arbitrarily pre-assigned small risk levels. These risks involve the errors associated with wrongly rejecting the test hypothesis (Type I) and the errors of failing to reject the test hypothesis when a discrepancy of some pre-assigned magnitude actually exist. He discusses tests of hypotheses involving the properties of commonly used statistical distributions:

(i) means of normally distributed variables,

(ii) variances of normally distributed variables,

(iii) means of binomially distributed variables, or

(iv) means of exponentially distributed variables.

The optimum procedure for testing hypotheses about normal means is based on the test statistics $t = (\bar{x} - \mu_0)\sqrt{n}/\sigma$. The decision rule for computing a critical value of this test statistic is formulated as follows: if the actual test statistic based on experimental results is less than the critical value, one may reject the test hypothesis with the error of erroneously rejecting this hypothesis when it is true limited to, at most, the α probability level.

In the tests of hypotheses about normal means, two variables are defined: μ is the value of the true but unknown population mean and μ_0 is the least favourable value of this mean consistent with the test hypothesis. To derive a formula for a required sample size, Mace (1974) assumed that the least favourable value of the test hypothesis is true, that is

$\mu = \mu_0$. Then the test statistic will be normally distributed with mean zero and variance unity. In order to satisfy the relation

$$Pr\{t < C \text{ if } \mu = \mu_0\} = \alpha,$$

he suggested choosing the critical value of the test statistic $C = u_{1-\alpha}$ where $u_{1-\alpha}$ is the upper limit of a cumulative standard normal probability integral.

For the remaining formulas, Mace (1974) assumed that the researcher can control the error of failing to reject the test hypothesis when it is false (Type II error defined by β) in order to obtain sufficient discriminating power in the test. This mean that when some alternative value of the true mean μ_1 is selected, it will be less than the test hypothesis by the standardized distance $(\mu_1 - \mu_0) / \sigma$, such that, if the true mean is really μ_1, the test procedure will erroneously fail to reject the test hypothesis at the β probability level. When the population standard deviation is known, the required sample size is

$$n = \left[\frac{(u_\alpha + u_\beta)\,\sigma}{\mu_0 - \mu_1}\right]^2 \qquad (2.1)$$

where u_α and u_β are the lower limits of the cumulative standard normal probability integrals, σ is the known common standard deviation, and $\mu_0 - \mu_1$ is the width of the tolerable region in which the discriminating power of the test is less than $1 - \beta$.

To test the hypothesis where the mean of one normally distributed variable is at least a given amount larger than the mean of a second normally distributed variable when there are independent samples from population with a known common standard deviation, the required sample size is

$$n = 2\left[\frac{(u_\alpha + u_\beta)\,\sigma}{\delta_0 - \delta_1}\right]^2 \qquad (2.2)$$

where u_α and u_β are the lower limits of the cumulative standard normal probability integrals, σ is the known common standard deviation, and $\delta_0 - \delta_1$ is the width of the tolerable region in which the discriminating power of the test is less than $1 - \beta$.

However, if the standard deviations from the two samples are unequal, the required sample size will be

$$n_1 = \sigma_1(\sigma_1 + \sigma_2)\left[\frac{(u_\alpha + u_\beta)}{\delta_0 - \delta_1}\right]^2 \qquad (2.3)$$

$$n_2 = \sigma_2(\sigma_1 + \sigma_2)\left[\frac{(u_\alpha + u_\beta)}{\delta_0 - \delta_1}\right]^2 \qquad (2.4)$$

where σ_1 and σ_2 are the known population standard deviations and u_α, u_β, δ_0 and δ_1 are the same as for equation (2.2). The test of highest discriminating power for a given total sample size $n_1 + n_2$ under the tests of hypotheses about the difference between two normal means with two independent samples from populations with unequal variances will be obtained when $n_1/n_2 = \sigma_1/\sigma_2$.

When testing the mean of the differences between paired observations from a bivariate normal distribution is at least as large as a given amount, the required sample size is

$$n = \left[\frac{(u_\alpha + u_\beta)\,\sigma_d}{\delta_0 - \delta_1}\right]^2 \qquad (2.5)$$

where σ_d is the known population standard deviation of differences which may be obtained from the known covariance matrix of the bivariate normal distribution, that is, $\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho_{12}\,\sigma_1\,\sigma_2}$, and u_α, u_β, δ_0 and δ_1 are the same as for equation (2.2).

To test that the ratio of the variances of two normally distributed variables is no larger than a certain standard, the required sample size for each sample is

$$n_1 = n_2 = 1 + \left[\frac{u_\alpha + u_\beta}{\ln \lambda}\right]^2 \qquad (2.6)$$

where u_α and u_β are the lower limits of the cumulative standard normal probability integrals and λ is the square root of the pre-assigned alternative value of the true variance ratio to the pre-assigned least favourable value of the true variance consistent with the test hypothesis.

Sawyer (1982) tried to estimate the sample size required to achieve a given level of accuracy in a prediction arising from multiple regression. He showed that when a prediction equation is based on a sample from a multivariate normal population, the mean error of prediction, relative to its minimum asymptotic value, can be closely approximated by a simple function of the number of predictor variables and the required sample size.

He considered the mean absolute error prediction, that is, the mean deviation of the distribution of prediction error, as an alternative measure of prediction accuracy. He showed that in the context of sampling from a multivariate normal population, the distribution of prediction error is approximately normal. Hence, the mean absolute error is approximately $\sqrt{2/\pi}$ times more than root mean square error. Thus, under normal approximation, the mean absolute error is equal to the product of $\sigma\sqrt{2/\pi}$ (its asymptotic value as the required sample size $n \rightarrow \infty$) and an inflation factor $K = \sqrt{(n+1)(n-2)/[n(n-p-2)]} > 1$ due to estimating coefficients. The inflation factor is a function of n and the number of predictors p. On solving for n in terms of p and K, he found that

$$n = \frac{K^2(p+2)-1+\sqrt{[(p+2)K^2-1]^2-8(K^2-1)}}{2(K^2-1)} \qquad (2.7)$$

Although n is not a linear function of p, he pointed out that it can be approximated by

$$n = \frac{2K^2-1}{K^2-1} + \frac{K^2}{K^2-1}\,p \qquad (2.8)$$

The first study by Mace (1974) is aimed at determining the properties of the variables involved in any experiment. The second study by Sawyer (1982) focused on forecasting using multiple regression equations. The latter also involves a sample having a multivariate normal population. These two studies provide a good guideline in determining the size of a sample if the properties of the variables are known. These studies attempt to determine the minimum sample size used in an experiment. However, the aim of this thesis is to investigate the characteristics of the estimated coefficients at various sample sizes.

2.2 Sample Size Determination In Clinical Trials

The purpose of this section is to provide a brief review of sample size determination in the literature on clinical trials. Clinical statisticians have also tried to derive formulas to determine the "optimal" sample size. This is because knowing the number of patients required to detect important differences will not waste resources due to an excessive number of patients being used.

The fundamental considerations in clinical trials are similar to those considered in the studies presented in the previous section. When conducting a statistical test in clinical trials, two types of error must be considered: Type I (rejecting the true hypothesis) and Type II (not rejecting the false hypothesis). These will have probabilities α and β respectively. Donner (1984) provided a summary of the formulae derived by other researchers in clinical trials. There are several approaches used in the design of clinical trials. These formulae are based on the comparison of an experimental treatment, *E*, with a control treatment, *C*. As in the previous section, the standard formulae for sample size depends on the chosen probabilities α and β associated with a type I error (falsely declaring a treatment difference) and a type II error (falsely declaring no treatment difference), respectively. The quantities Z_α and Z_β are defined as the values of the standardized normal deviates corresponding to α and β, where $1 - \beta$ is the trial power.

The approach to determining the requirements of the sample in terms of risk difference is a straightforward application of the traditional sample size formulae for comparing two proportions. To determine the sample size for this aspect of clinical trials, Donner (1984) lets

P_C = the anticipated event rate (in years) among control group patients

P_E = the anticipated event rate (in years) among experimental group patients

$\delta = P_E - P_C$ = the difference in event rates regarded as scientifically or clinically important to detect.

The null hypothesis is H_0: $P_E = P_C$, and the sample size required is

$$n = \frac{\left\{ Z_\alpha \sqrt{2\overline{P}(1-\overline{P})} + Z_\beta \sqrt{P_E(1-P_E) + P_C(1-P_C)} \right\}^2}{\delta^2} \qquad (2.9)$$

where $\overline{P} = (P_E + P_C)/2$.

The test statistic used is the chi-square contingency test.

To derive the formula for the sample size required in terms of relative risk, the following is denoted:

Let $R = P_E / P_C$ = relative risk regarded as clinically or scientifically important to detect.

The null hypothesis is H_0: $R = 1$, and the formula is

$$n = \frac{\left\{ Z_\alpha \sqrt{2\overline{P_R}(1-\overline{P})} + Z_\beta \sqrt{P_C\{1 + R - P_C(1+R^2)\}} \right\}^2}{\left[P_C(1-R)\right]^2} \qquad (2.10)$$

where $\overline{P_R} = \frac{1}{2} P_C(1 + R)$.

The chi-square contingency test is the test statistic used.

An increasing number of clinical trials seek to show that an experimental therapy is equivalent in efficacy to a control therapy, rather than superior. This often occurs when the experimental therapy is 'conservative' and the standard control therapy is invasive or toxic. In this case, the null hypothesis may specify that the success rate P_C on the control therapy is higher than the success rate P_E on the experimental therapy by at least some amount θ. The alternative hypothesis specifies that P_C - P_E $< \theta$, which implies that the two therapies are equivalent. In other words, the null and alternative hypotheses are

$$H_0\colon P_C \geq P_E + \theta$$

$$H_1\colon P_C < P_E + \theta$$

Therefore, the sample size requirements for such clinical trials which are designed to show equivalence is

$$n = \frac{\left\{ Z_\alpha \sqrt{2\bar{P}(1-\bar{P})} + Z_\beta \sqrt{P_E(1-P_E) + P_C(1-P_C)} \right\}^2}{\left(P_E - P_C - \theta\right)^2} \qquad (2.11)$$

where $P_C < P_E + \theta$ and $\theta > 0$.

The test statistic used is the critical ratio for difference between two proportions.

The next approach assumes stratification of subjects into K risk categories, for example based on age, with n_j subjects randomly assigned to each of an experimental and control group within the jth stratum, $j = 1, 2, \ldots, K$. One wishes to compare event rates within each of the resulting 2×2 tables, and to obtain an overall comparison to test whether the assumed common relative odds equals unity. Let P_{Cj} and P_{Ej} denote the event rate among control and experimental group patients in the jth stratum. Then the common relative odds is OR = $P_{Ej}(1-P_{Cj})/P_{Cj}(1-P_{Ej})$, for all j. Define

$$\Delta = \log_e(\text{OR})$$

$$g_j = \frac{\Delta^2}{\frac{1}{P_{Cj}(1-P_{Cj})}+\frac{1}{P_{Ej}(1-P_{Ej})}}, j = 1, 2, ..., K.$$

f_j = fraction of observations contained in the jth table, j = 1, 2, ..., K.

The null hypothesis is H_0: OR = 1, and the formula used to determine the sample size that account for stratification of subjects is

$$n = \frac{\left(Z_\alpha + Z_\beta\right)^2}{\sum^{K} g_j f_j} \quad (2.12)$$

where $n = \sum^{K} n_j = \sum^{K} f_j n$

The unconditional large sample test and/or the Mantel-Haenszel chi-square test are the test statistic used.

To determine the sample size requirements in terms of time to some critical event, it is assumed that greater interest attaches to the time to some critical event, such as death or the recurrence of disease, rather than the occurrence or non-occurrence of the event. Thus the approach pertains particularly to studies that aim to compare survival rates arising from different treatments.

It is assumed that the time-to-event (survival time) has an exponential distribution with means μ_C and μ_E in the control and experimental groups, respectively. This is equivalent to the assumption that the 'hazard function' or instantaneous probability of death (recurrence of disease, et cetera) is constant within each group. It is also assumed that patients entered the trial to a Poisson process.

Let $\theta = \mu_E / \mu_C$ = ratio of mean survival times regarded as important to detect. If all patients are followed-up in each group until the occurrence of the critical event, that is, there are no censored observations, the null hypothesis is H_0: $\theta = 1$, and the following equation gives the required number of patients.

$$n = \frac{2(Z_\alpha + Z_\beta)^2}{[\log_e(\theta)]^2} \tag{2.13}$$

Test statistic used is Cox's *F*-test.

Schork and Remington (1967) proposed an approach which takes into account yearly 'shifts' of subjects from the experimental group to the control group. Subjects who drop-out effectively become characterized by the control group event rate P_C from that point onward. To apply this approach, one must anticipate, on the basis of past experience, a particular pattern of shift, that is, the percentage of subjects anticipated to shift from the experimental to control group per unit time, for example the number of years, of the trial.

Suppose that the anticipated yearly event rates in the experimental and control groups are P_{EY} and P_{CY} , respectively, and the yearly drop-out rate in the experiment group is d_i, $i = 1, 2, ..., L$, where L is the study duration in years. Then the effective T-year event rate in the experimental group is

$$P_E^* = \sum_{i=1}^{L} d_i \left[1-(1-P_{EY})^{i-1/2}(1-P_{CY})^{L-i+1/2}\right] + C\left[1-(1-P_{EY})^2\right] \tag{2.14}$$

where $C = 1 - \sum d_i$ is the proportion of experimental group subjects anticipated to complete the study. One may easily evaluate this formula for any given values P_{EY}, P_{CY} and d_i, $i = 1, 2, \ldots, L$ and then substitute P_E^* for P_E in equation (2.9).

It is common during the course of a clinical trial that some patients assigned to the experimental regimen 'drop out' or fail to adhere to the prescribed protocol, although their outcomes are still recorded. Since one must count such individuals against the experimental group in the statistical analysis, the effect of patient drop-out is to dilute the effective treatment difference. Several approaches have evolved for taking this problem into account in the calculation of sample size requirements.

Lachin (1981) has proposed a very simple method of adjusting sample size requirements for an anticipated drop-out rate d among patients in an experimental group. This approach characterizes drop-outs by the control event rate P_C, rather than the event rate P_E corresponding to their original group assignment. It follows that the effective value P_E^* of the T-year event rate P_E is $P_E^* = P_E(1 - d) + P_C d$, and the effective difference δ^* by $P_E^* - P_C = (1 - d)(P_E - P_C)$. Substitution of δ^* for δ in equation (2.9) implies division of the usual formula for sample size requirements by the factor $(1 - d)^2$ to inflate appropriately the number of patients entered into the trial.

Lachin (1981) also considered a general family of statistics that are normally distributed under a null hypothesis (H_0) as $N(\mu_0, \Sigma_0^2)$ and under an alternative hypothesis (H_1) as $N(\mu_1, \Sigma_1^2)$; where $\mu_1 > \mu_0$ or $\mu_1 < \mu_0$ and where Σ_0^2 and Σ_1^2 are some function of the variance σ^2 of the individual observations and the sample size N. In a clinical trial, the parameter μ is the treatment-control difference in the outcome of interest, for example, the mean difference on some measurable pharmacologic effect such as serum healing cholesterol, or the difference in the proportion displaying an event such as healing. In such cases, μ_0 is usually zero and μ_1 is specified as the minimal clinically relevant therapeutic difference.

When the statistical test is conducted, the probability of Type I error, α, is specified by the investigator. However, the probability that a significant result will be obtained if a real difference (μ_1) exists, that is, the power of the test, $1 - \beta$, depends largely on the total sample size N. As one increases N the spread of a normal distribution decreases, that is, the curves tighten; thus β decreases and the power increases. Thus if the statistical test fails to reach significance, the power of test becomes a critical factor in reaching an inference. It is not widely appreciated that the failure to achieve statistical significance may often be related more to the low power of the trial than to an actual lack of difference between the competing therapies. Clinical trials with inadequate sample size are thus doomed to failure before they begin and serve to confuse the issue of determining the most effective therapy for a given condition. Thus one should take steps to ensure that the power of the clinical trials is sufficient to justify the effort involved.

The problem in planning a clinical trial is to determine the sample size N required such that in testing the null hypothesis H_0 with stated probability of Type I error α, the probability of Type II error is a desired small level β. The parameters of the problem are α, β, μ_0, μ_1, Σ_0^2, Σ_1^2. Since the variances Σ^2 are functions of N, the sample size required is that which simultaneously satisfies the equalities $\Pr(Z > Z_\alpha) = \alpha$ if H_0 is true and $\Pr(Z > Z_\alpha) = 1 - \beta$ if H_1 is true; where Z_α is the standard normal deviate at the α significance level and where $Z = (X - \mu_0)\,\Sigma^{-1}$ is the simple statistic one would use in testing H_0; where $Z \sim (0,1)$ if H_0 is true. He stated that the sample size that satisfies these equalities also satisfies the equality

$$|\mu_1 - \mu_0| = Z_\alpha \Sigma_0 + Z_\beta \Sigma_1 \qquad (2.15)$$

Three basic questions one can ask are

1. What sample size is required to ensure power $1 - \beta$ of detecting a relevant difference μ_1?
2. What is the power (Z_β) of the experiment in detecting a relevant difference when a specific sample size N is employed?
3. What difference μ_1 can be detected with power $1 - \beta$ if the experiment is conducted with a specified sample size N?

Question 1 is employed in planning an experiment and can be determined by solving equation (2.15) for N once the expression for the variances Σ^2 have been obtained. In many cases, Σ^2 will be a function of the form $\Sigma^2 = \sigma^2/N$, where σ^2 is the variance of the individual measurements and N is the total sample size. In this case

$$|\mu_1 - \mu_0| = (Z_\alpha \sigma_0 / \sqrt{N}\,) + (Z_\beta \sigma_1 / \sqrt{N}\,) \qquad (2.16)$$

Solving for the total sample size N one obtains

$$N = \left[\frac{Z_\alpha \sigma_0 + Z_\beta \sigma_1}{\mu_1 - \mu_0} \right]^2 \qquad (2.17)$$

Question 2 is employed in evaluating the results of an experiment. To determine the power, one can solve for Z_β from equation (2.16)

$$Z_\beta = \frac{\sqrt{N}\,|\mu_1 - \mu_0| - Z_\alpha \sigma_0}{\sigma_1} \qquad (2.18)$$

Power (1 - β) can then be determined from the value Z_β by referring Z_β to tables of the normal distribution where $Z_\beta < 0.00$ indicate power < 0.50.

Question 3 can be employed in either case. The minimal difference with power given a sample size N is obtained by solving equation (2.18) for μ_1. Equation (2.18) can thus be written as

$$\sqrt{N}\,|\mu_1 - \mu_0| = Z_\alpha \sigma_0 + Z_\beta \sigma_1 \qquad (2.19)$$

which can then be solved for N or Z_β. Equation (2.19) can be employed in cases where the basic equations equivalent to equations (2.17) and (2.18) become cumbersome.

Lachin (1981) also demonstrated how these simple relationships can be employed with the Student's t tests, chi-square tests for proportions, analyses of survival time, and test for correlations. The Student's t test is used to test the hypothesis that the mean of a normal variable, v, equals some specified value H_0: $\mu_0 = v_0$ against some alternative H_1: $\mu_1 = v_1$, $v_1 \neq v_0$, when the variance is unknown. The test statistic is one of the form $t = \sqrt{N}\,(x - \mu_0)/S$ where x is the sample mean with standard error S^2/N, S^2 being the unbiased sample estimate of the variance on N - 1 degrees of freedom. The distribution of t becomes increasingly close to that of a standard normal variable as the degrees of freedom increases, at least 30 degrees of freedom being required for the approximation to be adequate. Thus, equations (2.15) through (2.18) can be employed to yield an approximate evaluation of sample size and power. However, he claims that this approach will tend to overestimate power for given S^2 and N, and thus it will tend to underestimate the required sample size, although this effect is increasingly negligible for increasing degrees of freedom. He suggested an adequate adjustment is obtained by the correction factor $f = (\mathrm{df} + 3)/(\mathrm{df} + 1)$, where df is the degrees of freedom. fN denotes patients that are actually employed after N is obtained from equation (2.17), or alternately, by N/f used in equation (2.18) when solving for power.

These studies have derived formulas to assist investigators in determining the minimum number of patients required in a research design to achieve statistical accuracy and not to waste important resources. There are, however, a large number of arbitrary assumptions which may affect the results of an experiment. The survival time of a patient is modelled under an exponential distribution, and patients enter the trial according to the properties of a Poisson distribution. Nevertheless, the formula do provide an approximation as to the correct sample size to use in clinical trials. In this thesis, determining the minimum or correct sample size required in labour economics is secondary. One of the aims of the thesis remains investigating the stability of estimated coefficients at different sample sizes.

CHAPTER 3: MONTE CARLO EXPERIMENT

The Monte Carlo method is associated with random sampling, which is defined as a process where games of chance are being carried out. The earliest use of random sampling dates back to 1777 when Comte de Buffon performed an experiment to determine the probability that a needle will intersect the straight lines drawn on a horizontal plane when the needle was thrown at random. W. S. Gossett (1908) used random sampling to assist in his discovery of the distribution of the correlation coefficient. During the Second World War, there was a surge of interest in the Monte Carlo method. According to Kalos and Whitlock (1986), Von Neumann, Fermi, Ulam, and Metropolis wrote papers which described the new procedure and how it could be used to solve problems in statistical mechanics, radiation transport, economic modeling, and various other fields. The objectives of this chapter are to review the literature on the Monte Carlo experiment, and to introduce the Monte Carlo methodology which can be applied to the labour supply function. The objective is to investigate the stability of the parameter for labour supply elasticity when different sample sizes are used.

This chapter is organized in the following manner. Section 3.1 provides a literature review on the Monte Carlo experiment and some of its applications in the field of statistics. One such application is the construction of probability distributions using the Monte Carlo technique. This section presents the three main areas of application of Monte Carlo: the use of random numbers; distribution sampling; and simulation.

In section 3.2, the Monte Carlo technique is described. This is followed by a discussion on the form of statistical testing used and the statistical issues which the Monte Carlo procedure can help to investigate. A model is required in order to apply the Monte Carlo technique. Section 3.3 specifies the labour supply function used for estimation. The issue of collinearity for this function and the endogeneity of the wage variable are discussed. Section 3.4 explains the term *sample selection* (or *selectivity*) *bias*. In order to correct for sample selection bias, a sample selection rule is defined in this section.

Section 3.5 provides a discussion on the effect of taxes on the labour supply function since taxes have been a great concern in determining the labour supply. The conclusions of other applied labour economists are presented. The decision to incorporate taxes into the labour supply function used in this thesis is based on these conclusions. Section 3.6 explains how the Monte Carlo technique can be applied to the labour supply function specified in this thesis.

3.1 Literature Review

The Monte Carlo procedure is a general technique for finding solutions to problems using random numbers or pseudorandom numbers. Random numbers are stochastic variables which are uniformly distributed and show stochastic independence. Pseudorandom numbers, on the other hand, are generated by applying a deterministic algebraic formula which results in numbers that are considered to behave as random numbers. Pseudorandom numbers are uniformly distributed and mutually independent.[2]

There are three common areas where Monte Carlo procedures have been applied. First, is in the solution to deterministic problems. A deterministic problem is referred to as a situation which there is a difficult non-probabilistic (that is, deterministic) problem to be solved and for which a stochastic process may be invented. This process produces a set of numbers which has moments or distributions satisfying the relations in the non-probabilistic problem (Morgenthaler, 1961). The second application of Monte Carlo procedure is for sampling distributions. The purpose is to find the distribution of a stochastic variable. The third application is in simulation. As defined by Morgenthaler (1961), to "simulate" means to duplicate the essence of the system or activity without actually attaining reality itself.[3]

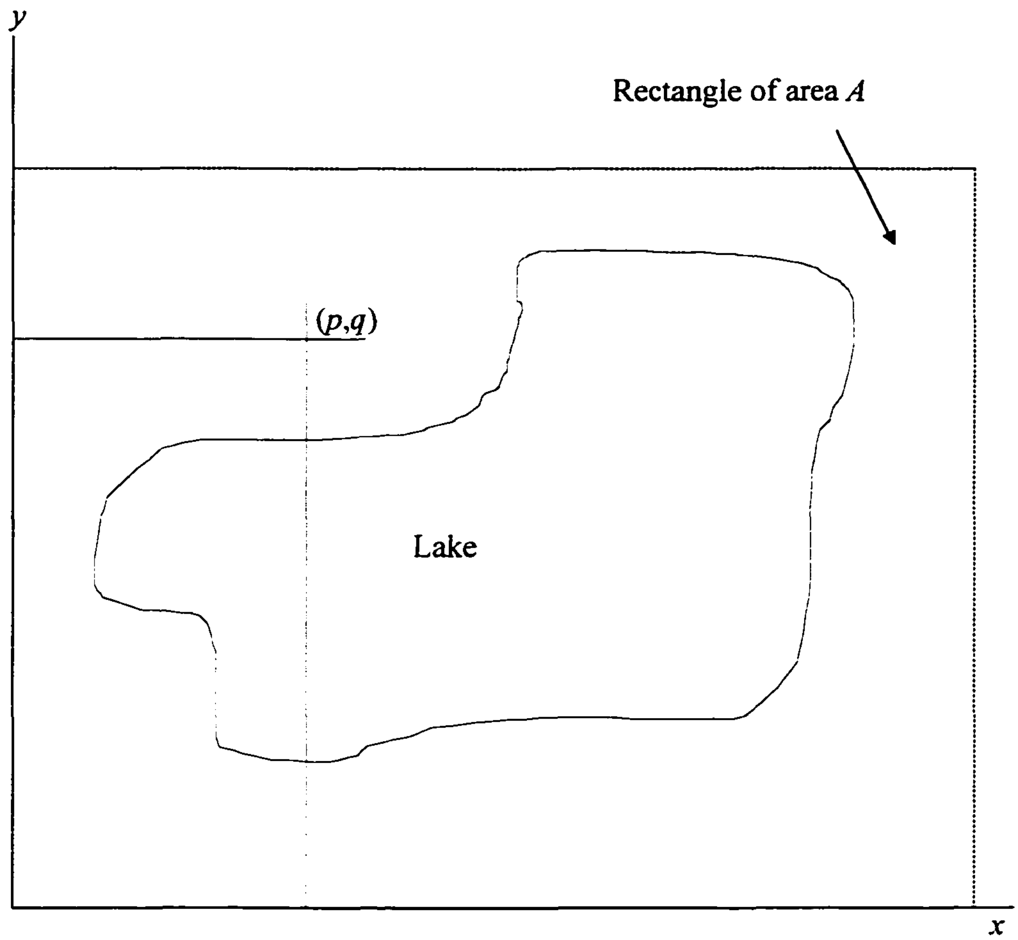
The first application for the use of Monte Carlo techniques concerns the solution of deterministic problems. Morgenthaler (1961) comments that the term Monte Carlo was introduced by Von Neumann and Ulam in the late 1940's. A deterministic problem can be solved by the Monte Carlo technique if the problem has the same formal expression as some stochastic process. *Stochastic process* means that a sequence of states is created whose evolution is determined by random events. Morgenthaler (1961) applied Monte Carlo techniques to calculate the surface water area of an irregularly shaped lake.

Figure 3.1 shows a lake enclosed in a rectangular area, A. Two points, p and q, are randomly selected from uniform distributions. Then the product

---

[2] Kleijnen, J. P. C. (1974), pp. 6.

Figure 3.1: An example of stochastic experiment to solve non-probabilistic problem.[4]



p and q are uniformly distributed.

Estimated area of water

$$= \text{Area of rectangle, } A \times \left(\frac{\text{number of points } (p,q) \text{ falling in the water}}{\text{total number of trials } N} \right)$$

[3] Kleijnen, J. P. C. (1974), pp. 9-12.
[4] Morgenthaler, G. W. (1961), pp. 369.

$$A\left(\frac{\text{number of points }(p,q)\text{ falling in the water}}{\text{total number of trials }N}\right) \qquad (3.1)$$

approximates the water area. The Monte Carlo procedure is applied in this example by increasing the number of trials, N, where points (p,q) are selected. The standard deviation can be approximated by

$$\sigma = \sqrt{\frac{\hat{a}(A-\hat{a})}{N}} \qquad (3.2)$$

where $\hat{a}$ is the estimated water area.

However, the foregoing example may be considered as an inefficient Monte Carlo application since no variance reduction techniques are employed. A variance reduction technique reduces the variance of the estimator by replacing the original sampling procedure by a new procedure that yields the same expected value but with a smaller variance. The efficiency of a variance reduction technique is usually measured by the decrease of the variance of the estimator of the mean. Referring to the aforementioned example, the variance can be minimized for a given sample size N, by making the rectangle enclosing the lake as small as possible, that is $(A - a)$ small. Alternatively, one can also increase the number of trials, N. This is obvious from the formula for σ in equation (3.2).[5]

The Monte Carlo method can also be applied to determine the value of an integral[6]. For example, τ is defined as the value of

$$\int_{u}^{\infty} \frac{1}{y} \gamma e^{-\gamma y} dy$$

[5] Morgenthaler, G.W. (1961), pp. 370.

or, $$\tau = \int_u^{\infty} \frac{1}{y} \gamma e^{-\gamma y}\, dy \qquad (\gamma, u > 0) \qquad (3.3)$$

The value of this integral is difficult to solve using direct partial integration or serial expansion[7]. The Monte Carlo procedure can be used to estimate the value of (3.3) in the following steps. First, sample a value of y from its exponential function and let f(y) be the density function of (3.3):

$$\begin{aligned} f(y) &= \gamma e^{-\gamma y} && \text{for } y \geq 0 \\ &= 0 && \text{for } y < 0 \end{aligned} \qquad (3.4)$$

Second, let g(y) be the other part of the integral in (3.3) where

$$\begin{aligned} g(y) &= 0 \quad \text{if} \quad y < u \\ &= \frac{1}{y} \quad \text{if} \quad y \geq u \end{aligned} \qquad (3.5)$$

Substitute the sampled value of y into g(y) as defined in (3.5). Then, the expected value of g(y) is given by (3.6):

$$\begin{aligned} E[g(y)] &= \int_{-\infty}^{\infty} g(y) f(y)\, dy \\ &= 0 + \int_u^{\infty} \frac{1}{y} \gamma e^{-\gamma y}\, dy \\ &= \tau \end{aligned} \qquad (3.6)$$

---

[6] Kleijnen, J. P. C. (1974), pp. 30-31.
[7] However, Kleijnen, J. P. C. (1974) shows that the solution is

$$\tau(\gamma, u) = \gamma \left[-c - ln(\gamma\, u) + \sum_{j=1}^{\infty} \left\{ (-1)^{j+1} \frac{(\gamma\, u)^j}{j!\, j} \right\} \right]$$

where c is the Euler constant

The Monte Carlo procedure implemented by repeating these two steps using different random numbers. If y_j denotes the observation on y sampled in replication j ($j = 1, ..., n$) then τ can be estimated by $\hat{\tau}$ defined in (3.7):

$$\hat{\tau} = \frac{1}{n}\sum_{j=1}^{n} g(y_j) \qquad (3.7)$$

Because of (3.6), $\hat{\tau}$ is an unbiased estimator of τ. As n increases, the variance of this estimator decreases, and the probability of correct estimation of τ increases.

In distribution sampling, the Monte Carlo technique is used to estimate the distribution of some parameters. For example, if one wishes to know the distribution form of a/b, where a and b ($b \neq 0$) are random variables from some unknown distributions. One may sample a and b in pairs and form a/b for each pair. The resulting shape of a histogram of the a/b values will yield an approximation to the distribution of a/b.[8]

Distribution sampling is used to evaluate features of a statistical distribution by representing them numerically and drawing observations from that numerical distribution. Kleijnen (1974: 32-33) presented an example where the Monte Carlo procedure can be applied to estimate the probability that a stochastic variable y is smaller than some constant r, or

$$p = P(y < r) \qquad (3.8)$$

It is assumed that output y is a function of inputs y_1 and y_2 specified as follows:

$$y = min\ (y_1, y_2) \qquad (3.9)$$

[8] Morgenthaler, G. W. (1961), pp. 370.

while y_1 and y_2 are assumed to have independent normal distributions with mean 100 and variance 400 and mean 90 and variance 100 respectively, that is,

$$y_1 \sim N(100, 400) \tag{3.10}$$

$$y_2 \sim N(90, 100) \tag{3.11}$$

For Monte Carlo estimation of p, it is convenient to introduce the variable x where:

$$\begin{aligned} x &= 1 \quad \text{if} \quad y < r \\ &= 0 \quad \text{if} \quad y \geq r \end{aligned} \tag{3.12}$$

The expected value of x is equal to p defined in (3.8) for

$$\begin{aligned} E(x) &= 1 \cdot P(y < r) + 0 \cdot P(y \geq r) \\ &= P(y < r) \\ &= p \end{aligned} \tag{3.13}$$

Assume an independent variable v which has mean 0 and variance 1 and if the central limit theorem[9] holds, this v is normally distributed, where

$$v = \sum_{i=1}^{12} k_i - 6 \tag{3.14}$$

where k is a random number and has mean 0.5 and variance 1/12, and i is the number of times the random number k is added. In this example, i equals to twelve, assuming that twelve is high enough to yield an approximately normally distributed variable. The Monte Carlo procedure is applied by generating twelve new random numbers in (3.14) to produce a new independent value of v.

If v has density $N(0,1)$, then $\sigma v + \mu$ has density $N(\mu,\sigma^2)$. In order to generate y_1 and y_2 defined in (3.10) and (3.11), the following are defined:

$$y_1 = 20v + 100 \qquad (3.15)$$

$$y_2 = 10v + 90 \qquad (3.16)$$

To keep y_1 and y_2 independent, different values of v in (3.15) and (3.16) are used. For every v substituted into (3.15) and (3.16), each set of y_1 and y_2 derived are substituted into (3.9) to determine the value of y. Equation (3.12) is used to evaluate the value of x. As a result of (3.13), p is estimated by the following

$$\hat{p} = \bar{x} = \sum_{i=1}^{N} x_i / N \qquad (3.17)$$

where N is the number of times when new values x are obtained.

It is important that the distribution of the input variables y_1 and y_2 is given for the Monte Carlo estimation of p. However, it is not necessary that this distribution has the form assumed in the above problem formulation. Two independent normal distributions are specified to simplify the analytical determination of p which was used to check the Monte Carlo estimate of p.

Following a similar process as the above example, the Monte Carlo procedure can investigate the robustness of some statistics. The more robust a statistic is, the less sensitive this statistic is to violations of any underlying assumptions. For example, the Monte Carlo procedure has been used to examine the distribution of t for nonnormal observations on x_i, where

---

[9] The *central limit theorem* which implies that the sum of a "large" number of independent stochastic variables with the same distribution and a finite mean and standard deviation is approximately normally

$$t = \frac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}} \quad (3.18)$$

Distribution sampling is also used to evaluate the "performance" of various regression analysis techniques for econometric models. Data used in these studies are generated randomly. Using these data, various regression techniques are applied to estimate the parameters of a model and the distributions of the resulting parameter estimators are compared to each other. Johnston (1973) describes several Monte Carlo studies on the performance of ordinary least squares (OLS), two-stage least squares (2SLS), limited information single equation (LISE) and full-information maximum likelihood (FIML) estimators applied to econometric model. Sample sizes in the range 15 to 40 are generated to reflect the small sample sizes which econometricians have to work with, especially in time series analysis. Applying the Monte Carlo procedure, arbitrary values are specified for the exogenous variables and these exogenous variables are combined with the disturbance values for each sample to generate values for the endogenous variables. The four estimating methods (OLS, 2SLS, LISE, and FIML) under study are then applied to each sample set of endogenous and exogenous values. This process is replicated a large number of times, and the resultant frequency distributions of estimates are studied in conjunction with the true values of the parameters in order to conjecture the small-sample properties of the estimators. Three criteria are used to study the parameters: bias, variance, and mean square error. These three criteria will be defined in section 3.2.

---

distributed.

In simulation, an experiment is carried out by means of an abstract model, which involves real people, and the behaviour of this model is followed over time. One kind of simulation is *operational gaming* where there is some form of conflict of interest among players or human decision-makers within the framework of the simulated environment The experimenter, by observing the players, may be able to test hypotheses concerning the behaviour of the individuals and/or the decision system as a whole (Morgenthaler, 1961). For example, in business games, simulation has been applied to determine the capacity of new truck-dock facilities at a central warehouse which is required to handle volume of truck traffic that was formerly accommodated at separate warehouse facilities. Truck arrival-time and service-time data were gathered and a program was prepared for computing the length of waiting-line and amount of service for various numbers of docks[10].

Apart from business games, simulations is also applied to military games. Military simulations were among the first of the large-scale digital simulations and are probably not more widely known because of security restrictions. Applications by almost all military study groups and weapons-system manufacturers range from vulnerability and armament-effectiveness studies for particular devices to large-scale studies dealing with the mechanics of recovery of the U.S. economy and estimates of civil defense capability subsequent to an all-out nuclear attack. One example of applying simulation to military games is the determination of the number of missiles needed to meet missile-site operational requirements. A certain number of missiles must always be ready to fire, and

---

[10] Morgenthaler, G. W. (1961), pp. 395.

yet periodic maintenance and checking are needed. Graphical model are used to simulate the flow of missiles through the maintenance cycle, with the position and status of each missile continuously recorded along a time scale. [11]

While simulation has been used extensively by management scientists, economists have also applied simulation technique in building economic models. An economic model consists of regression equations that do not include stochastic error terms. When estimating parameters of the regression equations these error terms are accounted for. However, once the parameters are estimated, the error terms are suppressed. The endogenous variables are calculated from the exogenous variables and the same set of error terms without using any random numbers.[12]

The Monte Carlo procedure is also used to compare economic models and estimators, and to analyze the convergence in the distribution of the estimators. Duan, Manning, and Rogers (1987) apply the Monte Carlo technique to analyze and compare the sample selection model with the two-part model. In this experiment, two different specifications are examined: in the first specification, there is only one regressor, y_1, and it appears in both equations with each coefficient equal to 1; in the second specification, there are two regressors, y_1 and y_2, and the variable y_1 enters only the probit equation, while the second equation contains only y_2. In both specifications, y_1 and y_2 are each drawn randomly from a uniform distribution. By specifying the coefficients to be equal to 1, the Monte Carlo procedure is applied. The y's are centered to yield three different probabilities of a positive outcome: 25, 50, or 75 percent. In each sample, there are 1,000

---

[11] Morgenthaler, G. W. (1961), pp. 393-94.

observations, and following the Monte Carlo procedure, the errors are randomly drawn from a bivariate normal distribution with variance 1 for each error term, and correlation coefficients equal to either +0.50 or +0.90. Each model is evaluated in terms of mean raw bias, mean squared raw error, and relative bias for the same set of design points. Another area of interest in this study is the prediction capability of these models throughout the range of the independent variables. For example, they found that when a model is biased upward by 10 percent in the first half of the data, and downward by 10 percent in the other half of the data, the model exhibits a disturbing bias pattern that cannot be detected in the average bias.

The analysis of the convergence in the distribution of estimators can be performed by the Monte Carlo procedure. Hiemstra and Kelejian (1991) apply the Monte Carlo procedure to specify a model in which the convergence in the distribution of the estimators are analyzed by the maximum likelihood methods. The specification of the model is associated with 'rare' events which can be related to oil spills, unemployment duration and child abuse. Such 'rare' events are sometimes unrecorded, and when recorded, their magnitudes are sometimes uncertain. Their results are based on 1,000 replications of each experiment and the measures used to assess the estimators are: the sampling means, standard deviations of the estimators, and the median of the asymptotic standard deviation as determined from an information matrix. The proportion of replications for which the absolute difference between the estimator and the true value of the parameters is less than a certain magnitude, and the Kolmogorov-Smirnov test

[12] Kleijnen, J. P. C. (1974), pp. 13-14.

statistic are used to investigate the consistency and asymptotic normality properties of the estimators.

The Monte Carlo procedure is applied in many studies to generate random numbers. In an attempt to assess the accuracy and reliability of generalized cost functions, where these functions are used to estimate the underlying technology and shadow prices, Parker (1994) uses data for which the "truth" is known. He claims that using real data is not helpful in determining the accuracy of the estimator, since the data generating process is not known. Therefore, he applies the Monte Carlo approach to generate data, and then estimates the underlying parameters as if they are their true values. This experiment is based on 1,000 trials of 100 observations of positive independent variables which are randomly and uniformly generated, and then normalized to a unit mean. For each trial, new coefficients are randomly and uniformly generated around some values and these coefficients are compared to the parameter point estimates. Other parameters such as the elasticity of substitution, the shadow price ratio, and the unknown input share are varied significantly in order to assess their marginal impact on inaccuracy. The error terms are assumed to be normally distributed, with mean zero and respective standard deviations; standard deviations are varied uniformly around 0.1 to assess their effects. To compare the performance of the mean calculated elasticity of substitution and shadow price ratio to the "true" values, log normalized square errors are calculated for each estimation. The mean errors, which is averaged over the 100 observations for each trial, are put into positive and proportional terms, and their logarithms are taken to derive a measure with an approximately normal distribution.

3.2 Methodology

A Monte Carlo experiment is an artificial controlled experiment that allows one to check if regression analysis is giving consistent estimates of the parameters. An estimate is considered a consistent estimator if it approaches the true value as the sample size gets larger. A simple possible Monte Carlo experiment can be described in a few steps.[13]

Consider a model in the form of:

$$Y = \beta_0 + \beta_1 X + u \tag{3.19}$$

First, the true values for β_0 and β_1 and the values of X in each observation are chosen. The values of X are chosen using the random number generating process, and the true values of β_0 and β_1 are chosen arbitrarily. Second, for the disturbance term, the random number generating process is used to provide this random factor in each observation. Third, the values of Y in each observation are generated using the relationship (3.19) and the values of β_0, β_1, X and u. Fourth, using only the values of Y thus generated and the data for X, a regression analysis is used to obtain estimates of the parameters. These new parameters will be called α_0 and α_1 in place of β_0 and β_1 respectively. A test of significance of the estimates, such as a simple *t*-test, can be conducted to examine if α_0 is a good estimator of β_0 and if α_1 is a good estimator of β_1, and this can provide some idea of whether the regression technique is working properly.

To obtain β_0 and β_1, many researchers (Manning, Duan, and Rogers, 1987, and Smith, 1971) have arbitrarily assumed the true values. In this thesis, the true values are

[13] These steps are replicated from Dougherty, C. (1992), pp. 76-80.

obtained by ordinary least squares (OLS) estimation using the total sample size of 12,680 observations, assuming this is the population size. This method of obtaining the true values of the parameters provides a good guideline as to where the actual true values will lie. Moreover, these values can also be considered as arbitrary because these number can be assumed by anyone. Using smaller sample sizes, many new parameter estimates of α_0 and α_1, are obtained. According to Dougherty (1992), a simple *t*-test can be conducted to examine if α_0 is a good estimator of β_0 and if α_1 is a good estimator of β_1. But some may argue that it is incorrect to test if α_i is a good estimator of β_i since the smaller sample size is a subsample of the total sample size (or the assumed population size) and they are not independent of each other. However, as mentioned, the values of β_i can be considered as arbitrary and the problem of independent samples no longer exists.

Besides deriving the true values of the parameters and controlling the distribution of the error term, the Monte Carlo experiment can also address a few issues which can be investigated:

(1) the effect of parameters' stability as sample size changes, and

(2) the convergence in the distribution of the estimated parameters

In studying individual parameters three important criteria are usually distinguished: bias, variance, and mean square error.[14] Bias is the difference between the mean of the sampling distribution of estimates and the true value. If β denotes the true

[14] In the literature of Monte Carlo, many researchers, such as W.G. Manning *et al* and V.K. Smith, have used these criteria in evaluating their studies.

value of the parameter and $\bar{\hat{\beta}}_R$ the mean of the sampling distribution of estimates for sample size R, then

$$\text{Bias} = \bar{\hat{\beta}}_R - \beta \tag{3.20}$$

The true bias of the true value of the parameter is zero. Therefore, to test for the significance of the bias of the mean of the sampling distribution of estimates for sample size R, $\bar{\hat{\beta}}_R$, a t-test is used and the null and alternative hypotheses are given by

H_0: The true bias equals to the bias in (3.20).

H_A: This true bias does not equal to the bias in (3.20).

For a sample size R, the sample variance of N estimates of β_j is the variance of each β_j estimate around their mean (Griffiths, Hill, and Judge, 1993), that is,

$$\bar{v}^2_{\hat{\beta}_j} = \frac{\sum_{j=1}^{N}(\hat{\beta}_j - \bar{\hat{\beta}})^2}{N-k} \tag{3.21}$$

where N is the number of iterations and k is the number of parameters. This is compared to the true sampling variability of β. The variance of β is given by

$$v2 = \frac{\sigma^2}{R} \tag{3.22}$$

where σ^2 is the residual sum of squares or the population variance. This variance, σ^2, is obtained from the OLS estimation with sample size 20,160. To test for the significance of the variance of the parameter $\hat{\beta}_j$, a chi-square test is used and the null and alternative hypotheses are given by

H_0: $v^2 = \bar{v}^2_{\hat{\beta}_i}$

H_A: $v^2 \neq \bar{v}^2_{\hat{\beta}_i}$

For N estimates of the sum of squares error, $\hat{\sigma}^2$, the average value is given by

$$\bar{\hat{\sigma}}^2 = \frac{\sum_{j=1}^{N} \hat{\sigma}_j^2}{N} \tag{3.23}$$

and the sample variance of this variance, $\hat{\sigma}^2$, is

$$\bar{v}^2_{\hat{\sigma}^2} = \frac{\sum_{j=1}^{N} (\hat{\sigma}_j - \bar{\hat{\sigma}}^2)^2}{N-k} \tag{3.24}$$

This $\bar{v}^2_{\hat{\sigma}^2}$ for a sample size R is compared to the variance of $v^2_{\sigma^2}$ using the variance of the true value of σ^2, that is,

$$v^2_{\sigma^2} = \frac{2\sigma^4}{(R-1)} \tag{3.25}$$

To test for the significance of the variance (v^2) of this variance (σ^2), a chi-square test is used and the null and alternative hypotheses are given by

H_0: $v^2_{\sigma^2} = \bar{v}^2_{\hat{\sigma}^2}$

H_A: $v^2_{\sigma^2} \neq \bar{v}^2_{\hat{\sigma}^2}$

Mean square error (MSE) is the variance of the estimates around the true value of the parameter being estimated, that is,

$$\text{MSE} = \frac{\sum_{j=1}^{N} (\hat{\beta}_j - \beta)^2}{N} \tag{3.26}$$

The rationale of the mean square error criterion is that estimated values near the true value are "good," those far away are "bad," irrespective of the direction of the discrepancy, and so all discrepancies are squared and averaged. The mean square error is equal to variance plus the square of the bias. Therefore, it is clear that a biased estimate may thus show a smaller mean square error than an unbiased one if it more than compensates for its bias by having a smaller variance.

Given N samples and the unbiased estimates of $\hat{\beta}_j$ and $\hat{\sigma}^2$, the N values obtained for these estimators can be expected to average out fairly close to the true underlying parameters, β and σ^2, respectively. To get an idea of how much $\bar{\hat{\beta}}_R$ can vary, the proportion of the estimates that fall inside or outside some particular ranges can be worked out using the normal distribution which can be computed as follows:

$$P(\text{L.L.} < \bar{\hat{\beta}}_R < \text{U.L.}) = P\left(\frac{\text{L.L.}-\beta}{\text{se}(\bar{\hat{\beta}}_R)} < \frac{\bar{\hat{\beta}}_R-\beta}{\text{se}(\bar{\hat{\beta}}_R)} < \frac{\text{U.L.}-\beta}{\text{se}(\bar{\hat{\beta}}_R)}\right)$$

$$= P\left(\frac{\text{L.L.}-\beta}{\bar{v}_{\hat{\beta}_j}} < Z < \frac{\text{U.L.}-\beta}{\bar{v}_{\hat{\beta}_j}}\right) \qquad (3.27)$$

where $Z = \dfrac{\bar{\hat{\beta}}_R-\beta}{\text{se}(\bar{\hat{\beta}}_R)}$ is a standard normal random variable with mean zero and variance one, and L.L. and U.L. are the lower limit and upper limit of the true parameter β. These limits can be defined as

$$\beta \pm \text{s.e.}(\beta)\, z_{\alpha/2} \qquad (3.28)$$

The result can be interpreted as the probability that β_R will be expected to lie within these limits, or that it would lie outside this range with the probability 1 - P(L.L. < β_R < U.L.).

In addition, two more measures are used in evaluating the performance of the estimator: relative precision and decentralization. Relative precision is the ratio of the mean of the sampling distribution to the root mean squared error. Decentralization is used for the number of times the estimated coefficient has the wrong sign (Smith, 1971).

The convergence in the distribution of the estimated parameters can be observed from graphs which plot the different sample sizes against the biases for the respective sample sizes. From these graphs, one can observe the sample size when the biases approach zero or become statistically insignificant, that is when the estimated parameter approaches its true value. When this sample size is reached, it will be the minimum sample size required when using the labour supply function.

3.3 Model Specification

In order to apply the Monte Carlo procedure, a model is required. This section specifies the model for this experiment, which is a labour supply equation that is based on a semi-log functional form and is specified as:

$$H_i = \beta_0 + \beta_1 \ln(GW_i) + \beta_2 M_i + \beta_{ij} X_{ij} + u_i \qquad (3.29)$$

where H_i = total number of hours of paid employment in 1986

GW_i = hourly gross wage rate

X_{ij} = a vector of control variables (including age, years of schooling, number of children less than six, number of children between five and nineteen)

β_0, β_1, β_2, β_{ij} = parameters to be estimated

M_i = the inverse Mills ratio calculated from a probit equation predicting quantity constraint

u_i = a random error term

A more detailed discussion of the variables are presented in Chapter 4.

Several factors determined this choice of functional form. Most importantly, these, or similar models, are those most frequently found in the literature pertaining to labour supply. Other models derived explicitly from a specification of the preference function could be analyzed, but the introduction of a new or rarely encountered functional form would introduce an additional source of discrepancies with previous studies. Stern (1986) provides the specification of the indirect utility function which yields the labour supply function used in this paper. Furthermore, the linearity in the parameters allows for relatively simple estimation schemes and makes possible extensive testing of hypotheses under consideration.

There are other applied labour economists who treat wage rate as an endogenous variable. Mroz (1987) found that the use of a woman's wage rate on her current (time of interview) job as an instrumental variable to control for measurement error in the hourly earnings induces a positive bias in the estimated wage effect. This suggests that the average hourly earnings measure combines measurement error and wage rate endogeneity, resulting in ordinary least squares point estimates that are not significantly different from the two-stage least squares estimates which control for these two sources of potential bias.

On the other hand, Blundell, Duncan, and Meghir (1992) find that the tax system is the major explanatory factor determining endogenous net wage rates and that the gross wage is exogenous. Hausman (1979) and Leuthold (1978) also treated wage rates as an exogenous variable.

3.4 Sample Selection Rule[15]

Randomly selected sample taken from the entire population are sometimes used in estimating labour supply functions, but in many cases either choice or necessity dictates that something other than a pure random sample be used. Sometimes, the investigator wants to study only the labour supply behaviour of a population sub group. For example, these groups might include poor people, married women, teenagers, or men. In other cases, data are available only for a particular population subgroup. For example, survey data sets sometimes refer only to persons below "poverty" level of income, wage data are usually available only for persons currently at work, and so on. No particular difficulties arise if the subgroup is selected according to endogenous factors -- that is, factors that may be determined along the labour supply, such as current earnings or current income. If they are not, then a variety of complications may arise.

[15] Killingsworth (1983), pp. 78-87.

The basic reason why these complications arise is that when a sample is selected on the basis of endogenous factors the error term may not be a mean-zero random variable in the resulting sample even if it is a mean-zero random variable in the population as a whole. As a result, direct application of simple regression methods such as ordinary least squares (OLS) may be inappropriate because they provide valid results only if the error term is a mean-zero random variable within the sample. Rather, simple regression will suffer from what is known as *sample selection* (or *selectivity*) *bias.*

In practice sample selection bias may arise for two reasons. First, there may be self selection by the individuals or data units being investigated. For example, individuals can choose to be in the workforce or not. Thus, the wage rates of persons who are not at work are not observed. Second, sample selection decisions by analysts or data processors operate in much the same fashion as self selection. For example, analysts or data processors can choose to include or exclude persons who are not working when fitting a labour supply regression. The decision will result in sample selection bias.

Under the sample selection rule for the labour supply function, Killingsworth[16] noted that one should include an observation *i* in the sample to be analyzed, if and only if $R_i > T_i$, where R_i is the value of a sample selection variable, and T_i is a particular cutoff value.

Consider the following simple regression equation:

$$H = \alpha + \beta X_i + \varepsilon \qquad (3.30)$$

where H = total number of hours worked

[16] Killingsworth (1983), pp. 83 - 85.

X_i = independent variables

α, β = parameters to be estimated

ε = random error term

Suppose one selects a regression sample on a purely random basis, so that the sample selection variable *R* is a randomly assigned number. In this case, then, *R* is necessarily uncorrelated with ε, so that $E[\varepsilon \setminus X_i]$ is zero. In other words, if *R* is a random number (which means that *R* is uncorrelated with ε), then $E[\varepsilon \setminus X_i, R > T] = E[\varepsilon \setminus X_i]$, and by assumption, $E[\varepsilon \setminus X_i] = 0$ for each and every observation in the population. Here, the application of least squares is entirely appropriate.

Suppose that the sample selection variable *R* is some variable that is exogenous to labour supply, in the sense that this *R* is uncorrelated with ε. (Examples of exogenous *R* variables might be age, sex, or race; increases in ε cannot be associated with increases in any of these variables, since they are all fixed). For example, suppose that *R* refers to age and that *T* is 44 for all observations, so that the selection rule is "select if and only if the observation is older than 44 years of age." This kind of selection is certainly not "random"; indeed, it is highly "systematic" and "nonrandom," as understood in day to day english usage. However, this kind of selection is nevertheless random with respect to labour supply, in the sense that such an *R* is uncorrelated with the labour supply error term ε and is therefore exogenous to labour supply. Thus, for this kind of selection rule as well as for random selection, $E[\varepsilon \setminus X_i, R > T] = E[\varepsilon \setminus X_i]$, and, by assumption, $E[\varepsilon \setminus X_i] = 0$ for each and every observation in the population. Again, then, application of least squares is entirely appropriate.

Suppose that the sample selection variable is endogenous to labour supply, in the sense that R is correlated with the labour supply error term ε. One example will be that R = H and $T = 0$. Here the selection rule will select only persons who are working. In this case, it is obvious that $E[\varepsilon \setminus X_i, H > 0] = E[\varepsilon \setminus X_i, \varepsilon > -(\alpha + \beta X_i)]$ will be nonzero for each observation in the population.

The nature of OLS bias has been considered by, among others, Goldberger (1981) and Greene (1981). Under the assumption that all independent variables and the dependent variable are multivariate normally distributed in the population (this therefore excludes dummy variables as regressors), Goldberger obtains strong result that the OLS regression coefficients are biased downward in the sense that the OLS coefficient is a scalar multiple of the "true" labour supply coefficient vector, where the scalar lies within the 0-1 interval. Moreover, Greene shows that to obtain consistent estimates of the true labour supply parameters in equation (3.30), all one need do is divide each element of OLS coefficient vector by the proportion of observations for which H > 0.

Although Goldberger shows analytically that this remarkable result does not hold when the multivariate assumption is relaxed, Greene finds that with a variety of nonnormal situations (including dummy variable regressors), this simple adjustment of OLS estimates provides a surprisingly robust approximation to maximum likelihood estimates of the labour supply parameters in equation (3.30). The OLS-based estimates of standard errors, however, are inconsistent, and as Greene shows, they cannot be easily adjusted. Hence, one possible way of dealing with sample selectivity is to do probit estimations of the labour force participation decision based on all observations, then do an OLS estimation of the hours worked equation, limiting sample to workers, and finally, to obtain consistent estimates of the hours worked parameters by multiplying the OLS coefficient vector by the sample proportion of observations for which H > 0.

There is now general agreement about the potential importance of the missing wage problem -- where wages are missing for nonworkers. There is less agreement about the particular solutions to the wage-imputation problem that have been offered in the literature. Each solution invokes different assumptions about unobserved counterfactuals: what wages would have been had nonworking persons worked. Different assumptions are likely to be appropriate for different problems and data sets.

3.5 Incorporation of Taxes

Taxes on wage and property income can significantly affect labour supply, in part because taxes often introduce a wedge between average and marginal after-tax wage rates. Moreover, much public policy discussion has focused on the labour supply effects of changing statutory provisions of the income tax code. Second-generation research on this topic has been substantial, with noteworthy contributions made by H. S. Rosen (1976), G. Burtless and J. A. Hausman (1978), Hausman (1979; 1980; 1981a, b; 1985), T. J. Wales and A. D. Woodland (1979), J. J. Heckman and T. E. MaCurdy (1980), and A. Nakamura and M. Nakamura (1981).[17] Studies have suggest that while the presence of taxes considerably complicates the analysis of labour supply, empirical analysis can

[17] Killingsworth (1983), pp. .

proceed provided one is willing to make tradeoffs concerning computational complexity and flexibility of the functional form.

Koster's (1967) pioneering study of the effects of taxes on labour supply found very weak effects for male hours-of-work equations for those who are working and somewhat stronger, but still small, effects on participation. His estimates are confirmed by MaCurdy et al. (1990). Mroz (1987) finds similar weak tax effects on female hours of work for working women.

3.6 The Application

This section presents the application of the Monte Carlo procedure on the model of this experiment, which is the labour supply equation. Recall the labour supply function from section 3.3:

$$H_i = \beta_0 + \beta_1 \ln(GW_i) + \beta_2 M_i + \beta_{ij} X_{ij} + u_i \qquad (3.31)$$

where H_i = total number of hours of paid employment in 1986

GW_i = hourly gross wage rate

X_{ij} = a vector of control variables (including age, years of schooling, number of children less than six, number of children between five and nineteen)

$\beta_0, \beta_1, \beta_2, \beta_{ij}$ = parameters to be estimated

M_i = the inverse Mills ratio calculated from a probit equation predicting quantity constraint

u_i = a random error term

The inverse Mills ratio is calculated from the probit model and then substitute back into the above equation as an explanatory variable. To obtain a clearer understanding of how the probit analysis is applied to the labour supply function, the labour supply function can be written as:

$$H_i = \beta_0 + \beta_1 \ln(GW_i) + \beta_2 X_{i2} + u_i \quad \text{if } H_i = \beta_0 + \beta_1 \ln(GW_i) + \beta_2 X_{i2} + u_i > 0$$

$$H_i = 0 \quad \text{if } H_i = \beta_0 + \beta_1 \ln(GW_i) + \beta_2 X_{i2} + u_i \leq 0$$

If only workers are observed, that is where $H_i > 0$, then

$$E\,[H_i / H_i > 0] = \beta_0 + \beta_1 \ln(GW_i) + \beta_2 X_{i2} + k_i$$

where $k_i = E\,[u_i / H_i > 0] = \sigma \dfrac{f(z_i)}{F(z_i)}$

$$= \sigma \lambda_i$$

where λ_i is the inverse Mills ratio,

σ is the standard error of the regression,

$f(z_i)$ is the standard normal density function,

$1 - F(z_i)$ is the proportion of the sample with $H_i > 0$, and

$F(z_i)$ is the cumulative normal density function.

The probit model in this thesis calculates the probability that an individual is quantity constrained, that is, would prefer additional weeks of work. Thus, the dependent variable (H_i) from the above example is replaced by the a dummy variable which determines if an individual is underemployed or not. This dummy variable will be further explained in details in the next chapter. This way of allowing for the possibility that individuals may face quantity constraints which prevent them from working all the hours which they want, at prevailing wage rates is outlined by Ham (1982). When those who are not satisfied with their total weeks of work are identified from the data, individuals who are not constrained in weeks of work can be considered as a selective sample from the labour force. This form of sample selection bias is corrected by introducing the inverse Mills ratio in the labour supply function.

However, it must be noted that in the case of the probit two-stage method, the standard errors from the second stage explicitly underestimate the correct standard errors.[18] This conclusion is supported by Lee et al. (1980) who compared the correct standard errors with the incorrect standard errors from the second stage (OLS) of the probit two-stage method. An alternative two-stage procedure is suggested by Heckman (1980) whereby estimates of the standard normal density function and the cumulative normal density function are obtained. This procedure is equivalent to the inverse Mills ratio. The inverse Mills ratio is substituted into equation like (3.31) which is then estimated by OLS. This will produce consistent estimates of the parameters. In short, the parameter estimates obtained from the second stage of the probit two-stage method as suggested by Heckman (1980) are biased but consistent.

The data consist of 20,160 observations, including both workers and nonworkers. Only seven out of the twenty independent variables from the probit model have been used in the semi-log labour supply equation, which is estimated by OLS. This is to prevent any collinearity problem. The independent variables which appear in both stages are the regional variables and the age variables (all are dummy variables). Collinearity problems

[18] Maddala, G. S. (1983) pp. 238.

exist when the independent variables calculated from the probit are exactly the same as the next equation to be estimated in the OLS. In this case, the next equation will have collinearity problem between the wage variable and the inverse Mills ratio. The exclusion of variables in the probit model from the labour supply function also satisfy the condition for identification in Heckman's model (Maddala, 1983: 233).

After using the probit analysis to derive at the inverse Mills ratios, linear regression is conducted on the following equation:

$$\begin{aligned} H_i = {} & \beta_0 + \beta_1 \ln(GW_i) + \beta_2 AT_i + \beta_3 PQ_i + \beta_4 PR_i + \\ & \beta_5 BC_i + \beta_6 AGE35_i + \beta_7 AGE45_i + \beta_8 U_i + \beta_9 PC_i + \\ & \beta_{10} DC_i + \beta_{11} S_i + \beta_{12} EE_i + \beta_{13} PS_i + \beta_{14} PD_i + \\ & \beta_{15} UNI_i + \beta_{16} JT_i + \beta_{17} M_i + u_i \end{aligned} \tag{3.32}$$

where H_i = total number of hours of paid employment in 1986

GW_i = hourly wage rate

AT_i = Resident of Newfoundland, Prince Edward Island, Nova Scotia, or New Brunswick

PQ_i = Resident of Quebec

PR_i = Resident of Manitoba, Saskatchewan, or Alberta

BC_i = Resident of British Columbia

$AGE35_i$ = Males who aged between 35 and 44

$AGE45_i$ = Males who aged between 45 and 54

U_i = Member of a union or wages were covered by collective agreement negotiated by a union

PC_i = Covered by a pension plan connected with the job(s)

DC_i = Number of dependent children (own or others) under 15 years of age

S_i = Males who are not married

EE_i = Obtain none or elementary education

PS_i = Obain some post-secondary education

PD_i = Obtain post-secondary certificate or diploma education

UNI_i = Obtain university education

JT_i = Job tenure (stop week minus start week)

M_i = the inverse Mills ratio calculated from a probit equation predicting quantity constraint

u_i = a random error term

This equation is estimated using the ordinary least squares (OLS) with the total sample size of 12,680 for male workers. Estimates of the parameters are obtained and they are assumed to be the true parameters of the model. Using the random number generating process in *Shazam* (1993), the random numbers for the disturbance term for each observation in the data are generated. Once all the required values on the right hand side of equation (3.32) are obtained, the values for the H_i can be generated.

Using the generated H_i and the data for w_i, X_{i2}, and M_i, equation (3.32) is estimated using OLS to obtain the estimates for the parameters. A test is then conducted to see if the latter are consistent estimators of the "assumed true" values of the parameters. Many tests can be conducted using the same sample size but with different random numbers for the disturbance term.

For the purpose of this experiment, the sample size will be reduced to a point where the generated parameters are not good estimators of the "assumed true" values of the parameters. Numerous experiments are performed for each subsample to avoid any bias constraints. Subsamples are randomly drawn using the random number generating process. In other words, the data in each subsample are always different for any two experiment.

CHAPTER 4: DATA

This chapter discusses the data used in this thesis. The data is obtained from the Labour Market Activity Survey (LMAS) of Statistics Canada. Section 4.1 discusses the methodology which the LMAS used to obtain the data. This section also includes a detailed discussion of the variables that are incorporated into the labour supply function; and the definition of the base period of the labour supply function to be estimated.

Section 4.2 provides a brief discussion on sampling error pertaining to the survey. The major sources of non-sampling error arise from total and partial non-response problem. A conclusion on whether total and partial non-response have contributed significantly to sampling error is then derived.

4.1 The LMAS and The Definition of Variables

There are a total of thirty variables in both the probit model and the labour supply equation, excluding the inverse Mills ratio. The data are drawn from the Labour Market Activity Survey (LMAS). The LMAS was conducted by Statistics Canada with the cooperation and support of Employment and Immigration Canada. The LMAS questionnaires collected information on the annual labour market participation of Canadians and the characteristics of up to five jobs held in each of the calendar years from 1986 to 1990. LMAS provides measures of the dynamic nature of Canadian labour market which are conceptually consistent with the Labour Force Survey (LFS). It also provides information on the characteristics of the paid jobs which are not available on the LFS.

The LMAS was administered to a sub-sample of the dwellings in the LFS sample, and therefore its sample design is closely tied to that of the LFS. The LFS is a monthly household survey whose sample of individuals is representative of the civilian, non-institutionalized population 15 years of age or older in Canada's ten provinces. Specifically excluded from the survey's coverage are residents from the Yukon and Northwest Territories, persons living on Indian Reserves, full-time member of the Canadian Armed Forces, and inmates of institutions. These groups together represent an exclusion of approximately two per cent of the population aged 15 or over. In the LMAS, persons 15 years of age and younger, and persons 70 years of age and older were not issued the questionnaires. This was to reduced respondent burden within the LMAS households surveyed.

Altogether there were 32,761 male respondents, aged between 15 and 69, for the reference year 1986. In our data, only male respondents who were in their prime age (that is age 24 to 54) were selected. It is known that most men in this age range are either in or want to be in the workforce. After selecting out of males from this age group, there were only 20,160 observations left.

The regional dummy variables are narrowed to only five main regions of residence. Respondents who were residing in Newfoundland, Prince Edward Island, Nova Scotia, or New Brunswick were categorized under the variable "Atlantic." Residents of Manitoba, Saskatchewan, and Alberta were categorized under the "Prairie" variable. For respondents who were residence of Quebec and British Columbia, they were categorized accordingly. Ontario does not appear as one of the variables because it is the base area for the regional dummy variables in the labour supply function. This is a rule of thumb to follow in order to prevent falling into the "dummy variable trap."[19] The data used were demographic data, and it is worth noting that the number of respondents from each province/region do not necessarily represent the population size in these provinces/regions. There is a weighting affiliated to each observation. Each observation of the dependent and independent variables is multiplied by the square root of the weight variable. The weights are normalized to sum to the number of observations.[20] Figure 4.1 shows the distribution of the male respondents, between 25 and 54 years of age, across the different regions. The figure shows that most of the male respondents come from the Prairie region and the Atlantic region since these two regions consist of a few provinces. Among the three provinces which stand on their own, most of the male respondents from this age range come from Ontario.

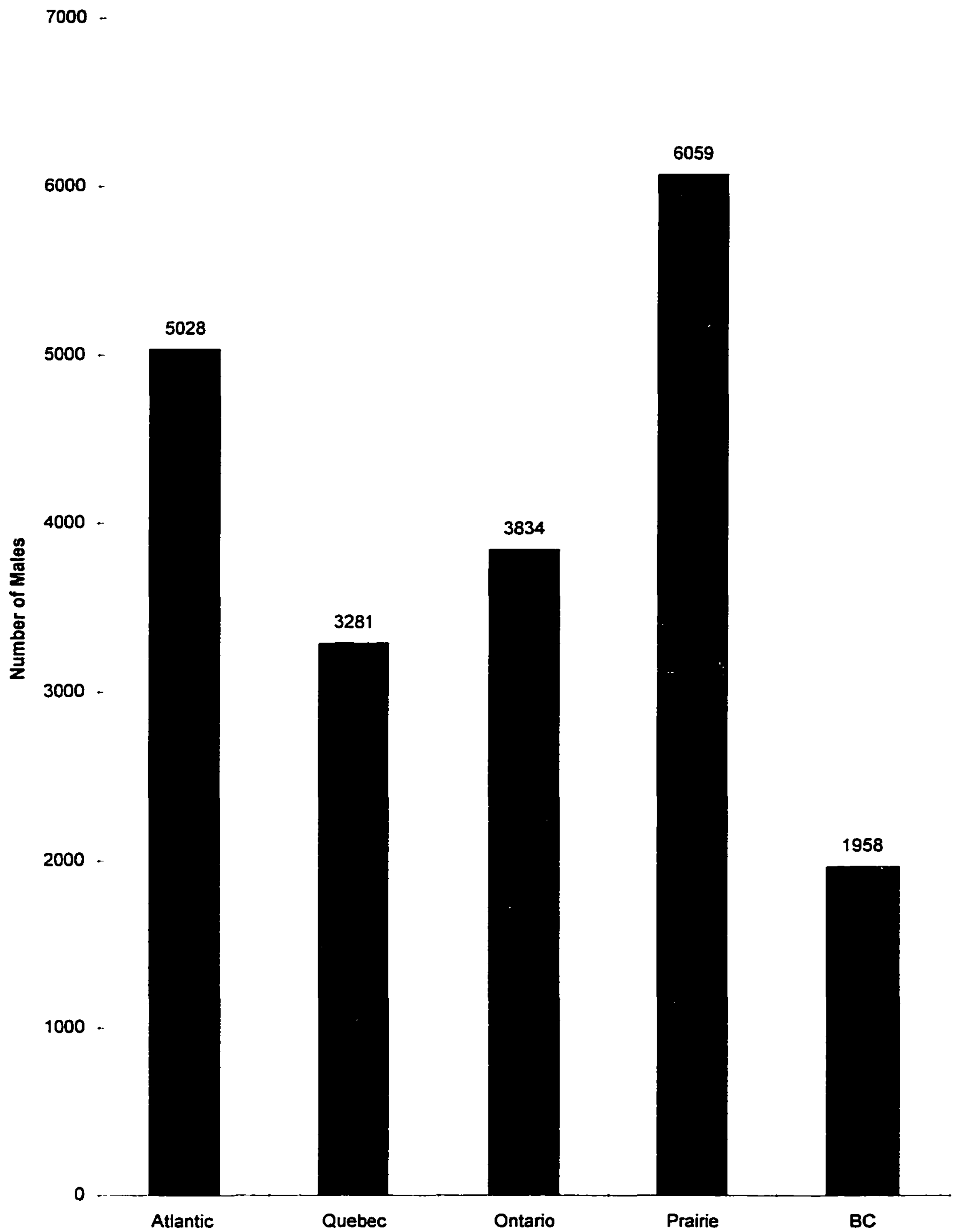
The data for 1986, which relates to 1986 calendar year, was collected in January, February, and March 1987. In the LMAS, "work" refers to any duties performed for pay or profit, including unpaid work in a family farm or business. Pay includes cash payment or "payment in kind", whether or not the payment was received in the year the duties were performed. Work included any periods of paid leave, paid sabbatical, paid sick

---

[19] To avoid falling into what might be called the dummy variable trap, the general rule is: If a qualitative variable has m categories, introduce only m - 1 dummy variables.

[20] White, Kenneth (1993). *Shazam User's Reference Manual Version 7.0*. pp. 77.

Figure 4.1: Distribution Of Male Age 25-54 Across Canada In 1986



leave, et cetera. In the LMAS, the respondent must have worked at least one day at a job or a business to be counted as having held a job in any given calendar year. The LMAS questionnaires collected detailed information on up to five different jobs held in a calendar year or reference year. Where a respondent held more than five jobs in either reference year, information was collected for only the first five jobs at which the respondent worked.

For each job identified in the LMAS, information was collected on the start date of the first spell of employment experienced in the reference year and the date most recently worked at the job, and, if applicable, the reason for leaving the job. In the terminology of the LMAS, an individual is said to have "held" the job between those two dates. In the LMAS, jobs could be held as little as one day. Self-employed respondents who performed a variety of duties during their period of employment under the same company name would count as one employer. For self-employed person with more than one business, each business was treated as a separate employer.

The LMAS collected sufficient information to assign a labour force status of employed, unemployed, or not in the labour force for each week of the reference year. The labour force status were assigned using a hierarchical logic similar to that used in the monthly LFS. Thus, weeks in which a respondent reported any work at any job were assigned a code of "employed." A code implicating that an individual was employed was also assigned in a small number of weeks where no work was reported because the respondent's usual monthly work schedule at a job did not involve work in every week of the month.

The marital status category of "single" means never married; "married" includes living in a common law relationship; and "other" includes widowed, separated or divorced. In our data, we are only interested in male respondents who fall under the category of "single" and age between 25 and 54. The distribution of single male, between 25 and 54 years of age, is depicted in Figure 4.2. When compared to Figure 4.1, both figures have the same distribution.

The variable which accounts for the number of dependent children refers to the number of sons or daughters be it natural, adopted, or step-children. In Figure 4.3, we compare the number of male respondents to those who have at least one dependent child. About 61.6% of the male respondents from the Atlantic have at least one dependent child. For the rest of the regions, about 54.1% of them have at least one dependent child.

In the LMAS, a job is classified as full-time if the usual monthly work schedule involve 120 or more hours, and part-time if the job would normally require less hours of work per month. A person may have more than one part-time job with a cumulative of 120 hours or more per month but they do not technically have a full-time job according to the LMAS definition. Figure 4.4 shows the number of males who worked less than 120 hours per month across the regions and Figure 4.5 compares the total number of male respondents, age between 25 and 54, to those who worked less than 120 hours per month within the respective regions. The latter shows that, on the average, about 4.2% of the male respondents from each region, in the same age group, are considered working part-time according to the LMAS description.

Figure 4.2: Distribution Of Single Male Age 25-54 Across Canada In 1986

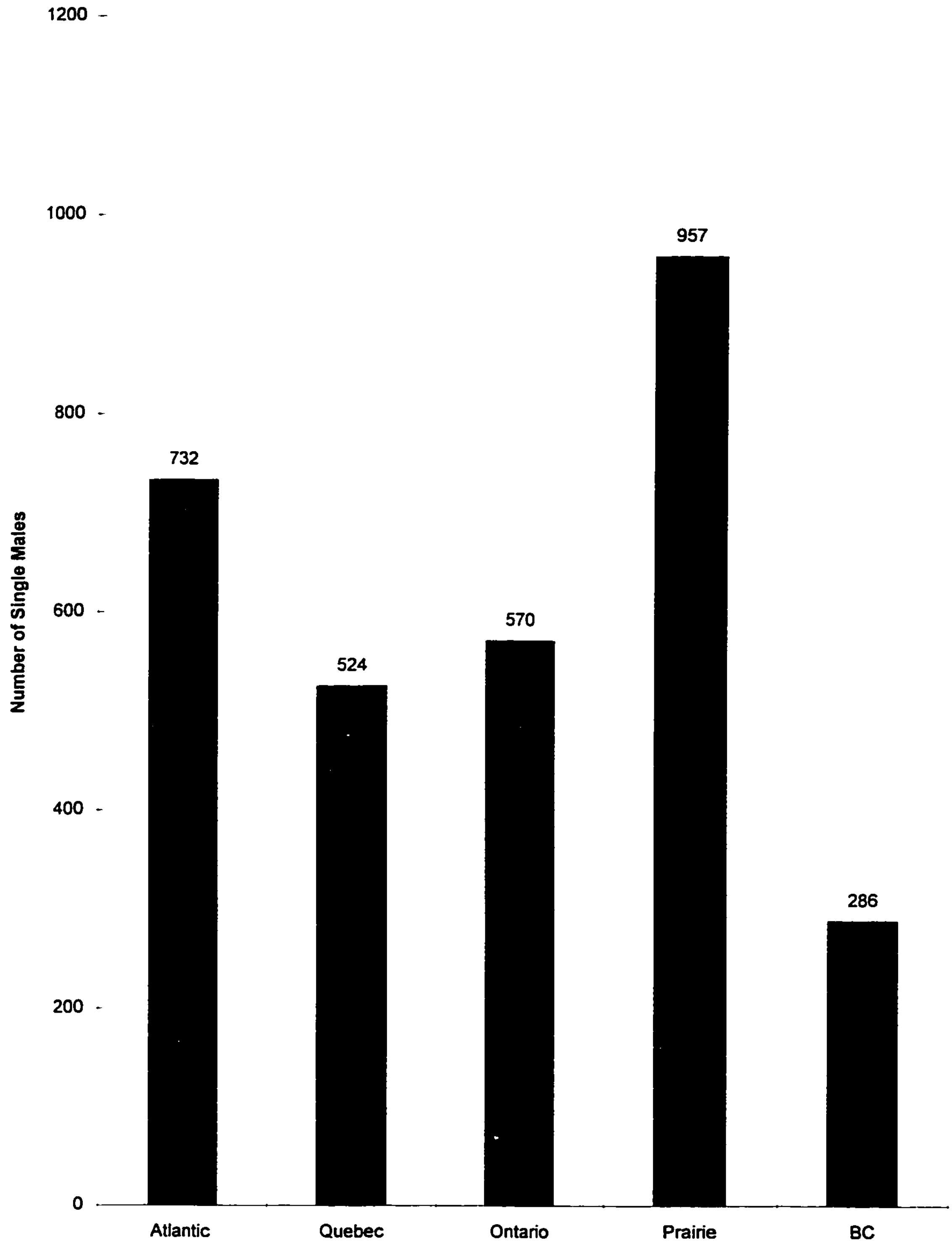


Figure 4.3: Comparison Between Those Who Have Dependent Children To The Total Respondents For Male Age 25-54 In 1986

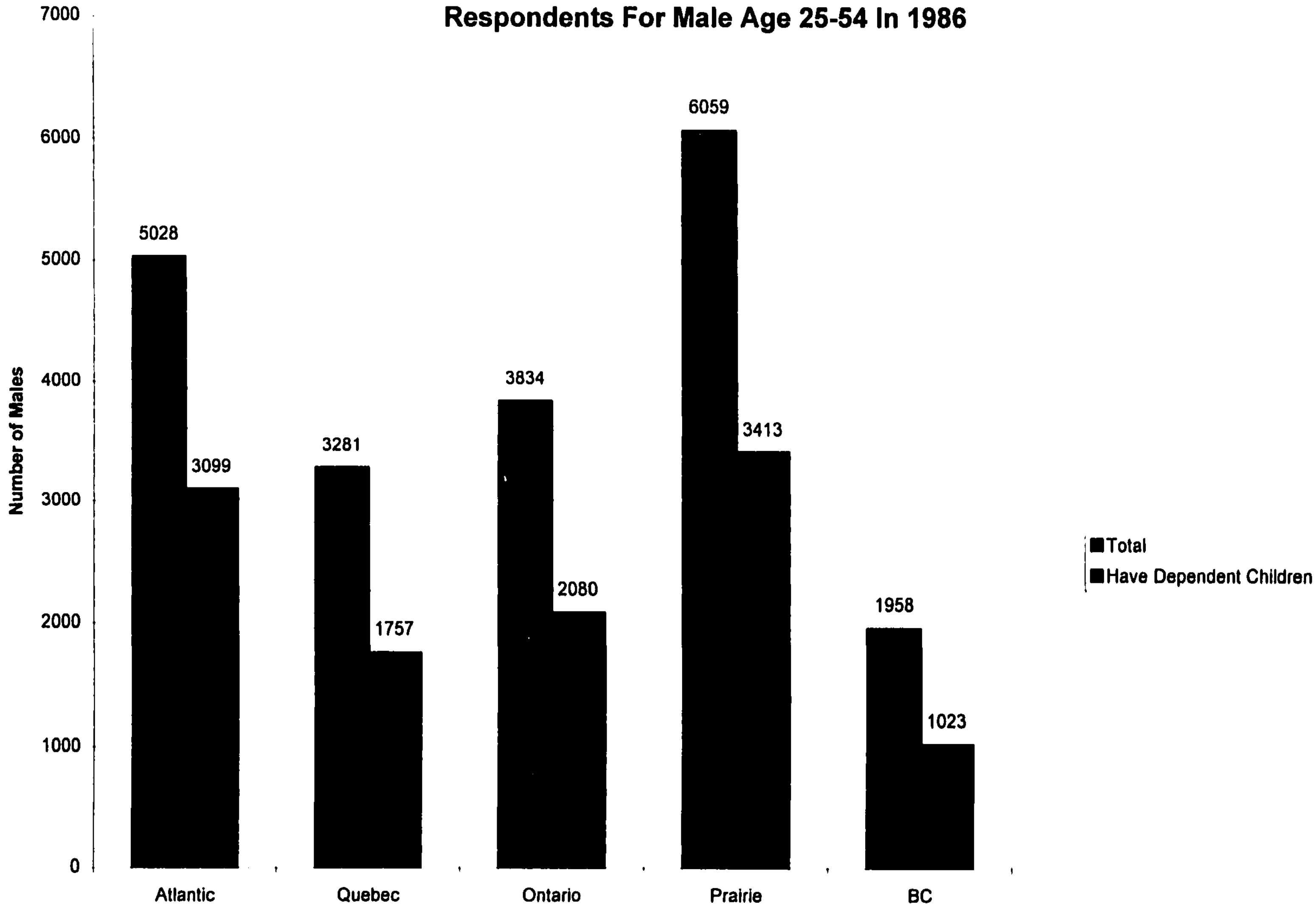


Figure 4.4: Male Age 25-54 Working Less Than 120 Hours Per Month In 1986

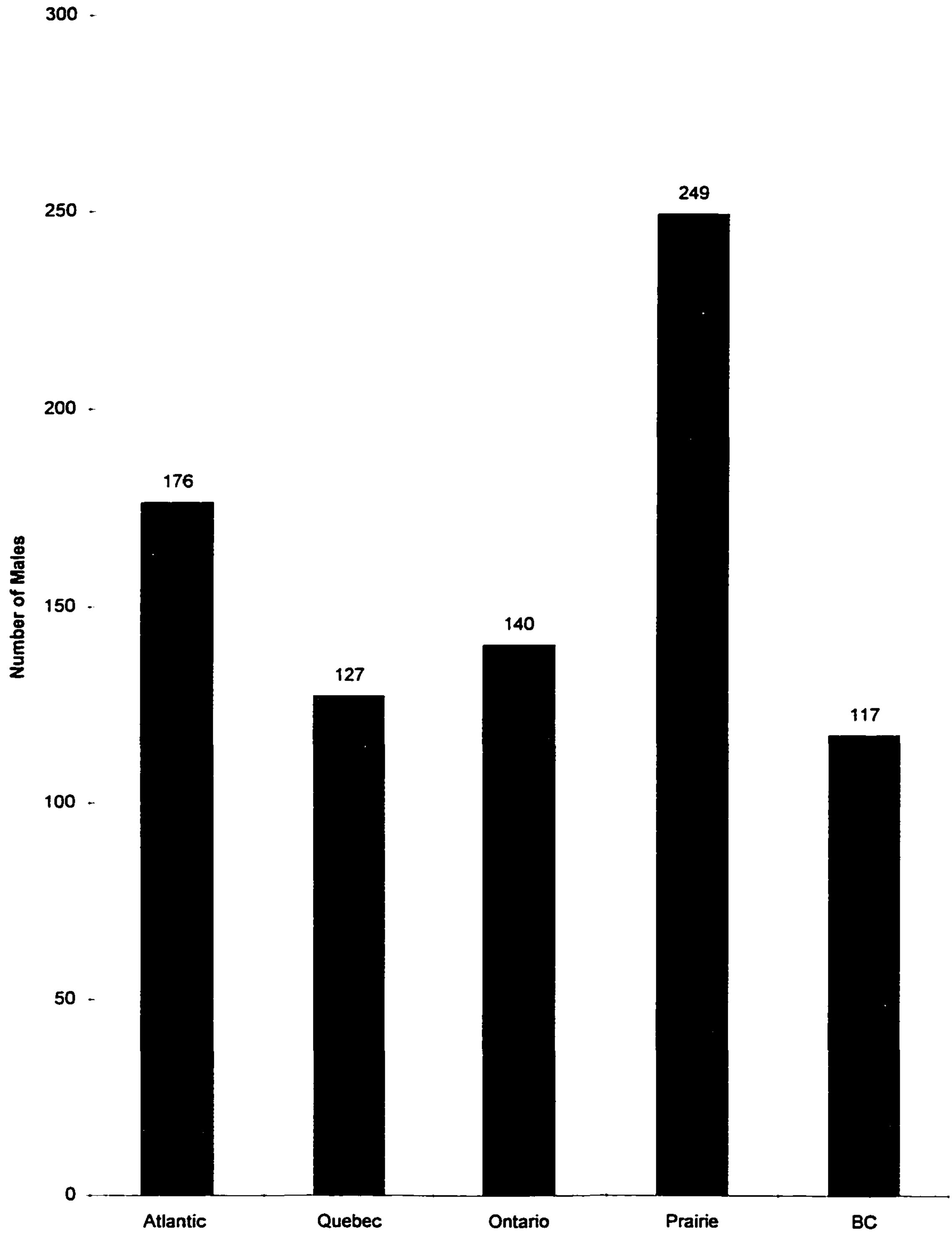
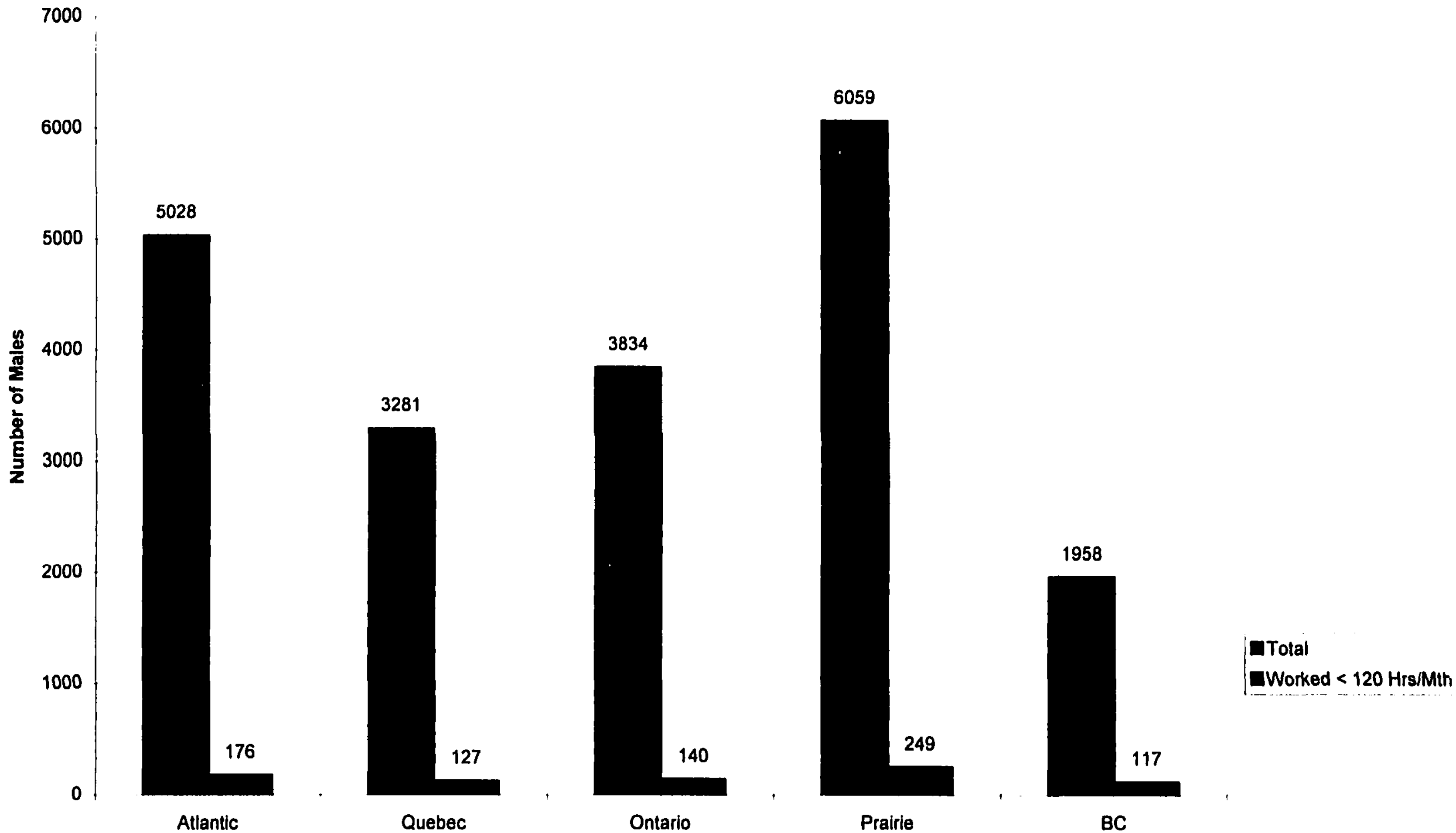


Figure 4.5: Comparison Between Those Working Less Than 120 Hours Per Month To Total Respondents In Each Region For Male Age 25-54 In 1986



According to the LMAS definition, the number of hours worked for each job in the reference year is derived by summing the number of hours worked in each employment spell, assuming the usual work schedule reported for the four week -- twenty-eight day -- 'lunar' month. A month-specific correction factor has been incorporated into the calculation to reflect the actual number of days in a month as compared to a 'lunar' month. These job-specific totals are subsequently summed to yield a total hours worked at all jobs in the reference year. To calculate the hourly wage rates at the aggregate level or for particular sub-groups of the population, for example, age, sex, industry, occupation, et cetera, the LMAS recommends the following methodology:

$$\text{Average hourly wage rate} = \frac{\Sigma\left(R(i) \times W(i) \times H(i)\right)}{\Sigma\left(W(i) \times H(i)\right)}$$

where R(i) = hourly wage rate for person i.

W(i) = the weighting for person i.

H(i) = the number of usual hours worked per week at the main job for person i.

This method compensates for the variation in hours worked among individuals by giving weights to those working longer hours. For workers who are not paid on an hourly basis, the hourly wage is calculated as pay per pay period, for example per week. This is divided by the usual hours per day times usual days per pay period. The selection of the hourly wage from this database is based on the hourly wage of the last job held by a respondent in the reference year. For instance, if a respondent held three jobs in the reference year, only the hourly wage of the third job is selected. However, respondents with "extreme" wages (lesser than $2 per hour and greater than $50 per hour) are removed from the sample. Any respondent who reports earning less than $2 an hour is believed to be a mistake because it is against the law for any employer to offer such low wage. For those respondents who earned more than $50 per hour, the regression equation used in this thesis is unable to determine their labour supply behaviour.

In the LMAS, there are five possible causes as to why a respondent is unemployed in the reference year, namely: lack of information, lack of skills, lack of education, lack of experience, and shortage of jobs. The LMAS defined the "lack of information" in four different ways. One survey question asked the respondents if their lack of information about available jobs had caused them trouble when looking for work; the second survey question, generally, asked if the respondents had no employers in the reference year; the third survey question asked if the respondent was not working at the end of the reference year; while the fourth survey question asked those who wanted to work, but did not look for work, if their lack of information had prevented them from working. If the respondent answered "yes" to any of the above four questions, we regard this respondent as "not having enough information about available jobs caused trouble when looking for work."

Similarly, the variables "lack of skill," "lack of education," "lack of experience," and "shortage of jobs" were asked in the same manner. Again, if the respondent answered "yes" to any of the four questions relating to these variables, the respondent is regarded as "not having the right skill for available jobs caused trouble when looking for work," "not having enough education for available jobs caused trouble when looking for work," "not having enough experience for available jobs caused trouble when looking for work," or "shortage of jobs in area caused trouble when looking for work."

One survey question asked the respondents if they had received income from sources such as unemployment benefits, welfare or worker's compensation. Every response to each of these sources of income is categorized accordingly. Figures 4.6 and 4.7 show the number of single males, age between 25 and 54, who received unemployment benefits or welfare in 1986 respectively. The Atlantic has the most male respondents, between the ages of 25 and 54, who received unemployment benefits and welfare, while Ontario and British Columbia have the least number of such respondents. This is not surprising at all since most jobs can be found in these two regions, and the Atlantic is known for its high unemployment rate. However, the highest number of male respondents, between 25 and 54 years of age, who received worker's compensation in 1986 come from the Prairie. Figure 4.8 shows the distribution of this variable for male respondents, age between 25 and 54, across the five regions.

The variable for job tenure describes the duration of the last job the respondent held in the reference year. The value corresponds to the week number since December 31, 1900. Week number 4435 corresponds to week one in 1986 which ends on the first Saturday of 1986. Week number 4487 starts on the last Sunday of 1986 which is week 53. Thus, if a respondent held 3 jobs in 1986 and worked only 6 weeks in the last job, the value corresponds to this variable for this particular respondent will be 6.

If a male respondent is a member of a union or his wage is covered by a collective agreement negotiated by a union, this individual is classified as belonging to a union.

Figure 4.6: Male Age 25-54 Who Received Unemployment Benefits In 1986

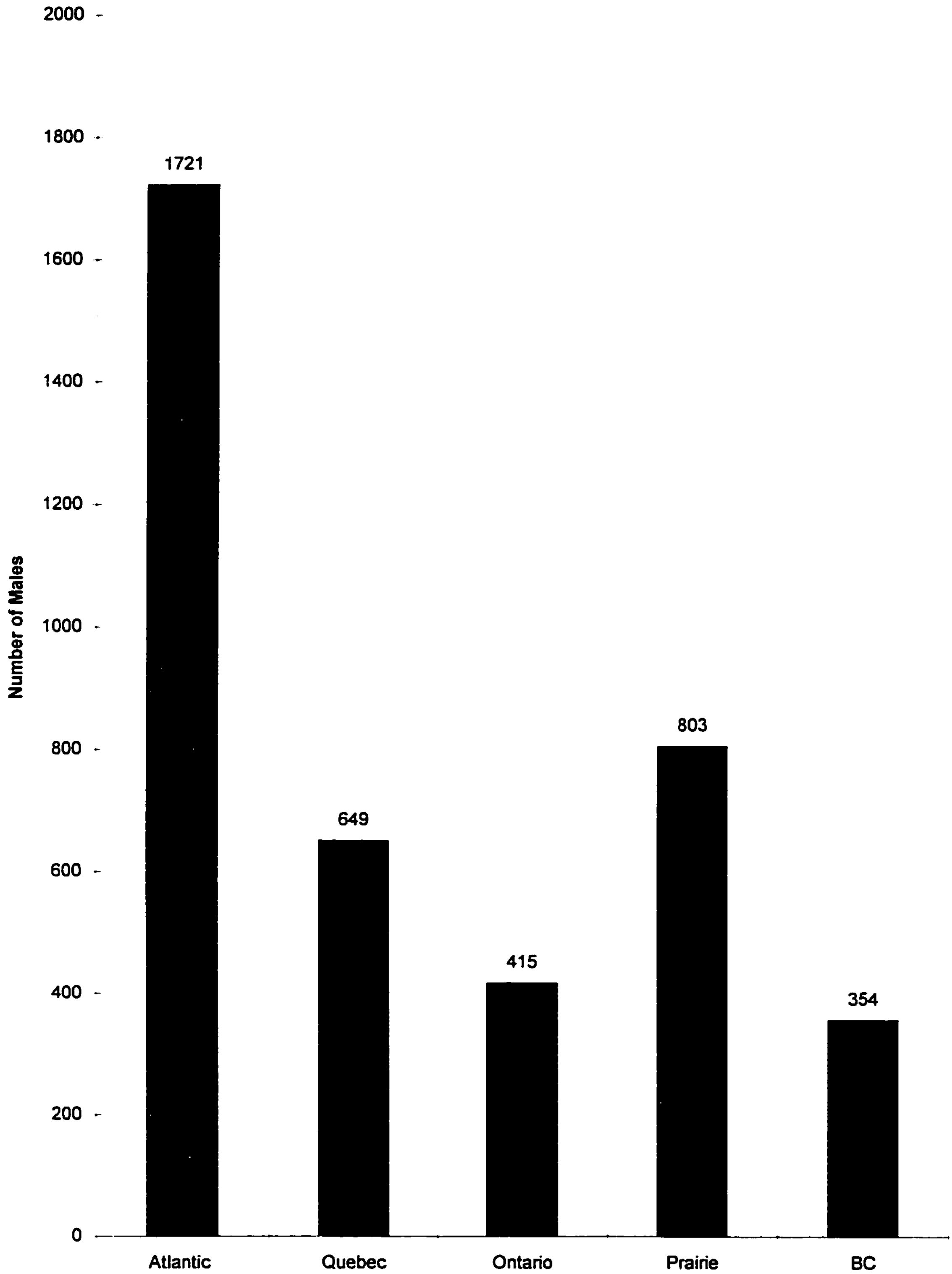


Figure 4.7: Male Age 25-54 Who Received Welfare In 1986

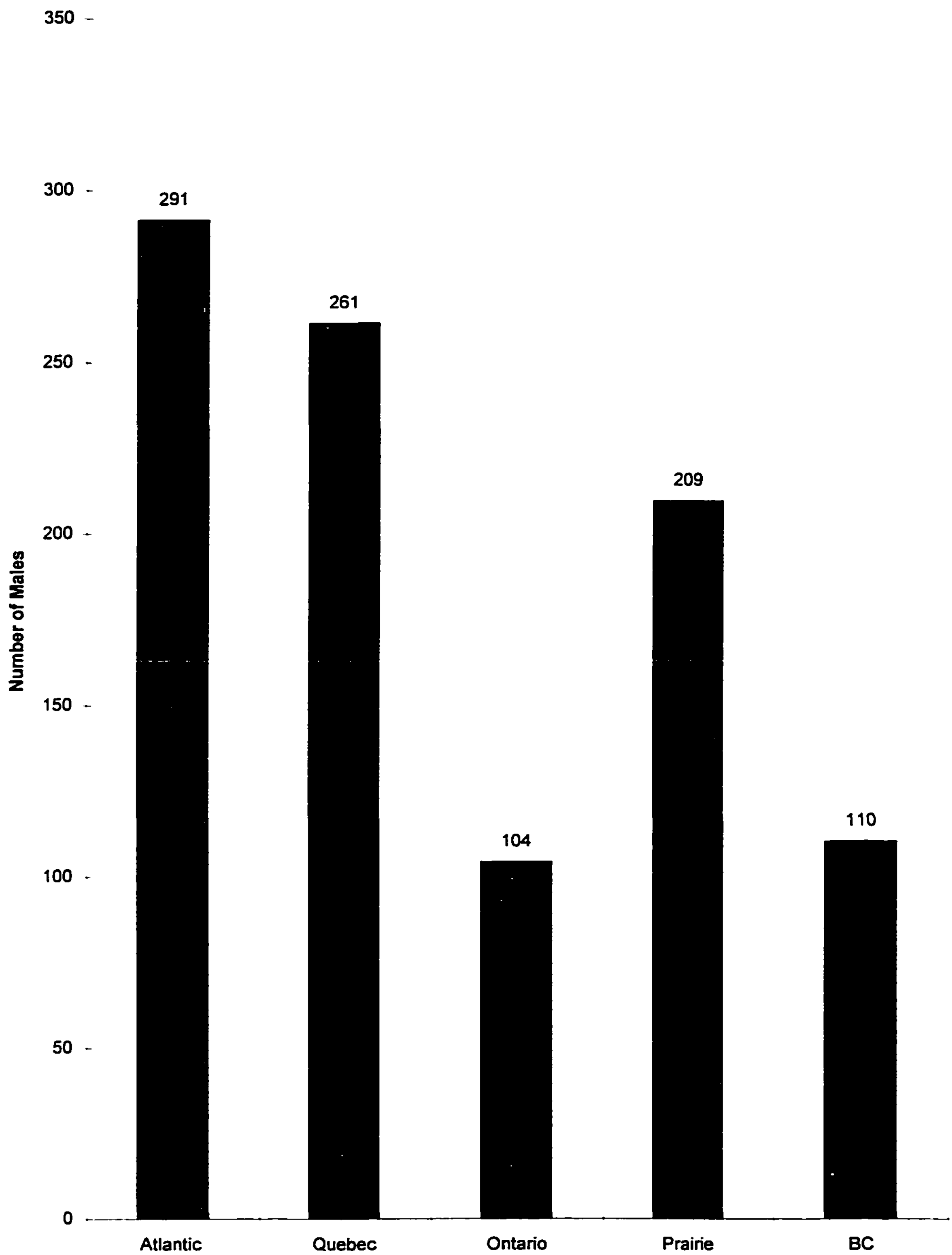


Figure 4.8: Male Age 25-54 Who Received Worker's Compensation In 1986

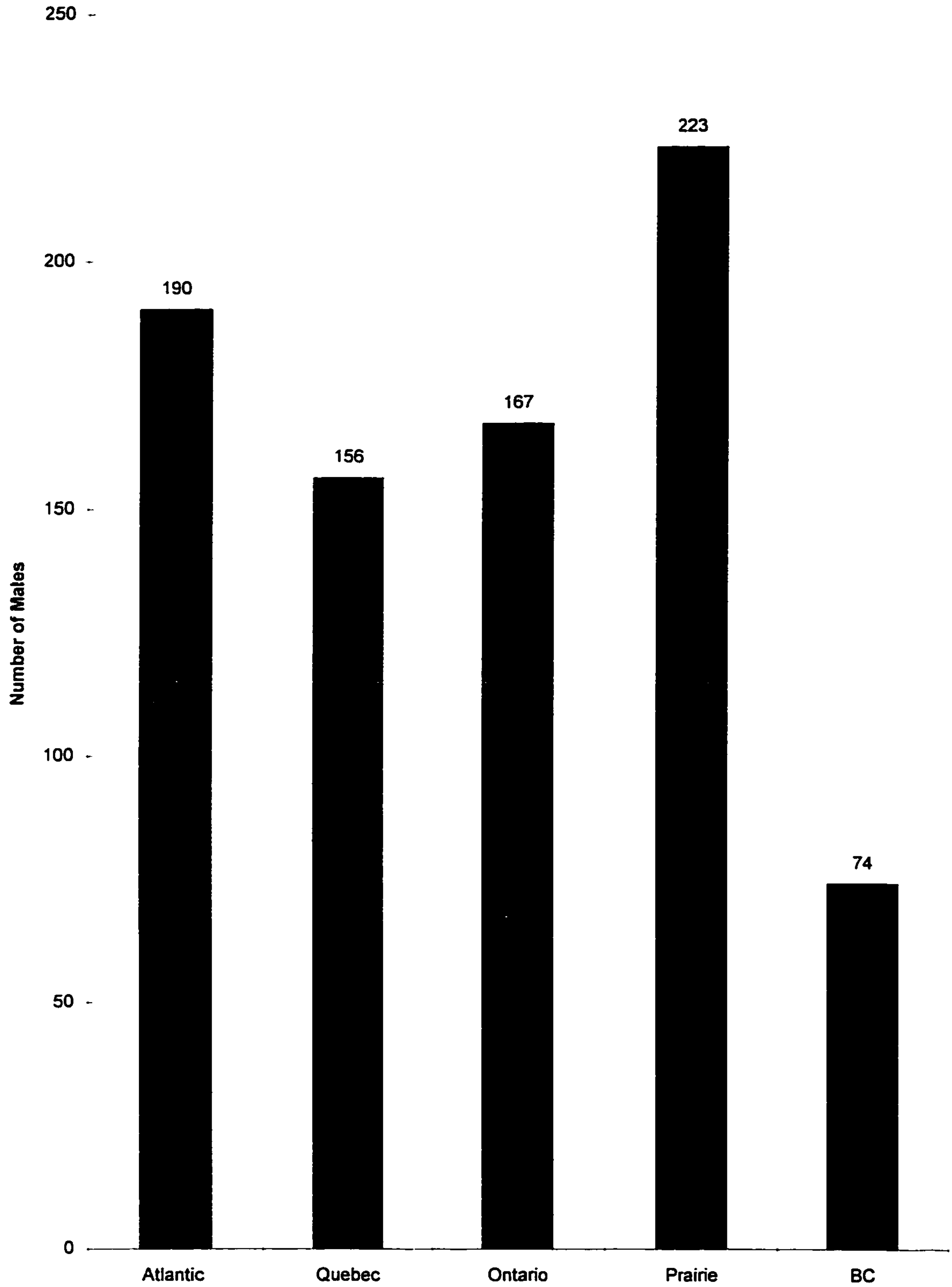


Figure 4.9 shows the number of male respondents, age between 25 and 54, belonging to a union. The Prairie has the most number of male respondents (215), from this age group, belonging to a union, while British Columbia has the least number of such respondents (62). Among those who worked, there were only a handful who were considered a member of a union according to the definition of the LMAS.

The dummy variable for pension coverage refers to those respondents who answered "yes" when they were asked if they were covered by a pension plan connected with their job(s) other than Canadian or Quebec pension plans, deferred profit sharing plans or personal savings plans for retirement. Figure 4.10 shows the number of male respondents, age between 25 and 54, who are covered by such pension plan. A comparison between those who are covered by such pension plan and those who worked is made in Figure 4.11. From this figure, it is worth noting that for every region, about half of the workers are covered by a pension plan connected with their job(s).

A respondent may have left his job(s) because of the following conditions:

(1) seasonal nature of the job(s), or

(2) non-seasonal economic or business conditions, or

(3) company moving or going out of business, or

(4) installation of or conversion to new equipment, or

(5) an on-call arrangement, or

(6) end of a temporary non-seasonal job, or

(7) sale of the business or farm.

Figure 4.9: Male Age 25-54 Who Were Members Of Union(s) Or Wages Were Covered By Collective Agreement Negotiated By Union(s) In 1986

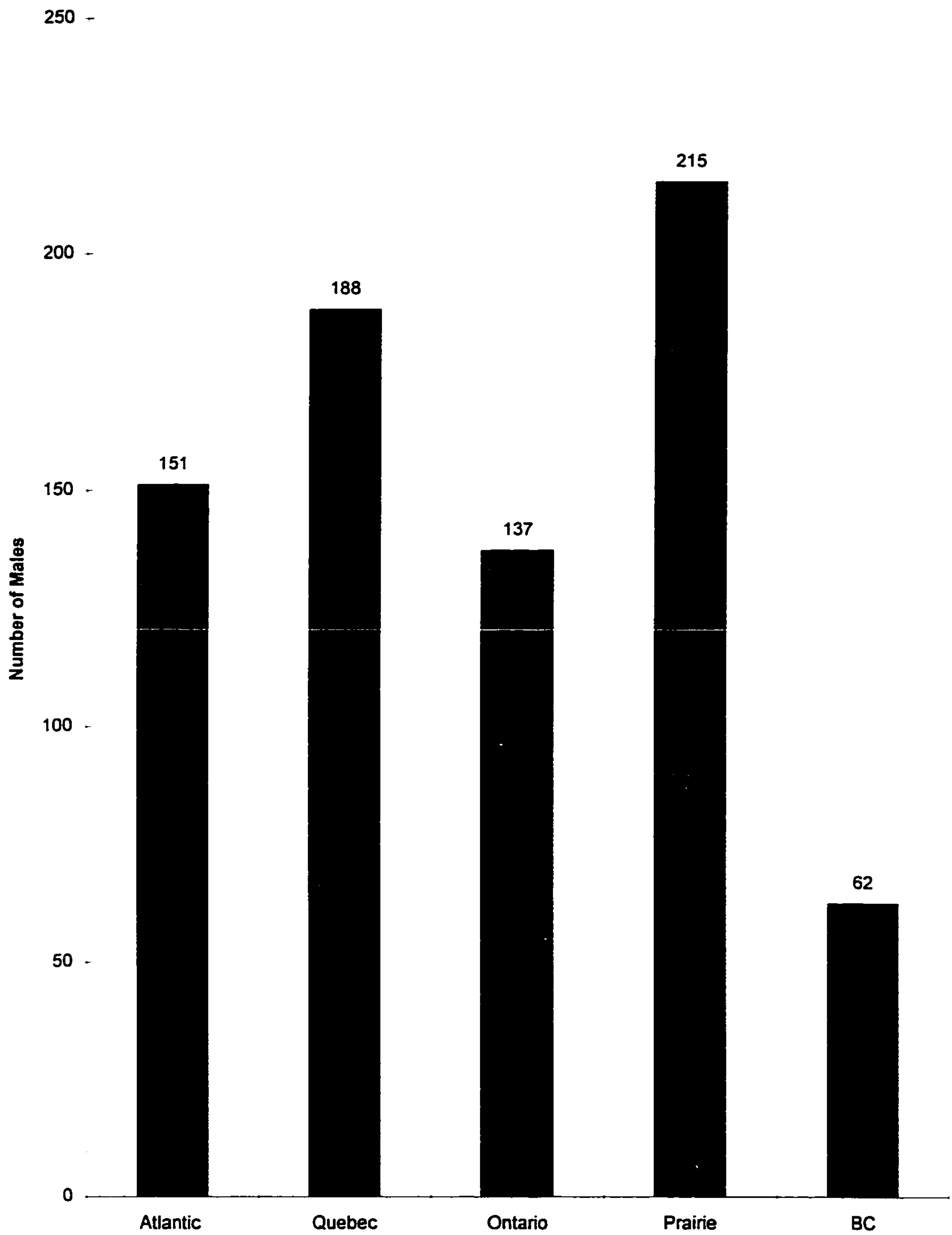


Figure 4.10: Male Age 25-54 Covered By A Pension Plan Connected With The Job(s) In 1986

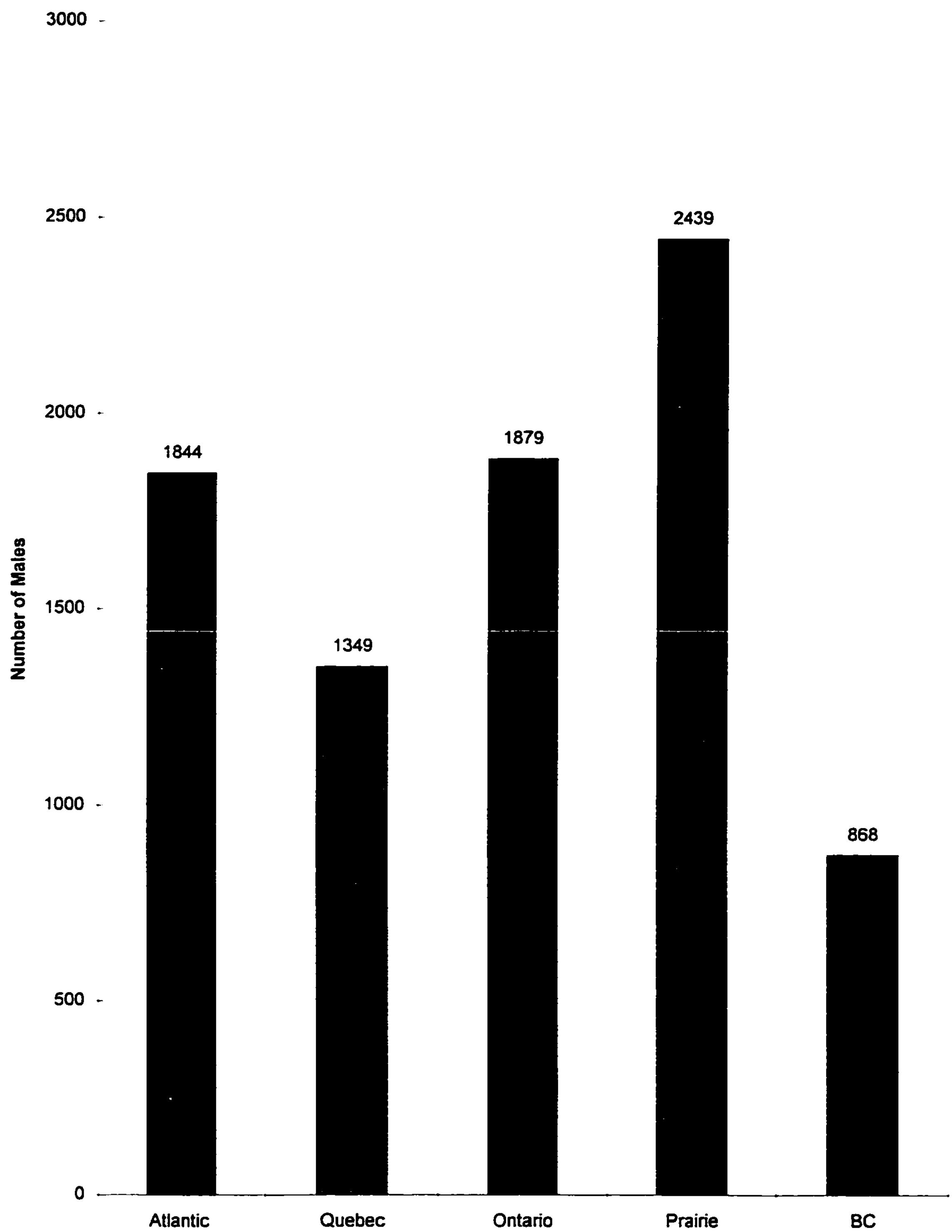
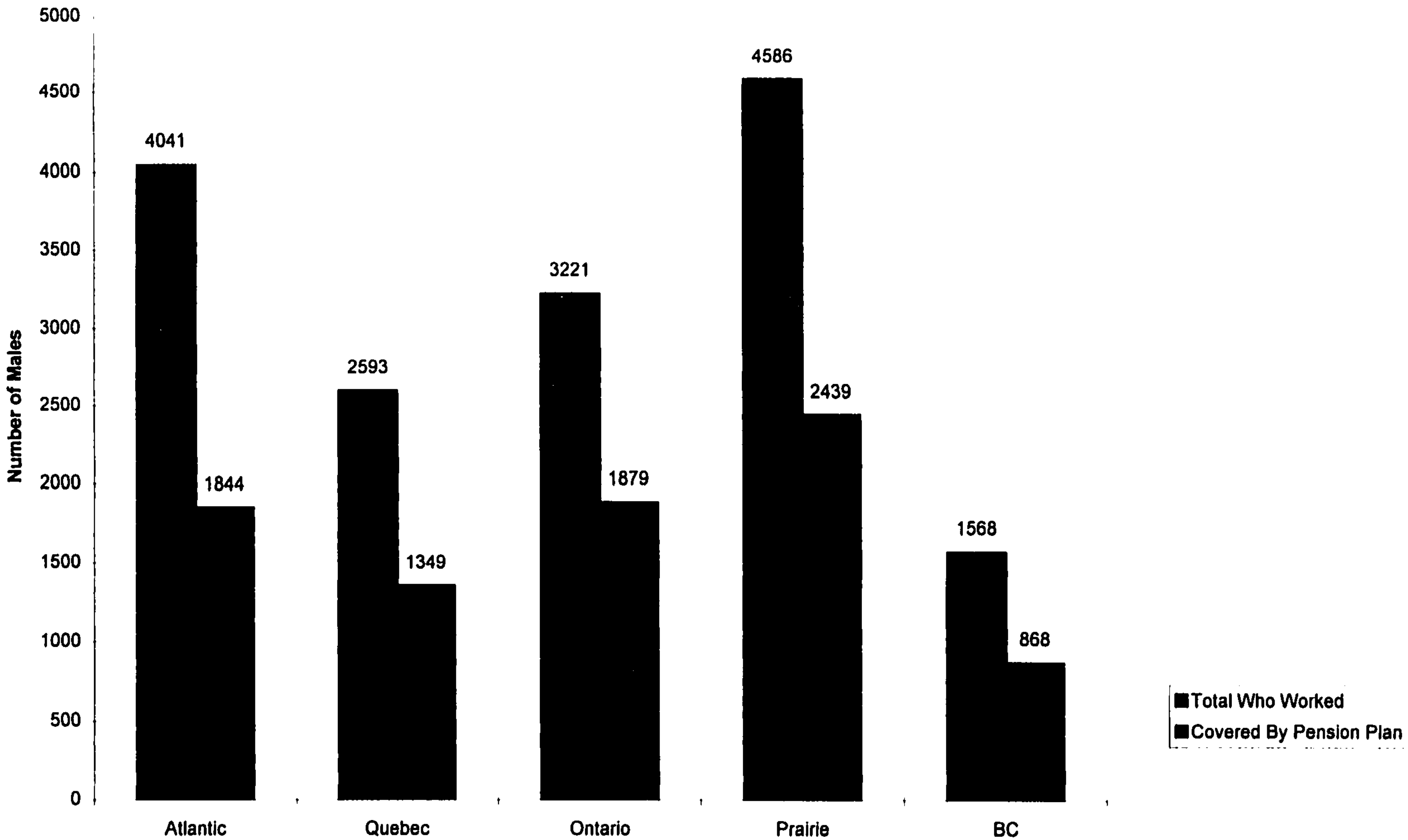


Figure 4.11: Comparison Between Those Who Worked And Those Who Are Covered By A Pension Plan For Male Age 25-54 In 1986



If any one of the above conditions explains the reason behind the individual leaving his job, he is classified as one who demands a job. Any other reasons that a respondent may give for leaving his job are ignored.

An individual may have wanted a job, or looked for work, but could not take a job because of his own illness or disability, or may have had to stop work because of his own illness or disability. These respondents are considered "ill" in one of the variables. Figure 4.12 shows the distribution of males with illness or disability across the five regions.

Not all respondents who worked are satisfied with the number of weeks worked. In the LMAS, respondents were asked if they were satisfied with the number of weeks worked in the reference year. Another question asked the respondents if they would like to work more weeks in the reference. If they answered "no" to the former and "yes" to the latter, these individuals are classified as underemployed. Alternatively, a respondent may be considered underemployed if he ended the year 1986 without a job, but looked actively for employment. Figure 4.13 shows the number of male respondents, between 25 and 54 years of age, who are considered underemployed across the regions.

The classifications of the dummy variables for the level of education attained for each respondent are similar to the way the individuals are categorized for their regions of residence. If a respondent has no education, or only elementary school education, the variable for "elementary education" will take the value of one; if he has a post-secondary education, the "post secondary" variable will take the value of one; if he has a post-secondary certificate or diploma education, the "post diploma" variable will take the

Figure 4.12: Number Of Male Age 25-54 Who Are Ill Or Disable In 1986

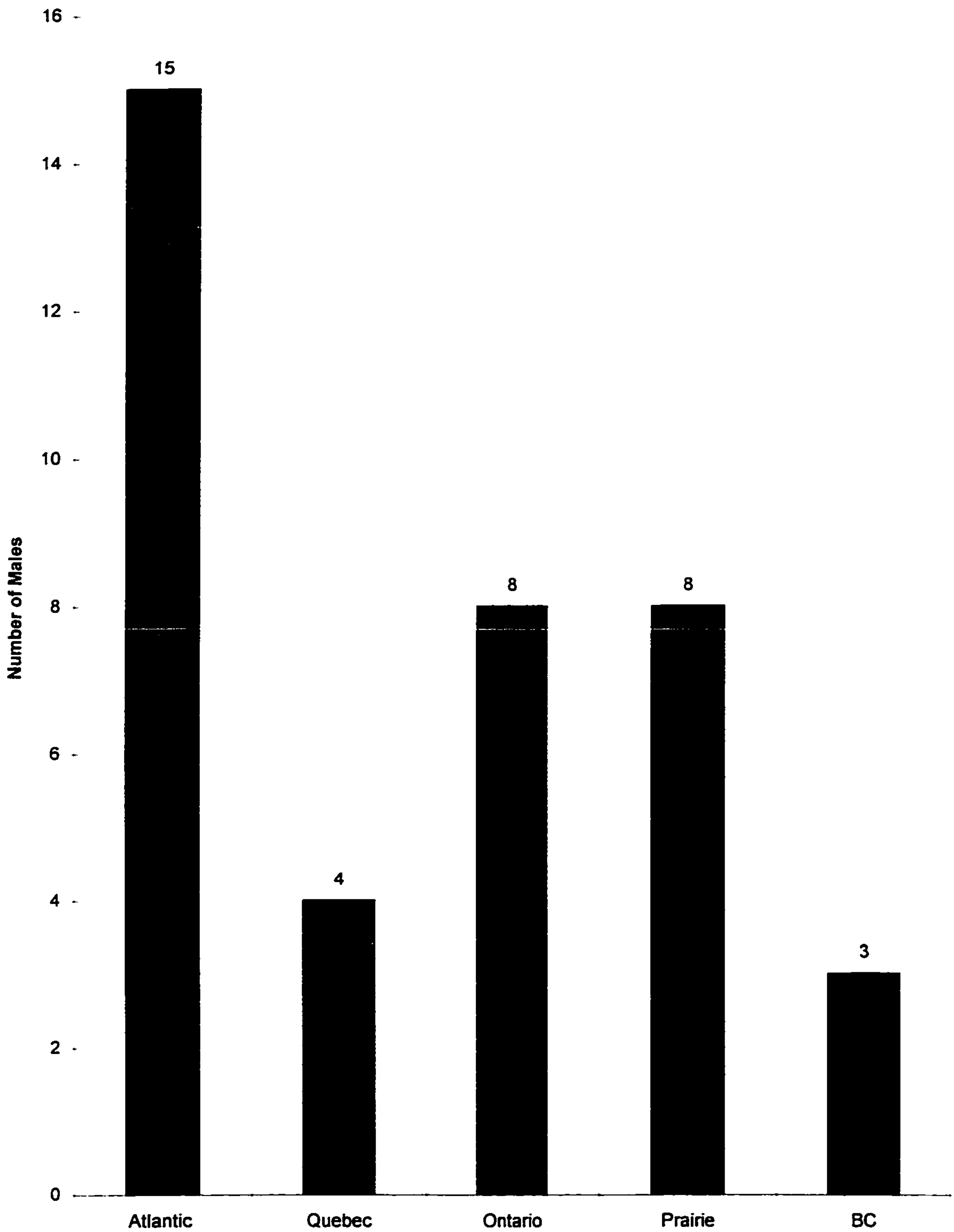
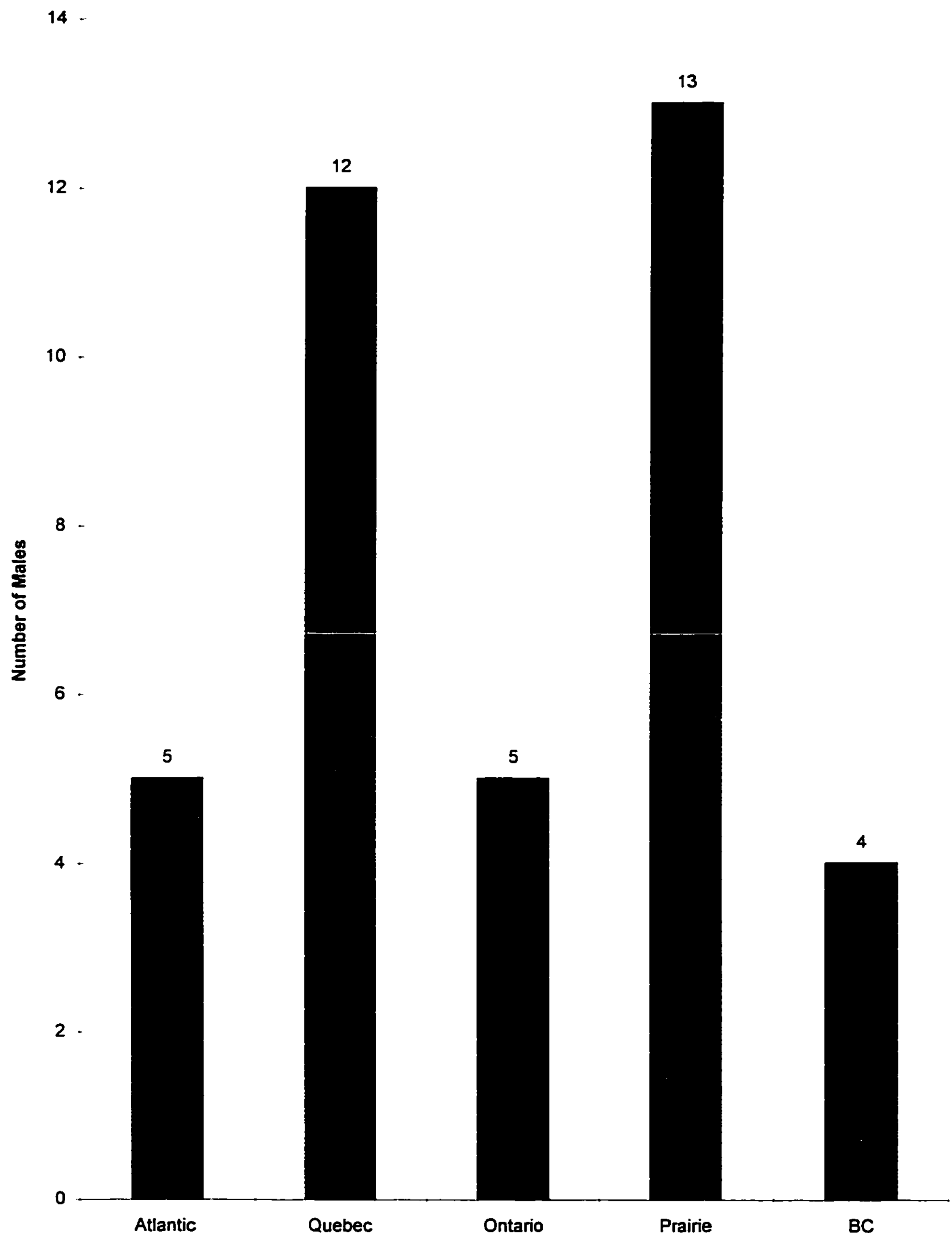


Figure 4.13: Number Of Males Age 25-54 Who Are Not Satisfied With Their Number Of Weeks Worked In 1986



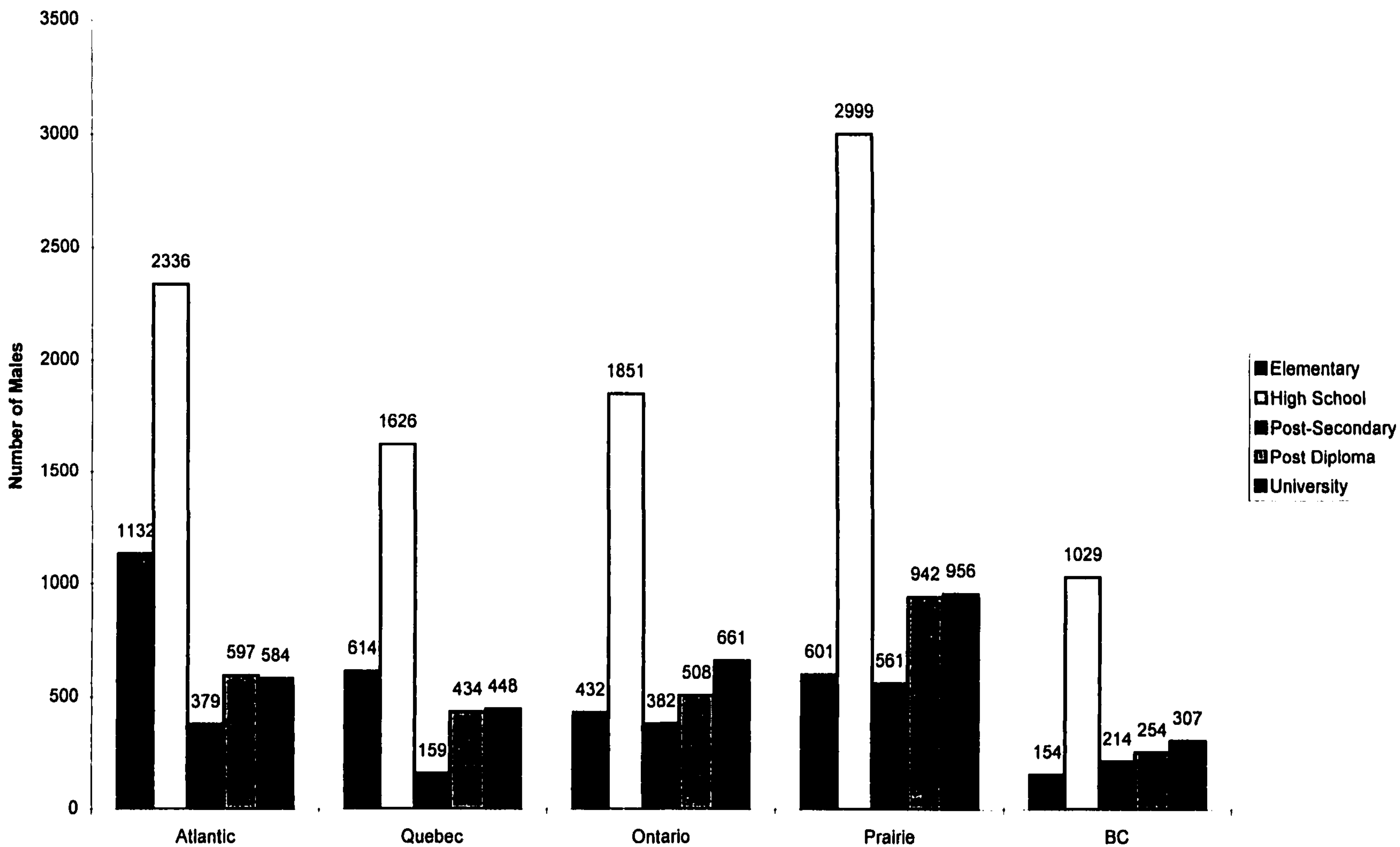
value of one; and if he has a university education, the "university" variable will take the value of one. Figure 4.14 shows the number of male respondents, age between 25 and 54, in each level of education across Canada.

When assigning codes for industry and occupation, the LMAS questionnaire collected information on the name of the employer, the kind of business, industry or service the employer was in, the kind of work done and the usual duties and responsibilities of a respondent in the job for each job held by the respondent. This information was used to assign industry and occupation codes using the 1980 version of Statistics Canada's Industrial and Occupational Classifications.

The unemployment rate represents the number of unemployed persons expressed as a percentage of the labour force. The unemployment rate for a particular group (age, sex, marital status, etc.) is the number unemployed in that group expressed as a percentage of the labour force for that group. The LMAS is unable to provide us with the industry unemployment rate. Therefore, the source of this data is the Canadian Socio-economic Information Management (CANSIM) as well as the Statistic Canada catalogue. However, these sources do not have all the 52 classifications which LMAS have. After a thorough and careful review, the 52 classifications from the LMAS are grouped accordingly to the 14 different industries gathered from CANSIM and the Statistic Canada catalogue.

The base period of the labour supply function is where all dummy variables take the value of zero. In this case, we are referring to those non-single male respondents who are employed, living in Ontario, age between 25 and 34, have a high school education,

Figure 4.14: Level Of Education Attained By Male Age 25-54 In 1986



are not underemployed, and are not receiving income from sources such as unemployment benefits, welfare or worker's compensation.

In brief recapitulation, Table 4.1 presents a list of the variables employed in the regression. There are, altogether, twenty-five dummy variables out of the thirty variables which are adopted. The remainder five variables are continuous variables. Table 4.1 presents a summary of how the values (of one and zero) are attached to the dummy variables.

4.2 Sampling Error

Total non-response has been a major source of non-sampling error in many surveys. In the LMAS, total non-response occurs because the household cannot be contacted, no member of the household was able to provide the information, or members of the household refused to participate in the survey in the survey year. Total non-response is handled by adjusting the household sampling weight of responding households to compensate for missing households or missing individuals within responding households. Analysis of the characteristics of LMAS non-respondents suggests that total non-response is not a major source of non-sampling error.

Partial non-response in the LMAS may occur if the respondent refuses to answer a question, does not understand a question, or in the case of a proxy reporting, the person reporting does not know the answer to a question. Generally, partial non-response was not a problem in the LMAS. As a result, it is unlikely that partial non-response contributed significantly to sampling error.

Table 4.1: Definition of Variables

Variables	Definition
Hours (H)	Total number of hours supplied in 1986
Gross Wage (GW)	Hourly wage rate of the last job held in 1986.
Part-Time (PT)	= 1 if Hours worked per month less than 120; = 0 otherwise.
Lack Information (LIN)	= 1 if "Not having enough information about available jobs caused trouble when looking for work; or no employers in the reference year; was not working at the end of the reference year; or those who wanted to work but did not look for work because the lack of information had prevented them from working"; = 0 otherwise.
Lack Skill (LSK)	= 1 if "Not having the right skills for available jobs caused trouble when looking for work; or no employers in the reference year; was not working at the end of the reference year; or those who wanted to work but did not look for work because the lack of skill had prevented them from working"; = 0 otherwise.
Lack Education (LED)	= 1 if "Not having enough education for available jobs caused trouble when looking for work; or no employers in the reference year; was not working at the end of the reference year; or those who wanted to work but did not look for work because the lack of education had prevented them from working"; = 0 otherwise.

Table 4.1: Definition of Variables (continued)

Variables	Definition
Lack Experience (LEX)	= 1 if "Not having enough experience for available jobs caused trouble when looking for work; or no employers in the reference year; was not working at the end of the reference year; or those who wanted to work but did not look for work because the lack of experience had prevented them from working"; = 0 otherwise.
Job Shortage (JS)	= 1 if "Shortage of jobs in the area caused trouble when looking for work; or no employers in the reference year; was not working at the end of the reference year; or those who wanted to work but did not look for work because of job shortage had prevented them from working"; = 0 otherwise.
Unemployment Insurance (UI)	= 1 if Received income from unemployment insurance benefits during 1986; = 0 otherwise.
Received Welfare (WF)	= 1 if Received income from social assistance or welfare benefits during 1986; = 0 otherwise.
Worker's Compensation (WC)	= 1 if Received income from worker's compensation during 1986; = 0 otherwise.

Table 4.1: Definition of Variables (continued)

Variables	Definition
Demand For Job(s) (DD)	= 1 if Left job(s) because of seasonal nature of job(s), or non-seasonal economic or business conditions, or company moving or going out of business, or installation of or conversion to new equipment, or an on-call arrangement, or end of a temporary non-seasonal job, or sale of the business or farm;
	= 0 otherwise.
Atlantic (AT)	= 1 if Resident of Newfoundland, Prince Edward Island, Nova Scotia, or New Brunswick;
	= 0 otherwise.
Quebec (PQ)	= 1 if Resident of Quebec;
	= 0 otherwise.
Prairie (PR)	= 1 if Resident of Manitoba, Saskatchewan, or Alberta;
	= 0 otherwise.
British Columbia (BC)	= 1 if Resident of British Columbia;
	= 0 otherwise.
Age 35-44 (AGE35)	=1 if Aged between 35 and 44;
	= 0 otherwise.
Age 45-54 (AGE45)	= 1 if Aged between 45 and 54;
	= 0 otherwise.
Industry Unemployment Rate (IUR)	Industry unemployment rate for industry of major employment

Table 4.1: Definition of Variables (continued)

Variables	Definition
Union (U)	= 1 if Member of a union or wages were covered by collective agreement negotiated by a union; = 0 otherwise.
Disability (DIS)	= 1 During the period wanted a job or looked for work but could not take a job because of own illness or disability, or stopped working because of own illness or disability; = 0 otherwise.
Underemployment (UE)	= 1 if Wanted to work more weeks, -i.e. (a) answer "no" to question "were you satisfied with the number of weeks worked in 1986" and "yes" to "did you want to work more weeks in 1986" or (2) ended the year 1986 without a job, but looking actively for employment; = 0 otherwise.
Pension Covered (PC)	= 1 if Covered by a pension plan connected with the job(s) (not counting Canadian or Quebec pension plans, deferred profit sharing plans or personal savings plans for retirement; = 0 otherwise.
Dependent Children (DC)	Number of own children under 15 and number of other children under 15.
Single (S)	= 1 if Males who are not married; = 0 otherwise.
Elementary Education (EE)	= 1 if Obtain none or elementary education; = 0 otherwise.

Table 4.1: Definition of Variables (continued)

Variables	Definition
Post Secondary (PS)	= 1 if Obtain some post-secondary education;
	= 0 otherwise.
Post Diploma (PD)	= 1 if Obtain post-secondary certificate or diploma education;
	= 0 otherwise.
University (UNI)	= 1 if Obtain university education;
	= 0 otherwise.
Job Tenure (JT)	Stop week minus start week
M	"Inverse Mills ratio" calculated from probit analysis of underemployment

CHAPTER 5: EMPIRICAL RESULTS

Killingsworth (1983) has noted that most of the results for men imply that the uncompensated wage elasticity is between -0.38 and +0.14, and that for women this figure is between -0.89 and +15.24. Such large discrepancies in the estimation of the wage elasticity in labour economics have lead some applied economists to look for both economic and statistical explanations. One important question is whether or not different sample sizes may result in variation in the estimation of the wage elasticity. To investigate this possible cause for the observed inconsistency in the estimated coefficients, the Monte Carlo technique is applied.

In this chapter, the results of the Monte Carlo simulation are reported. The simulation is performed using the econometric software package, *Shazam* (1993). Figures and tables are used to provide intuitive insights for explaining the variation in estimated elasticities and their implications. Section 5.1 focuses on what actually motivates the investigation of the instability of the estimated coefficients. It presents the graphs of five independently estimated coefficients and their corresponding standard deviations at each sample size. The purpose is to provide a visual presentation of the behaviour of these estimated parameters. These graphs depict the instability of the estimated coefficient at small sample sizes.

Section 5.2 presents the results of a Monte Carlo procedure performed on the semi-log labour supply function. The total sample size of 12,680 observations is divided into 26 different sample sizes. For each of these sample sizes, 1,000 estimates of the coefficient of the ln(GW) variable are obtained. Two important characteristics of the estimated coefficient are generated: they are the means of the 1,000 estimates and the standard deviations at each sample size. Each of these iterations is independently and randomly drawn from each estimation at each sample size. A simple *t*-test is performed on the means of these coefficient estimates to investigate if they are statistically different from the true value. To test if the 1,000 estimated coefficients at each sample are statistically significant from one another, a pairwise test is used.

Section 5.3 presents the results of using the confidence interval at each sample size. A confidence interval around the true value of the coefficient of ln(GW) can be constructed based on the true standard deviation. The purpose is to find out the percentage of the point estimators that fall within the confidence intervals. In using the confidence interval estimation, the probability of committing a Type I error can be controlled by choosing the level of significance or *p* value.

In estimating the labour supply function, the *t*-ratio of the coefficient of ln(GW) can sometimes be insignificant. This means some estimated coefficients generated from different sample sizes are not statistically different from zero. Section 5.4 discusses the percentage of the estimated coefficients that are statistically different from zero for the various sample sizes used. Even in repeated sampling, the estimated coefficients produce the wrong sign. This is called *decentralization*. This section also investigates the percentage of the estimated coefficients possessing the wrong sign at each sample size.

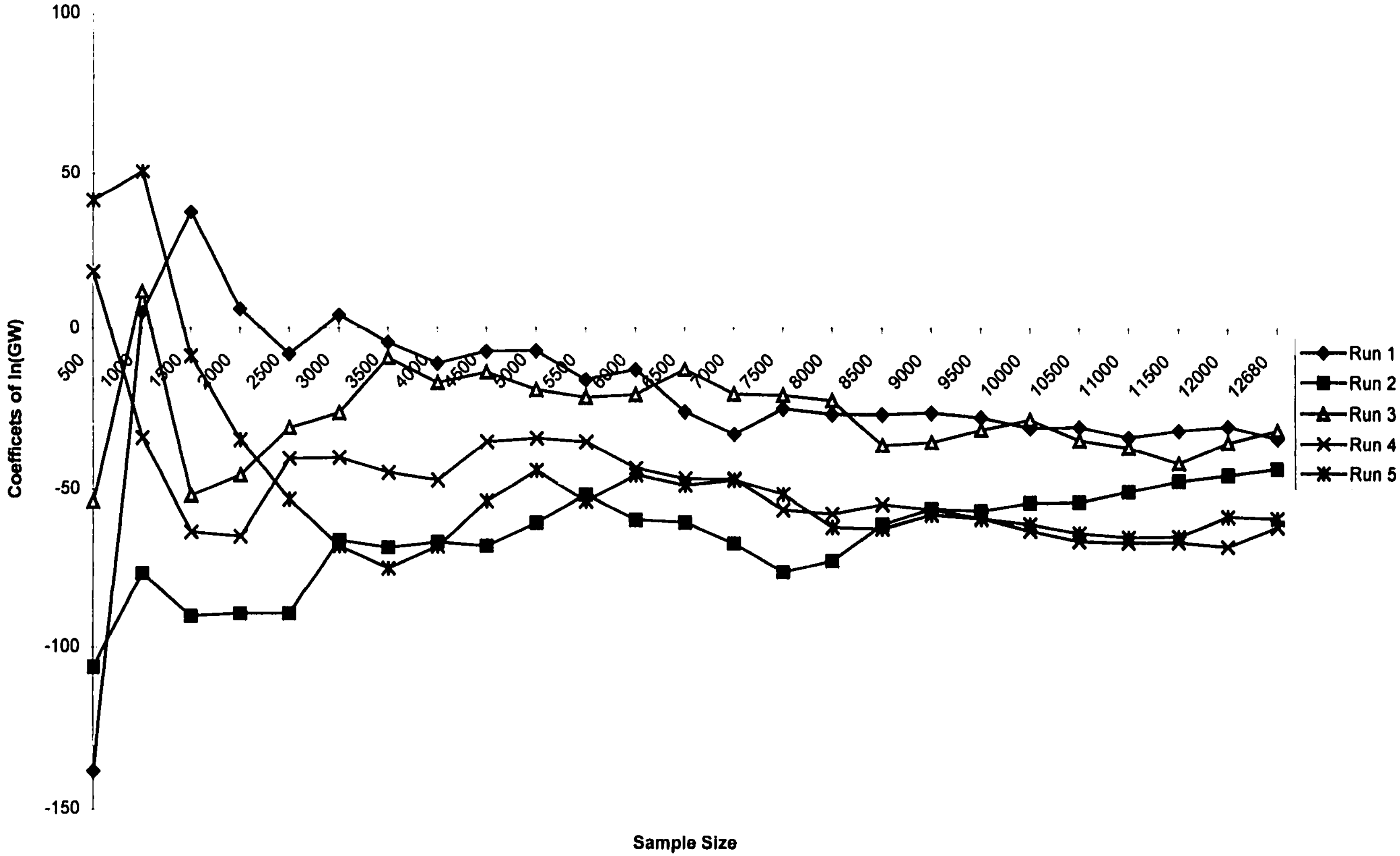
Concluding remarks regarding the Monte Carlo simulation are presented in section 5.5. It summarizes the results presented in the previous sections and highlights the behaviour of the estimated coefficients at small sample sizes.

5.1 The Motivation

In this section, five individual iterations at each sample size are presented. The smallest sample size begins with 500 observations, and the subsequent sample sizes are in increments of 500 observations over the previous sample size. At 12,000 observations the increment is 680 observations as this is when all observations are fully exhausted. Figure 5.1 depicts five different estimates of the coefficient of ln(GW). This graph shows that the estimated coefficients at small sample sizes fluctuate over a wider range than those at larger sample sizes. This is especially true for sample sizes from 500 up to about 3,500 as compared to the larger sample sizes. Once the sample size exceeds 6,000 observations, the estimates appear to converge towards the true value. For sample sizes larger than 6,000 observations, the estimated coefficients fluctuate much less than those in the smaller sample sizes.

It is worth noting that the estimated coefficient of the ln(GW) will not be exactly the same as the true value at the full sample size of 12,680 observations. This is because the values of the dependent variable (hours of work), generated by the Monte Carlo technique, will never be the same as the original dependent variable. The generated values of the dependent variable are random and the variation depends on the true

Figure 5.1: Coefficients Of ln(GW) As Sample Sizes Increase



variance of the residuals of the labour supply function. All that is required from the point estimates at 12,680 observations is that they are statistically indistinguishable from the true value.

Observing the standard errors of these estimated coefficients from the five iterations can also provide some insight in motivating the study of parameter instability. Figure 5.2 plots the standard errors of these estimates against the various sample sizes. The graph shows a smooth curve which is convex to the origin. It declines fast for smaller sample sizes up to about 5,000 observations and then stabilizes from 6,000 observations onwards. The standard deviations for the smaller sample sizes show some instability because they are rarely close to each other. Moreover, for the smaller sample sizes, the standard deviations are relatively larger than those derived from larger sample sizes. When the sample size starts to increase, one can observe that these standard deviations are almost identical.

Table 5.1 presents the summary statistics of some of the continuous variables used in the labour supply function for the different sample sizes. This table provides a guideline to the reading of the enormous data set by the computer program written. The minimum and maximum values columns show that these values change accordingly as the sample size increases. For example, at a sample size of 12,680, the maximum value of the variable of ln(GW) is 3.8699. Any smaller sample sizes will not have their maximum values for this variable larger than this amount.

Given the unstable behaviour of the estimated coefficients at small sample sizes as depicted in Figure 5.1, the corresponding standard errors of these estimated coefficients

Figure 5.2: Standard Error of the Coefficients of ln(GW)

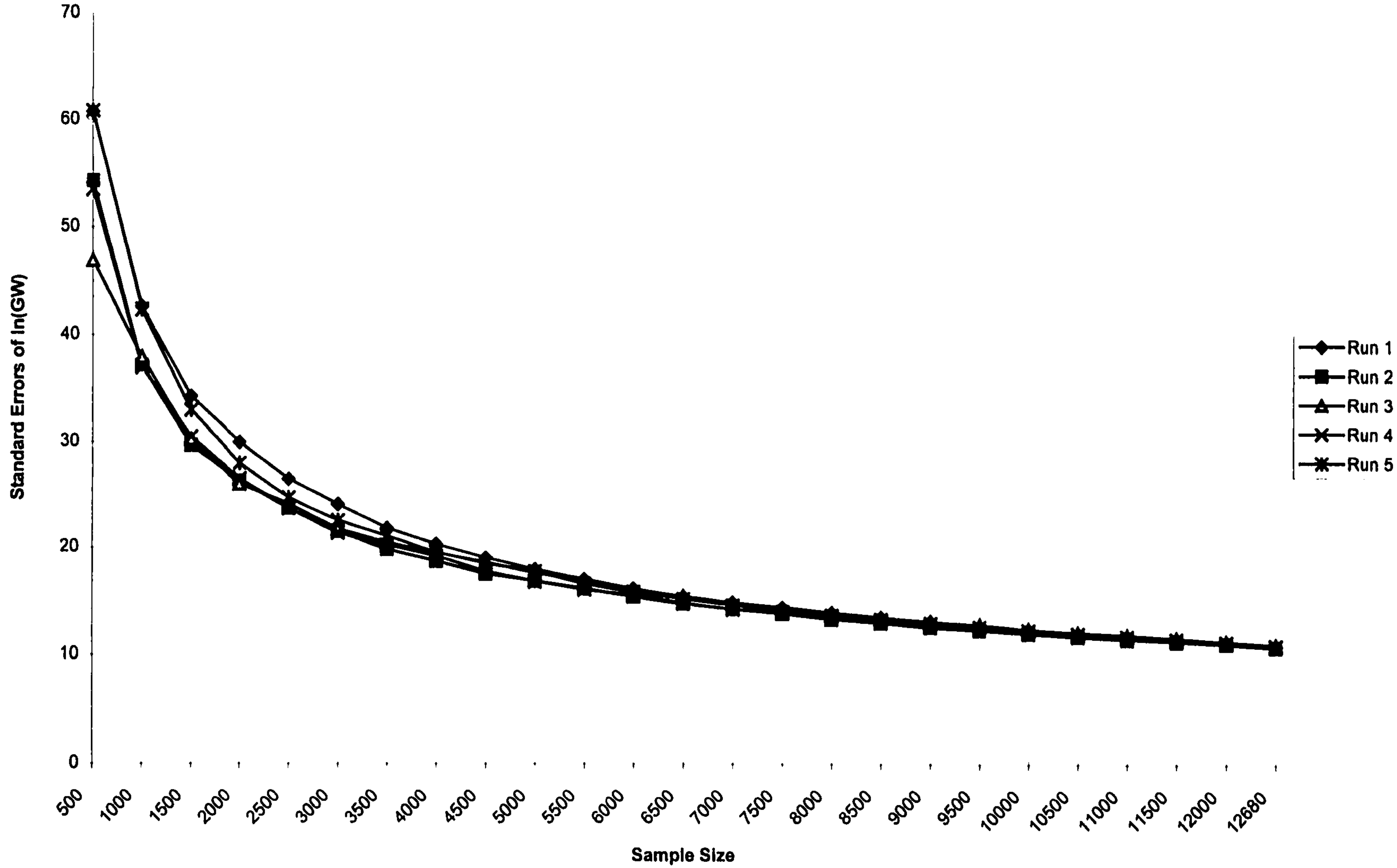


Table 5.1: Summary Statistics Of Data
By Sample Size

Variable	Sample Size	Mean	Standard Deviation	Minimum	Maximum
H	500	2091.0	479.08	349	4404
ln(GW)		2.5629	0.4317	0.8329	3.675
JT		464.56	376.26	9	2088
DC		1.1303	1.1183	0	5
H	1,000	2092.4	458.27	309	4404
ln(GW)		2.5525	0.4304	0.8329	3.8699
JT		475.46	394.52	9	2697
DC		1.0793	1.1179	0	5
H	1,500	2095.5	460.24	220	4404
ln(GW)		2.5519	0.4293	0.7608	3.8699
JT		484.51	399.93	6	2697
DC		1.0445	1.1153	0	5
H	2,000	2095.5	483.77	183	4404
ln(GW)		2.5519	0.4427	0.7608	3.8699
JT		476.05	397.07	6	3187
DC		1.0428	1.1071	0	5
H	2,500	2085.8	481.03	183	4404
ln(GW)		2.5491	0.4390	0.7608	3.8699
JT		473.61	392.81	6	3187
DC		0.9920	1.0935	0	5
H	3,000	2081.3	476.70	183	4404
ln(GW)		2.5580	0.4397	0.7130	3.8699
JT		477.45	396.82	6	3187
DC		0.9823	1.0965	0	5
H	3,500	2079.6	469.30	183	4404
ln(GW)		2.5496	0.4402	0.71230	3.8699
JT		474.11	395.66	5	3187
DC		0.9796	1.0885	0	7

Table 5.1: Summary Statistics Of Data By Sample Size (continued)

Variable	Sample Size	Mean	Standard Deviation	Minimum	Maximum
H	4,000	2078.8	473.96	183	4404
ln(GW)		2.5489	0.4369	0.7130	3.8699
JT		471.71	393.11	5	3187
DC		0.9801	1.0852	0	7
H	4,500	2081.6	482.99	183	4404
ln(GW)		2.5505	0.4377	0.7130	3.8699
JT		470.21	392.38	5	3187
DC		0.9808	1.0842	0	7
H	5,000	2085.8	478.31	171	4404
ln(GW)		2.5529	0.4387	0.7080	3.8699
JT		467.95	388.67	4	3187
DC		0.9728	1.0767	0	7
H	5,500	2083.9	485.38	171	4404
ln(GW)		2.5517	0.4396	0.7080	3.8699
JT		470.31	391.61	4	3187
DC		0.9772	1.0769	0	8
H	6,000	2080.6	486.75	171	4404
ln(GW)		2.5496	0.4397	0.7080	3.8699
JT		469.18	392.40	4	3187
DC		0.9752	1.0759	0	8
H	6,500	2080.9	485.02	171	4404
ln(GW)		2.5505	0.4413	0.7080	3.8699
JT		469.49	393.89	4	3187
DC		0.9686	1.0754	0	8
H	7,000	2079.9	481.64	171	4404
ln(GW)		2.5518	0.4401	0.7080	3.8699
JT		470.53	393.59	4	3187
DC		0.9680	1.0742	0	8

Table 5.1: Summary Statistics Of Data
By Sample Size (continued)

Variable	Sample Size	Mean	Standard Deviation	Minimum	Maximum
H	7,500	2081.9	478.67	171	4404
ln(GW)		2.5538	0.4438	0.7080	3.8699
JT		472.33	393.93	4	3187
DC		0.9714	1.0786	0	8
H	8,000	2078.8	479.24	171	4404
ln(GW)		2.5529	0.4425	0.7080	3.8699
JT		472.27	394.42	4	3187
DC		0.9661	1.0756	0	8
H	8,500	2077.6	485.16	171	4404
ln(GW)		2.5520	0.4431	0.7080	3.8699
JT		473.81	396.89	4	3187
DC		0.9622	1.0783	0	8
H	9,000	2077.6	483.80	171	4404
ln(GW)		2.5524	0.4434	0.7080	3.8699
JT		473.71	396.62	4	3187
DC		0.9614	1.0788	0	8
H	9,500	2077.3	483.30	171	4404
ln(GW)		2.5549	0.4422	0.7080	3.8699
JT		474.56	396.32	4	3187
DC		0.9642	1.0778	0	8
H	10,000	2077.3	482.32	171	4404
ln(GW)		2.5536	0.4434	0.7080	3.8699
JT		474.87	397.04	4	3187
DC		0.9644	1.0815	0	8
H	10,500	2079.2	480.63	171	4404
ln(GW)		2.5519	0.4448	0.7080	3.8699
JT		474.95	397.50	4	3187
DC		0.9627	1.0779	0	8

Table 5.1: Summary Statistics Of Data
By Sample Size (continued)

Variable	Sample Size	Mean	Standard Deviation	Minimum	Maximum
H	11,000	2078.4	481.18	171	4404
ln(GW)		2.5524	0.4431	0.7080	3.8699
JT		474.44	397.71	4	3187
DC		0.9650	1.0772	0	8
H	11,500	2075.6	482.52	171	4404
ln(GW)		2.5528	0.4441	0.7080	3.8699
JT		474.64	398.58	4	3187
DC		0.9658	1.0756	0	8
H	12,000	2075.1	483.07	171	4404
ln(GW)		2.5536	0.4438	0.7080	3.8699
JT		475.1	399.38	4	3187
DC		0.9662	1.0762	0	8
H	12,500	2075.5	483.33	171	4404
ln(GW)		2.5527	0.4423	0.7080	3.8699
J		473.96	399.14	4	3187
DC		0.9676	1.0751	0	8
H	12,680	2076.4	483.37	171	4428
ln(GW)		2.5525	0.4416	0.7080	3.8699
JT		474.47	398.89	4	3187
DC		0.9678	1.0751	0	8

H denotes the total number of hours worked.
ln(GW) denotes the logarithm of gross wage.
JT denotes job tenure.
DC denotes the number of dependent children.

as illustrated in Figure 5.2, and the summary statistics of the data, one wonders if sample size is the cause of this coefficient instability and difference in the standard deviations. Thus, the visual inspection of the behaviour of the estimated coefficients and their corresponding standard errors have prompted a further investigation of the properties of these estimates. The next section will present the evidence from the Monte Carlo simulation and the results of statistical testing.

5.2 Evidence From Monte Carlo

This section provides evidence from the Monte Carlo simulation. Different sample sizes are used to estimate the coefficient of ln(GW). Similar to the previous section, the smallest sample size used is 500 observations, and the subsequent sample sizes are in increments of 500 observations over the previous sample size until all observations are utilized. Thus, there are 26 different sample sizes, including the full sample of 12,680, that are used to obtain the estimated coefficients of ln(GW). For each sample size, there are 1,000 iterations performed on the semi-log labour supply function. Each of these iterations is independently and randomly drawn. A thousand estimates of the coefficient of the ln(GW) variable are obtained at each sample size, and the mean of these estimates, together with the standard deviation of this mean are generated. Two other characteristics of this coefficient are also generated: the mean of the standard deviation and the standard deviation of this standard deviation.

No results can be convincing evidence without first undergoing statistical testing. A simple *t*-test is performed on these coefficient estimates to investigate if they are statistically different from the true value (-47.938). Another test, called a pairwise test, is used to test if the 1,000 estimates at each sample size are statistically different from one another.

The mean of the 1,000 estimated coefficients for each sample size is depicted in Figure 5.3. The dashed line is the true value of the coefficient of the ln(GW) variable. To view the fluctuation of the mean of the estimated coefficient, this graph can be divided into 4 main segments. The first segment is the sample sizes between 500 and 2,000 observations. In this range, there is a small fluctuation in the mean of the estimated coefficients. The next segment is the sample sizes between 2,000 and 8,000 observations, where the means of these estimated coefficients fluctuate more than those in the first segment. The least fluctuation appears in the third segment, which occurs between 8,000 and 10,500 observations. The last segment is between 10,500 and 12,680 observations. In this segment, the means fluctuate at about the same intensity as in the first segment. However, the fluctuation in the fourth segment is around the true value, whereas in the first segment, the fluctuation is below the true value.

Do the erratic movements of the means across different sizes still lie within their confidence intervals? Figure 5.3 depicts the movement of these means in microscopic form. Figure 5.4 depicts the movement of these means with their confidence intervals built around them. As illustrated in Figure 5.4, the confidence interval shows a convergence as sample size increases. This graph also shows that, despite the fluctuation

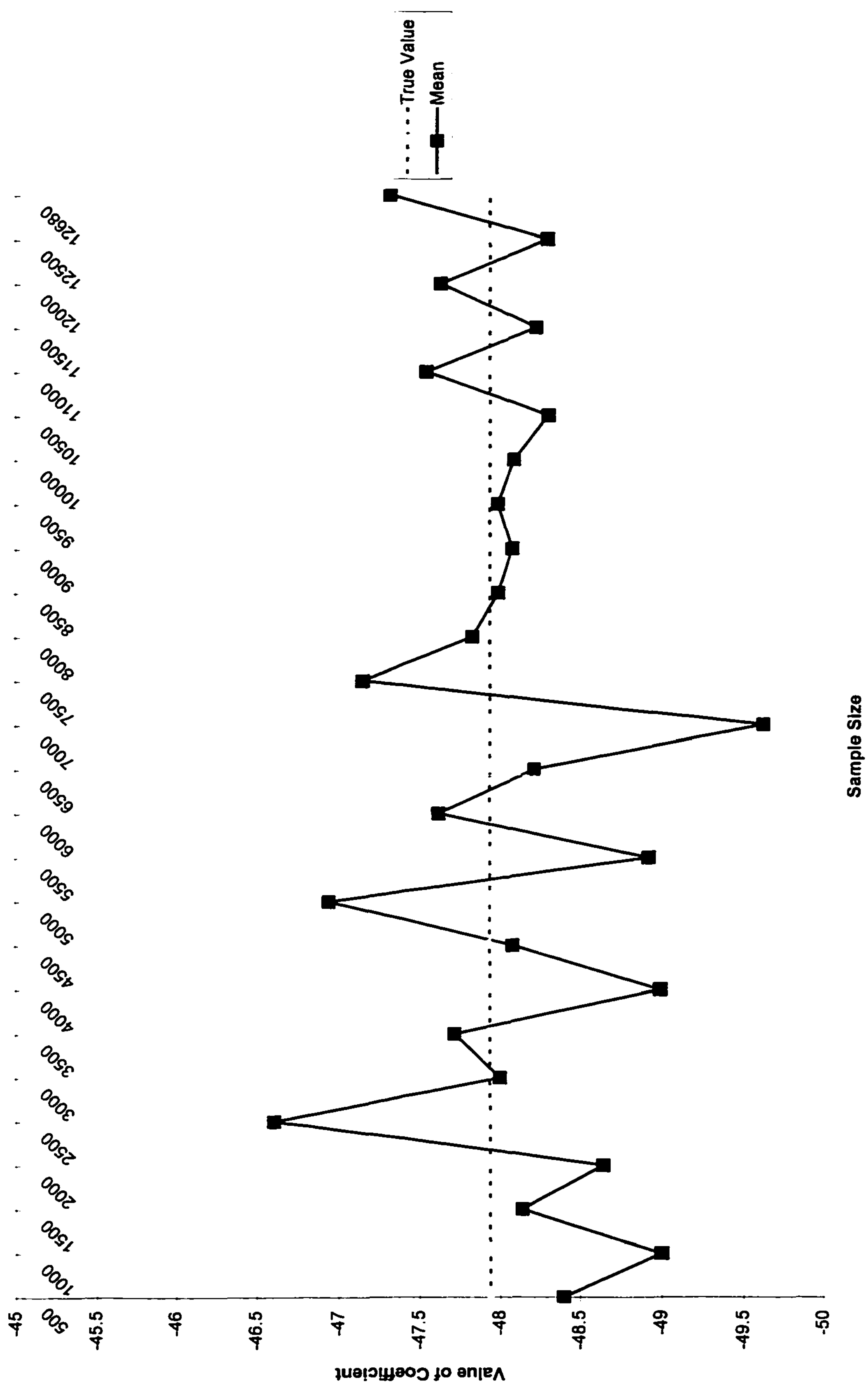
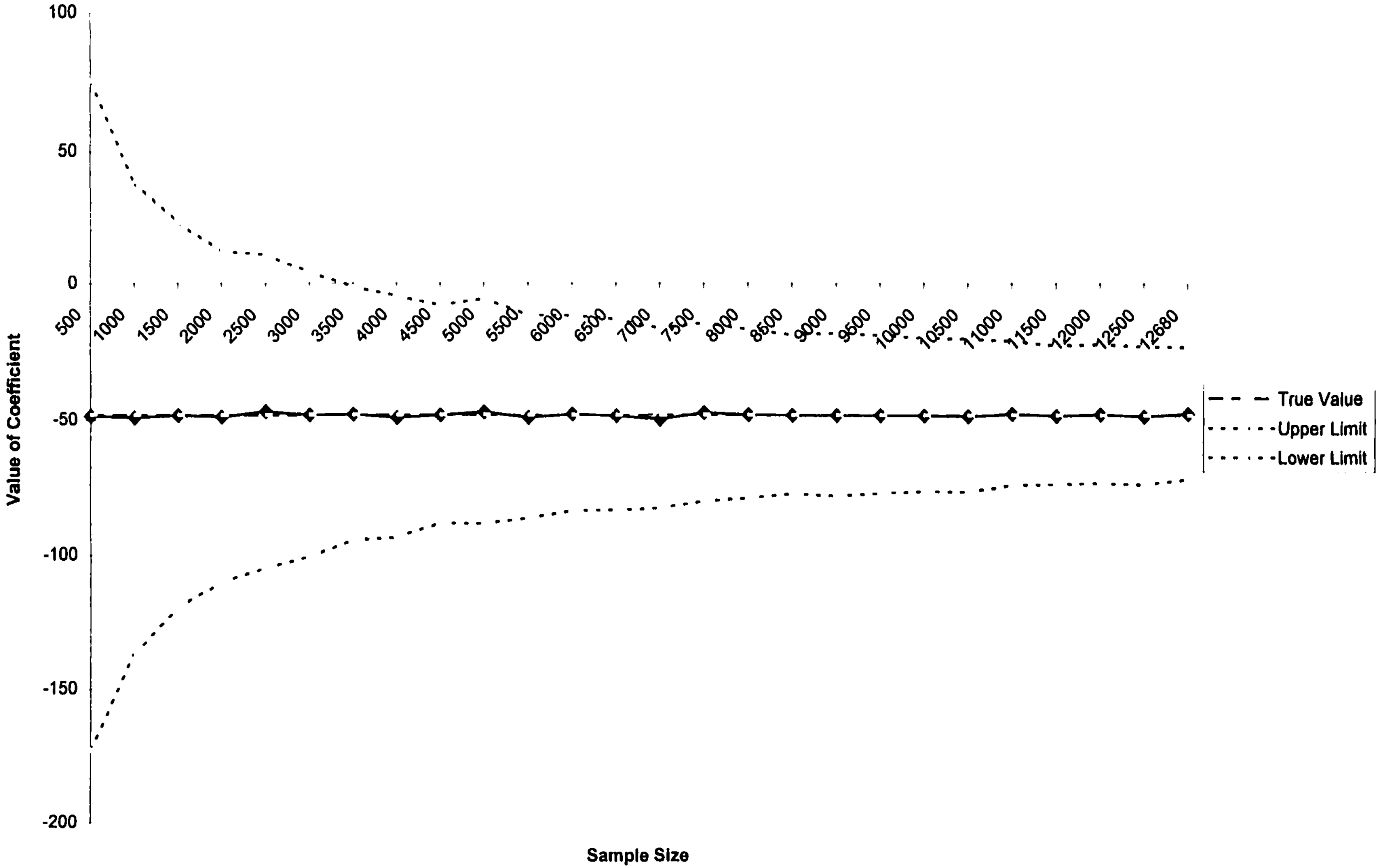


Figure 5.3: Mean Of Coefficient After 1,000 Iterations

Figure 5.4: Mean Of Coefficient After 1,000 Iterations With Confidence Interval



of the means of the estimated coefficients depicted in Figure 5.3, the means still lie within their confidence intervals at the 90% level of significance. This implies that the means of these estimated coefficients are not statistically different from the true value.

Figures 5.5 depicts the standard deviations of the means of the estimated coefficients. The graphs shows a negatively sloped curve which decreases dramatically for sample sizes smaller than 5,000, and then stabilizes as the sample size increases. The large standard deviations of the mean indicate that the estimated coefficients are unstable at small sample sizes, but become more stable as the sample size increases. Thus, this graph indicates that, for small sample sizes (less than 5,000 observations), the estimated coefficients are unstable. As the sample size increases, these estimated coefficients are more stable due to the lower standard deviations.

Figures 5.6 and 5.7 depict the means of the standard deviations and the standard deviations of these standard deviations respectively. Both graphs display a similar behaviour to Figure 5.5. The dashed line in Figure 5.6 is the value of the true standard deviation (10.941) of the ln(GW) variable. This graph implies that the standard deviations do approach the true standard deviation of the coefficient of ln(GW) as sample size increases. On the other hand, Figure 5.4 has shown that the means of the estimated coefficients fluctuate around the true value for all sample sizes and still lie within the confidence intervals.

Figure 5.7 depicts the standard deviations of those means illustrated in Figure 5.6. This graph also shows high variability in the mean of the standard deviations for small sample sizes. As the sample size increases, these standard deviations approach zero.

Figure 5.5: Standard Deviation Of The Mean

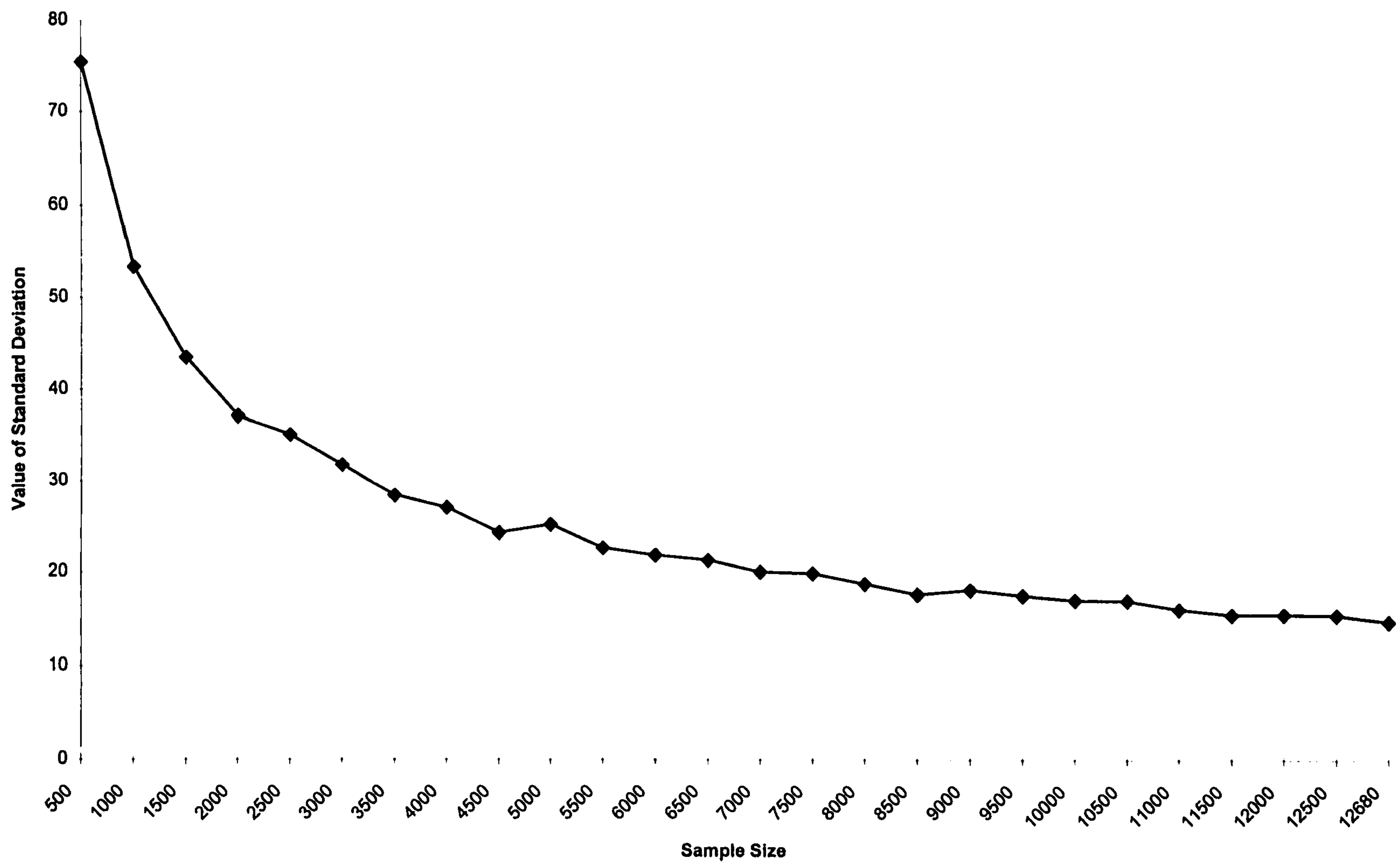


Figure 5.6: Mean Of The Standard Deviation

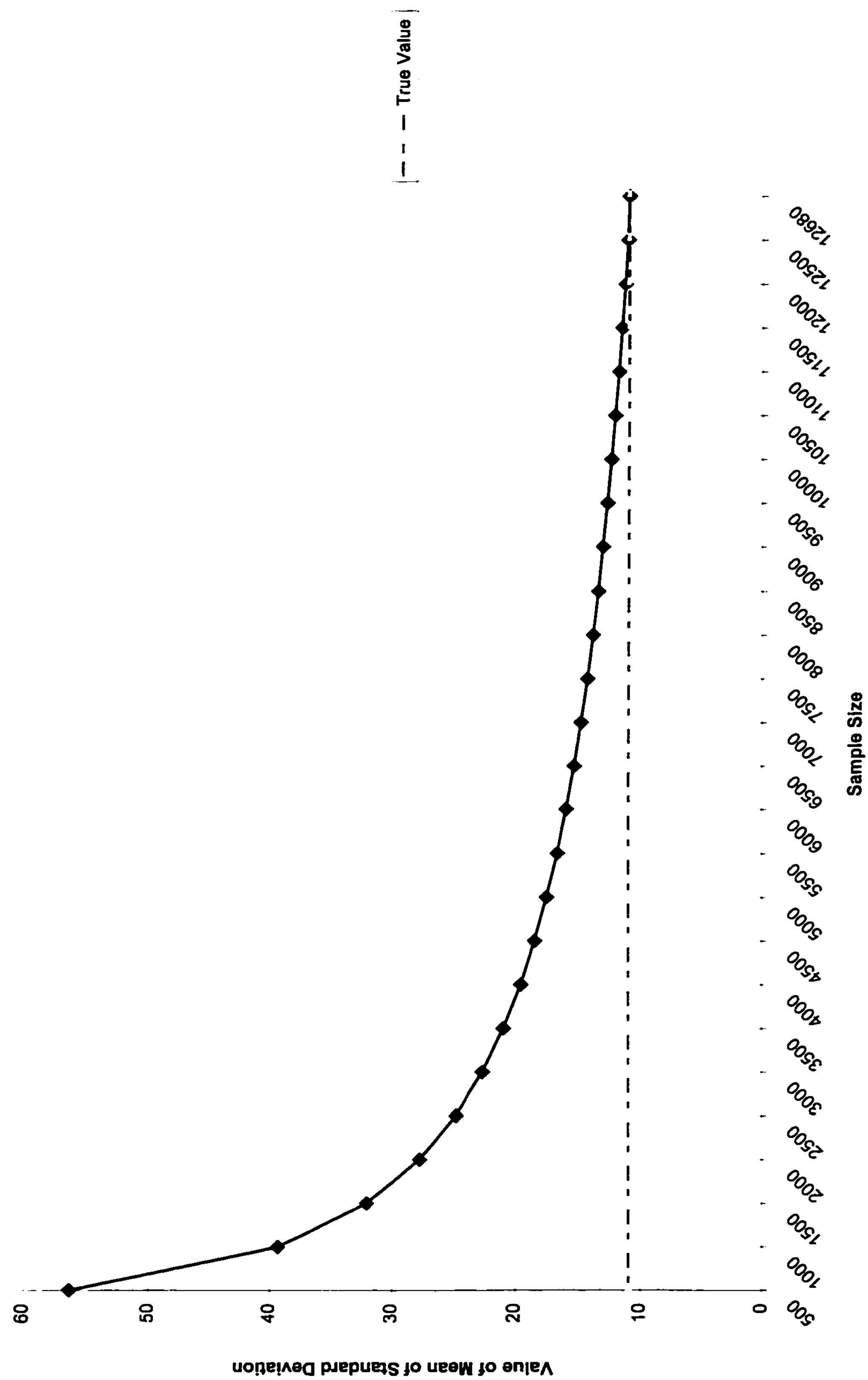
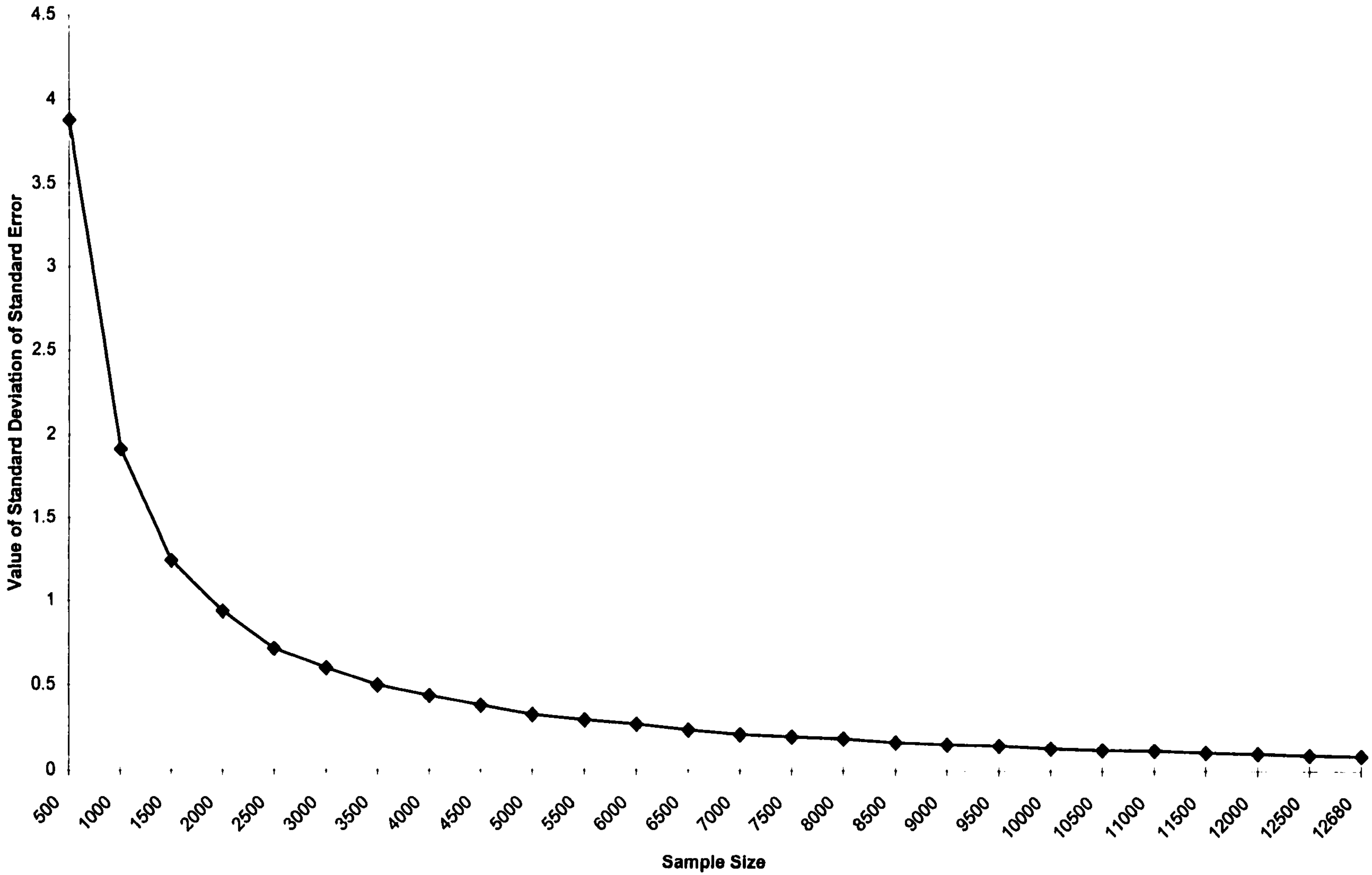


Figure 5.7: Standard Deviation Of The Standard Error



This behaviour implies that the standard deviations for the estimated coefficients are relatively more stable for larger sample sizes (greater than 5,000 observations) than for the smaller sample sizes.

From the graphs depicted, one can observe that the behaviour of the standard deviations are consistent regardless of the sample size. Figures 5.2, 5.5, 5.6, and 5.7 show the similar trend in the movement of standard deviations as sample size increases. All these graphs are downward sloping and convex to the origin. This implies that for small sample sizes, one can expect a high variability in the estimated coefficients in repeated sampling. But for sample sizes larger than 5,000, the variability begins to decline. However, these visual inspections of the properties of the estimated coefficient of the ln(GW) are not sufficient. Statistical testing is necessary to provide empirical evidence of the behaviour of these estimated coefficients.

The first test is the test of significance using a simple *t*-test on the means of these 1,000 estimated coefficients for all different sample sizes at the 1% level of significance. In this test, the null hypothesis is defined as the mean of the estimated coefficients at each sample size is equivalent to the true parameter (-47.938). The results of the *t*-test are illustrated in Table 5.2. For all sample sizes, the *t* values of the estimated coefficients are very small. This results in not rejecting the null hypothesis for all sample sizes, and implies that all these means are not statistically different from the true value because at the 1% level of significance these means are not statistically different from the true value. Hence, one can expect that at any lower levels of significance, the null hypothesis will also not be rejected at all sample sizes.

Table 5.2: Test Of Significance On The Mean Of The 1,000 Estimated Coefficients For Various Sample Sizes

$H_0: \bar{\beta}_N = \beta$

$H_A: \bar{\beta}_N \neq \beta$

Sample Size	*t* (calculated)	*p*-value
500	-0.00612	0.99512
1,000	-0.01992	0.98411
1,500	-0.00464	0.99630
2,000	-0.01903	0.98482
2,500	0.03788	0.96979
3,000	-0.00197	0.99843
3,500	0.00786	0.99373
4,000	-0.03915	0.96878
4,500	0.00594	0.99526
5,000	0.03953	0.96848
5,500	-0.04362	0.96522
6,000	0.01467	0.98830
6,500	-0.01325	0.98943
7,000	-0.08419	0.93292
7,500	0.03964	0.96839

Table 5.2: Test Of Significance On The Mean Of The 1,000 Estimated Coefficients For Various Sample Sizes (continued)

$H_0 : \bar{\beta}_N = \beta$

$H_A : \bar{\beta}_N \neq \beta$

Sample Size	*t* (calculated)	*p*-value
8,000	0.00577	0.99540
8,500	-0.00301	0.99760
9,000	-0.00776	0.99381
9,500	-0.00296	0.99764
10,000	-0.00896	0.99285
10,500	-0.02187	0.98256
11,000	0.02454	0.98043
11,500	-0.01910	0.98477
12,000	0.02000	0.98405
12,500	-0.02387	0.98096
12,680	0.04194	0.96655

* The *t*-test is defined as $t = \dfrac{\bar{\hat{\beta}}_N - \beta}{se(\bar{\hat{\beta}}_N)}$, where $N = 1, ..., 1{,}000$.

A pairwise test is performed on the estimated coefficients to test if they are statistically different from one another at each sample size. Table 5.3 provides a summary of the results from the pairwise test. This table presents the percentage of the 350 estimated coefficients of the ln(GW) variable generated at each sample size that are statistically different from each other estimated coefficient at the 1%, 5%, and 10% levels of significance. Thus, the null hypothesis is that the estimated coefficient generated from one run is equivalent to the estimated coefficient generated from another run, at a given sample size. At a level of significance, say 10%, one would expect to reject the null hypothesis 10% of the time. However, it is interesting that, for all sample sizes, the results violate this theoretical assumption. For example, at the 10% level of significance, the null hypothesis is rejected, on the average, about 23.3% of the time. At 5%, the average rejection rate is about 16.1%, while at 1%, the average rejection rate is about 7.1%. Thus, the results from the pairwise test suggest that, even at sample sizes larger than 10,000 observations, the rejection rate of the null hypothesis is larger than the level of significance chosen. To reinforce the findings from the pairwise test, a simple *t*-test is used.

The *t*-test is used to test if the estimated coefficients are statistically different from the true value. The results from the *t*-test are presented in Table 5.4. The significance level of 10% is chosen to investigate the behaviour of the estimated coefficients, and there are only 200 estimated coefficients generated for this *t*-test in each sample size. The null hypothesis is that each estimated coefficient at a given sample size is equivalent to the true value. Table 5.4 illustrates that the percentages of rejecting the null hypothesis

Table 5.3: Percentage Of Pairwise Tests That Reject The Null Hypothesis*

$H_0: \beta_i = \beta_j$

$H_A: \beta_i \neq \beta_j$

Sample Size	@ Pr = 90%** C.V. = 2.7055***	@ Pr = 95%** C.V. = 3.84146***	@ Pr = 99%** C.V. = 6.63490***
500	23.1	14.9	6.9
1,000	23.7	15.4	7.7
1,500	22.3	16.0	6.6
2,000	23.1	16.6	5.1
2,500	21.7	15.1	5.7
3,000	24.0	15.7	4.6
3,500	23.4	14.9	6.3
4,000	21.4	15.4	7.1
4,500	24.3	15.1	6.3
5,000	24.3	15.4	7.1
5,500	20.3	14.9	6.3
6,000	20.6	14.0	6.6
6,500	20.0	13.7	6.9
7,000	22.6	15.1	8.0
7,500	23.7	17.1	8.0
8,000	22.3	16.9	7.7

Table 5.3: Percentage Of Pairwise Tests That Reject The Null Hypothesis* (continued)

$H_0: \beta_i = \beta_j$

$H_A: \beta_i \neq \beta_j$

Sample Size	@ Pr = 90%** C.V. = 2.7055***	@ Pr = 95%** C.V. = 3.84146***	@ Pr = 99%** C.V. = 6.63490***
8,500	23.4	16.0	8.0
9,000	24.0	17.1	6.9
9,500	24.9	16.0	6.6
10,000	24.6	16.6	7.1
10,500	24.9	16.9	7.1
11,000	23.4	17.7	8.3
11,500	25.4	17.7	8.3
12,000	25.7	18.0	9.1
12,500	24.3	19.7	10.3

* The pairwise test is defined as $\chi_1^2 = \frac{(\hat{\beta}_i - \hat{\beta}_j)^2}{\hat{\sigma}_i^2 + \hat{\sigma}_j^2}$. This test is performed within each sample size and no test is performed across sample sizes.
** This is a two-tail test.
*** C.V. represents the Critical Value from the chi-square distribution for the respective probability levels.

Table 5.4: Percentage Of *t*-tests That Reject The Null Hypothesis*

$H_0 : \beta_i = \beta$

$H_A : \beta_i \neq \beta$

Sample Size	@ α = 10%** U.T. = 1.645*** L.T. = -1.645***
500	25.0
1,000	27.0
1,500	21.0
2,000	19.5
2,500	21.5
3,000	25.5
3,500	26.0
4,000	28.5
4,500	20.5
5,000	27.5
5,500	22.5
6,000	19.5
6,500	22.5
7,000	29.5
7,500	21.0
8,000	21.5

Table 5.4: Percentage Of t-tests That Reject The Null Hypothesis* (continued)

$H_0: \beta_i = \beta$
$H_A: \beta_i \neq \beta$

Sample Size	@ α = 10%** U.T. = 1.645*** L.T. = -1.645***
8,500	21.0
9,000	28.0
9,500	28.0
10,000	21.0
10,500	19.5
11,000	24.0
11,500	22.0
12,000	23.5
12,500	27.0
12,680	19.5

* The t-test is defined as $t = \frac{\hat{\beta}_i - \beta}{se(\hat{\beta}_i)}$, where i = 1, ..., 200.

** This is a two-tail test.

*** U.T. and L.T. represent Upper Tail and Lower Tail respectively.

are unstable across the sample sizes, but that these percentages are very close to each other. There is not a trend in these percentages as sample size increases. For example, at sample size 2,000, 6,000, and 12,680, the rejection rates are identical (19.5%). Comparing Table 5.4 to Table 5.3, a similarity in the results from the pairwise test and the *t*-test is observed. The average percentage of rejecting the null hypothesis in Table 5.4 is about 23.5%, which is relatively close to that in Table 5.3. Despite the large variance of the estimated coefficients at small sample sizes, the percentage of estimates that are statistically different from the true value is surprisingly close to that at the larger sample sizes. The *t*-test has shown that the percentage of rejecting the null hypothesis is similar at the various sample sizes.

Intuitively, the reason underlying the constant rejection rates of the null hypothesis in both the pairwise test and the *t*-test for various sample sizes can be explained using Figures 5.1 and 5.2. The intense fluctuation in the estimated coefficients for small sample sizes is accompanied by the corresponding large standard errors. The comparatively large difference in the estimated coefficients from the true value in small sample sizes will increase the value of the numerator in the *t*-test. Their standard errors, however, are large enough to increase the value of the denominator for all estimated coefficients in each sample size. On the other hand, the small variations in the estimated coefficients for the larger sample sizes will decrease the value of the numerator; and the corresponding small standard errors will decrease the value of the denominator. Therefore, the result is a constant rejection rate of the null hypothesis, and this explains the reason underlying these surprising results. A similar explanation can also be given for the results from the pairwise test.

From the visual investigation of these graphs, one can perceive that there exists unstable behaviour of the estimated parameter when the sample size is small (sample size fewer than 5,000 observations). However, it is not appropriate to conclude that the estimates generated by the small sample sizes are incorrect since the statistical testing carried out in the previous section shows that the percentages of rejecting the null hypothesis are consistent for all sample sizes, even though those results indicate that the statistical assumption is violated. The next section will provide another statistical test -- constructing a confidence interval -- to support the graphical evidence presented.

5.3 Evidence Using Confidence Intervals

The evidence from the Monte Carlo simulation using confidence intervals is presented in this section. The upper limits and lower limits of the confidence intervals are constructed around the true value of the coefficient of the ln(GW) and its true standard error. Figures 5.8, 5.9, and 5.10 illustrate the percentages of these parameters that lie outside the 90%, 95% and 99% confidence intervals respectively, and Table 5.5 presents these results in tabular form. The interpretation for Table 5.5 is that for a given confidence interval, 90% for example, in the long run 90 out of 100 estimated coefficients generated will lie within (-65.94, -29.94). In all the three graphs depicted in Figures 5.8, 5.9, and 5.10, the percentages of the estimated coefficients that lie outside the respective

Figure 5.8: Percentage Of Coefficents Which Lie Outside The 90% Confidence Interval

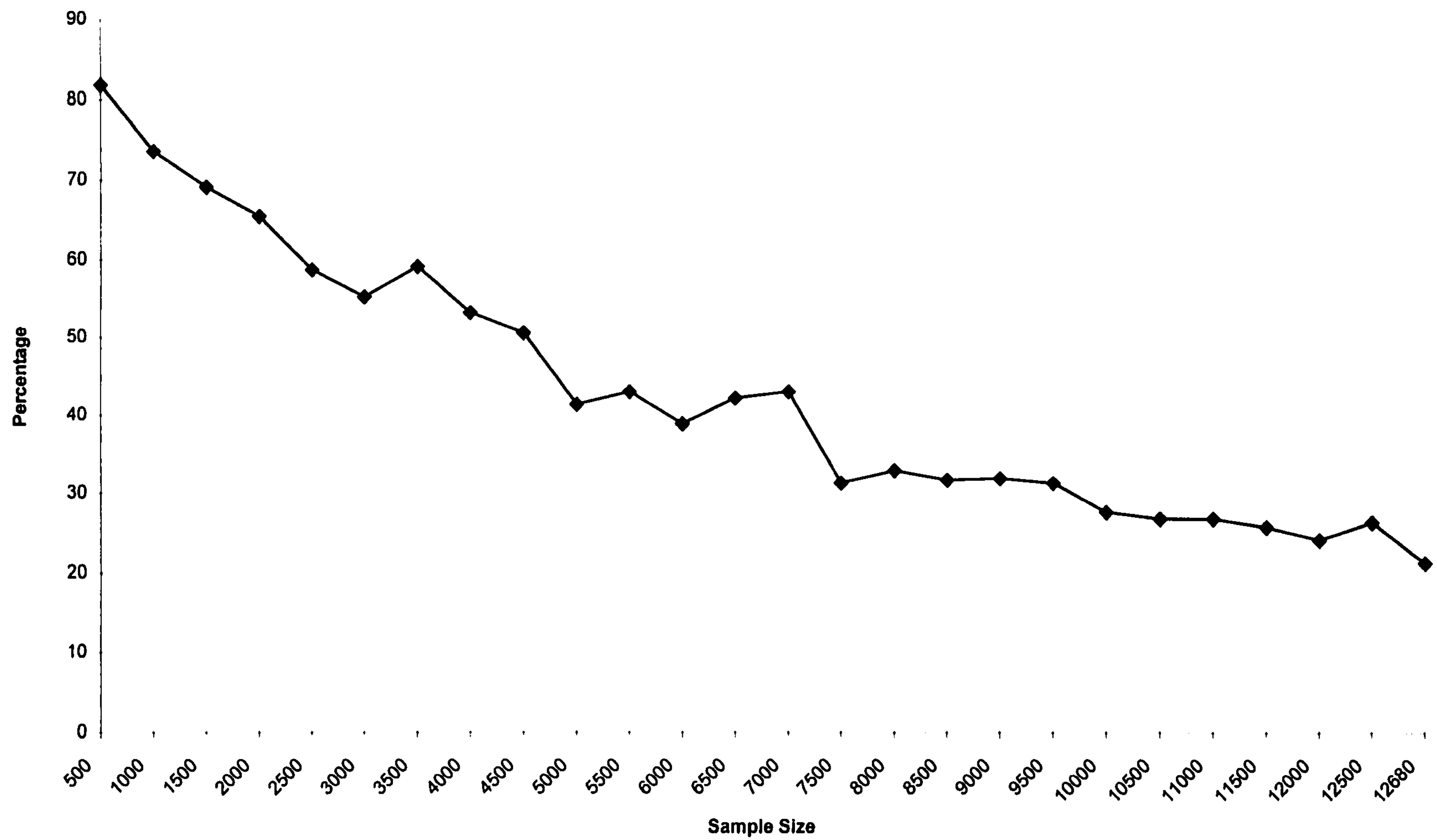


Figure 5.9: Percentage Of Coefficients Which Lie Outside The 95% Confidence Interval

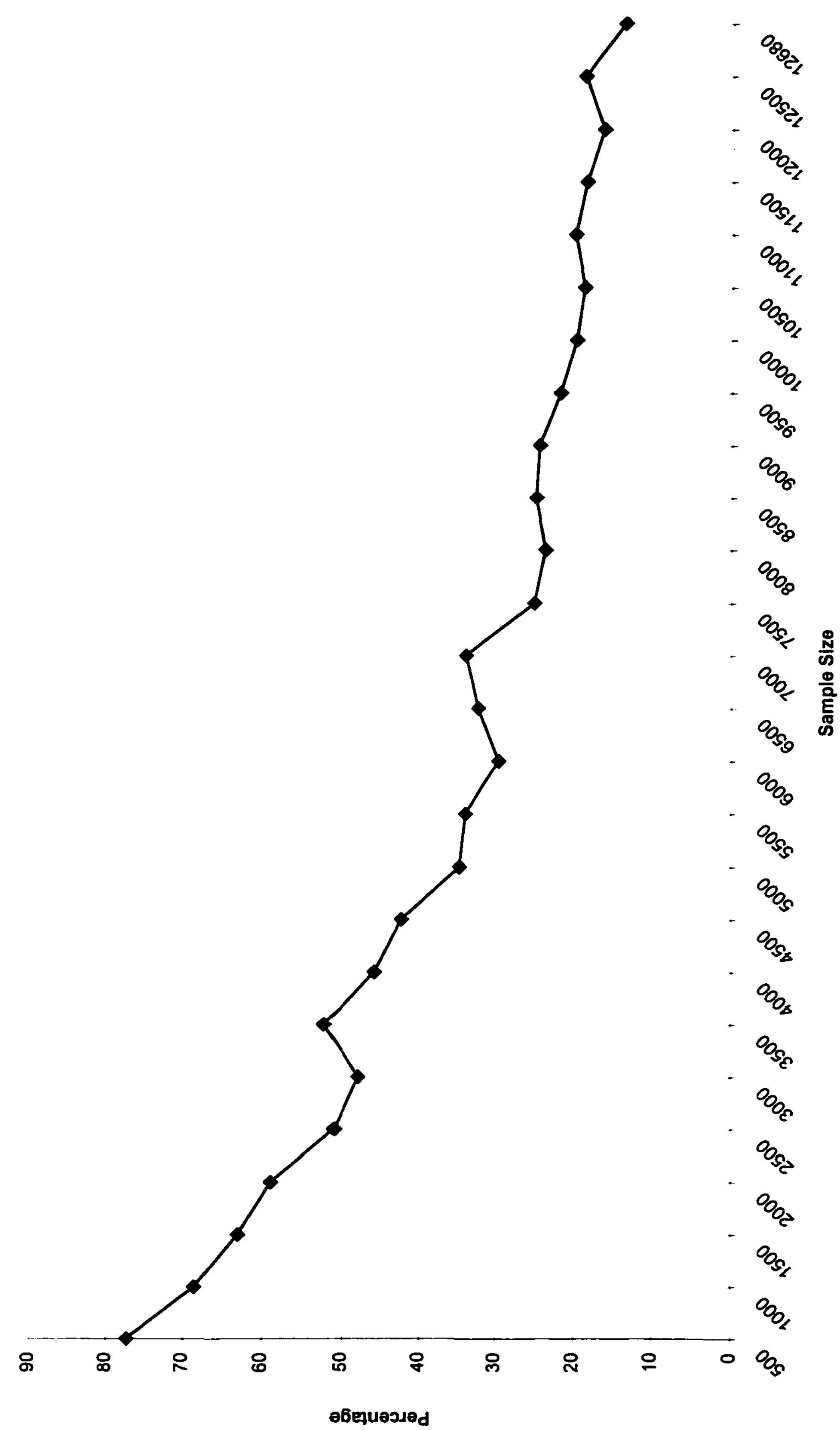


Figure 5.10: Percentage Of Coefficients Which Lie Outside The 99% Confidence Interval

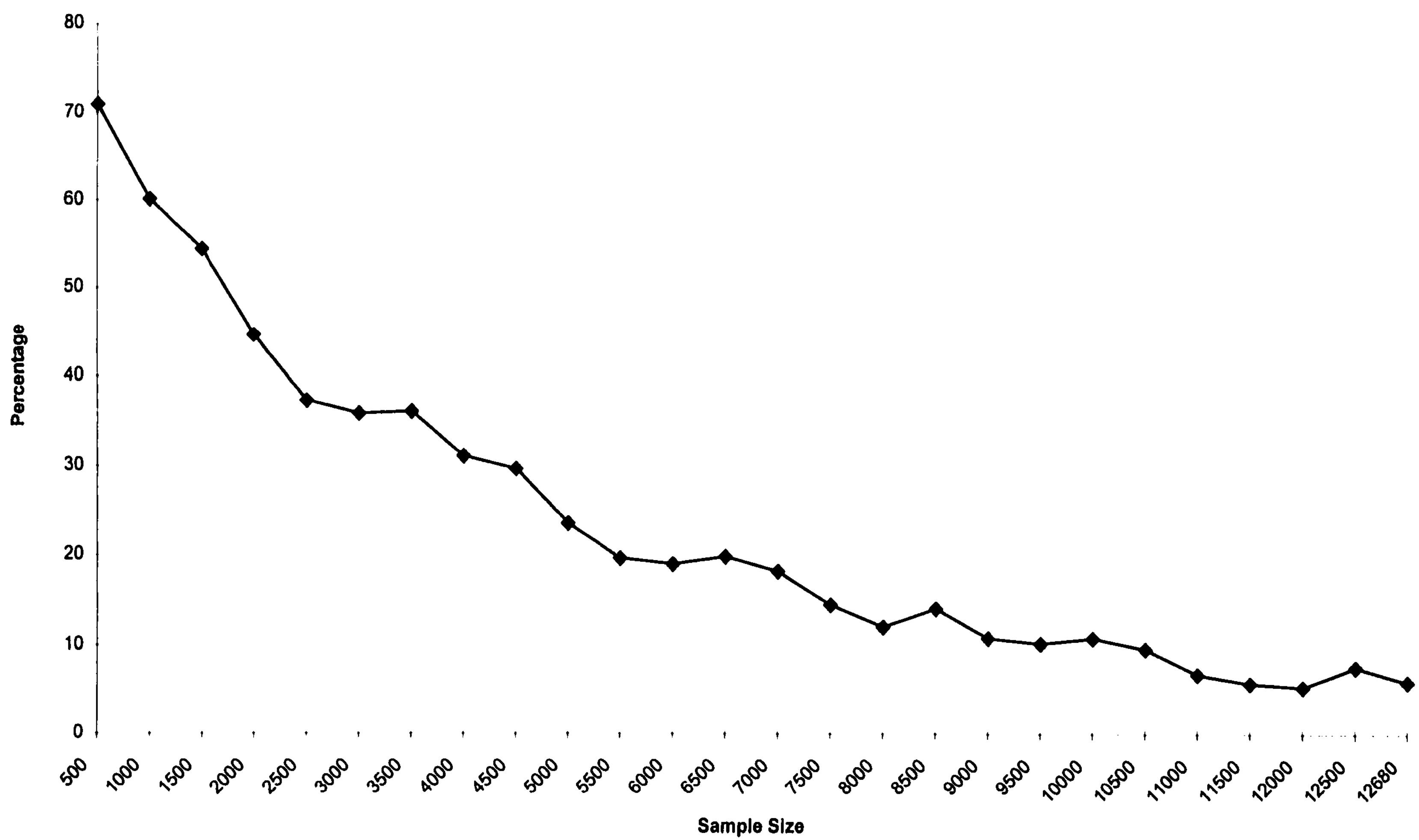


Table 5.5: Percentage Of The Estimated Coefficients That Lie Outside The Confidence Interval

Sample Size	@ 90% C.I.* (U.L. = -29.94)** (L.L. = -65.94)**	@ 95% C.I.* (U.L. = -26.49)** (L.L. = -69.38)**	@ 99% C.I.* (U.L. = -19.75)** (L.L. = -76.12)**
500	81.80	77.30	70.96
1,000	73.62	68.51	60.12
1,500	69.12	62.99	54.40
2,000	65.44	58.90	44.58
2,500	58.69	50.92	37.22
3,000	55.21	47.85	35.79
3,500	59.10	52.15	35.99
4,000	53.17	45.81	31.08
4,500	50.51	42.33	29.65
5,000	41.31	34.76	23.52
5,500	42.94	33.95	19.63
6,000	38.85	29.65	19.02
6,500	42.13	32.31	19.84
7,000	42.94	33.95	18.20
7,500	31.29	25.15	14.52
8,000	32.92	23.72	12.07
8,500	31.70	24.95	14.11

Table 5.5: Percentage Of The Estimated Coefficients That Lie Outside The Confidence Interval (continued)

Sample Size	@ 90% C.I.* (U.L. = -29.94)** (L.L. = -63.94)**	@ 95% C.I.* (U.L. = -26.49)** (L.L. = -69.38)**	@ 99% C.I.* (U.L. = -19.75)** (L.L. = -76.12)**
9,000	31.90	24.54	10.84
9,500	31.29	21.88	10.22
10,000	27.61	19.84	10.84
10,500	26.79	18.81	9.61
11,000	26.79	20.04	6.75
11,500	25.77	18.61	5.73
12,000	24.13	16.36	5.32
12,500	26.38	18.81	7.57
12,680	21.27	13.70	5.93

* C.I. represents Confidence Interval.
**U.L. and L.L. represent Upper Limit and Lower Limit respectively. They are calculated from equation (3.28).

confidence intervals are very high for small samples. For example, for the smallest sample size chosen, more than 70% of the estimated coefficients are outside the confidence interval. This violates the theoretical assumption of statistics when the level of significance is chosen. However, the graphs depict a slow convergence of the estimated coefficients towards the level of significance, although they are never close. Even when the total sample size of 12,680 observations is used, the percentage of the estimated parameter that lie outside the 90%, 95%, and 99% confidence intervals are 21.27%, 13.70%, and 5.93% respectively. This indicates that the variability of the estimated coefficients is relatively large at small sample sizes when compared to the variability of the estimated coefficients at the larger sample sizes. Thus, even at a very large sample size of more than 12,000 observations, there are more estimated coefficients which lie outside the confidence intervals than expected. This implies that, regardless of the size of a sample, there still exists an unstable behaviour with respect to the estimated coefficient.

5.4 Significance of The Estimated Coefficient and Decentralization

The labour supply function in this thesis is widely used by other labour economists in the literature. Theoretically, the estimated coefficient of ln(GW) is statistically different from zero. However, investigating the *t*-ratio of the estimated coefficients of ln(GW) shows that there is a substantial percentage of the estimated coefficients are not statistically different from zero at the small sample sizes. Table 5.6

Table 5.6: Percentage Of Estimated Coefficients That Are Statistically Different From Zero

H_0 : The coefficient of Lnwage is not statistically different from zero.
H_A : The coefficient of Lnwage is statistically different from zero.

Sample Size	Reject The Null (Out of 200 Iterations)	In Percentage
500	63	31.5
1,000	68	34.0
1,500	94	47.0
2,000	104	52.0
2,500	112	56.0
3,000	132	66.0
3,500	135	67.5
4,000	156	78.0
4,500	161	80.5
5,000	149	74.5
5,500	160	80.0
6,000	174	87.0
6,500	176	88.0
7,000	183	91.5
7,500	176	88.0
8,000	186	93.0

Table 5.6: Percentage Of Estimated Coefficients That Are Statistically Different From Zero (continued)

H_0 : The coefficient of Lnwage is not statistically different from zero.
H_A : The coefficient of Lnwage is statistically different from zero.

Sample Size	Reject The Null (Out of 200 Iterations)	In Percentage
8,500	183	91.5
9,000	187	93.5
9,500	181	90.5
10,000	192	96.0
10,500	185	92.5
11,000	191	95.5
11,500	194	97.0
12,000	194	97.0
12,500	194	97.0
12,680	197	98.5

illustrates the percentage of the estimated coefficients that are statistically different from zero at various sample sizes. For a sample size of 500, there are only 63 estimated coefficients out of a total of 200 (31.5%) that are statistically different from zero. As the sample size gets larger, this percentage also increases. At about a sample size of 6,000, there are only 26 estimated coefficients (13%) that are not statistically significant from zero. With a sample size of 12,680, there is a total of 197 estimated coefficients out of the 200 generated (98.5%) that are statistically different from zero.

A final measure of performance of the estimator is decentralization. Here, decentralization means the number of times the estimated coefficient has the wrong sign -- that is, does not conform to the a priori expectation. Table 5.7 presents this result for 489 iterations, and Figure 5.11 depicts this result in graphical form. When 500 observations are used, about a quarter of the 489 estimated coefficients have the wrong the sign. Decentralization decreases drastically up to the sample size of 2,500 observations. Thereafter, there is a small decrease in decentralization up until about 9,000 observations, where almost no estimated coefficient has the wrong sign. However, there still exists a notable number of estimated coefficients that have the wrong sign for any sample size smaller than 5,000 observations. This result implies that the variability in the estimated coefficients at a sample size smaller than 5,000 observations is relatively large compared to those coefficients obtained at a larger sample size. Therefore, using any sample size smaller than 5,000 observations increases the probability of obtaining an estimated coefficient with a wrong sign.

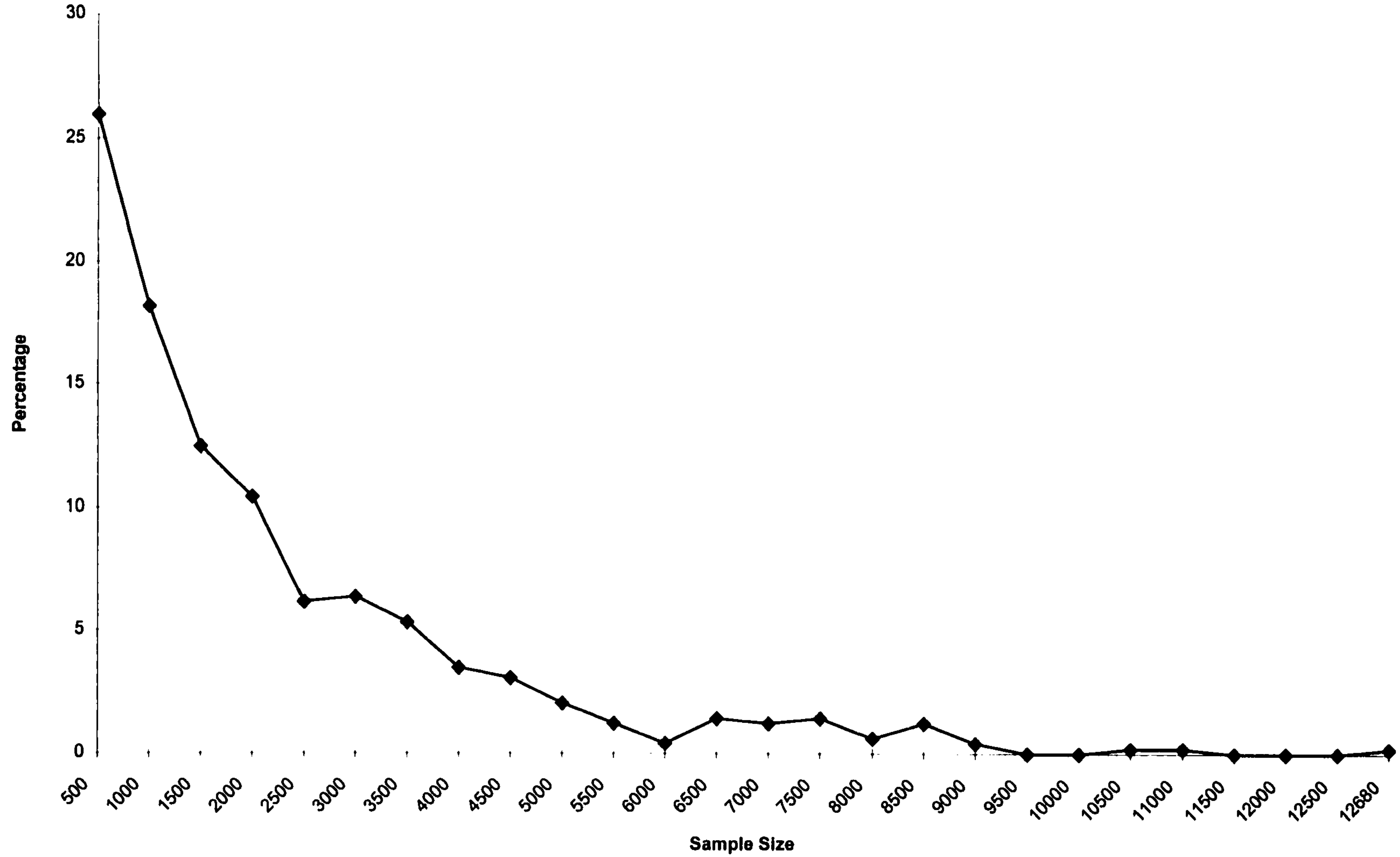
Table 5.7: Percentage Of Estimated Coefficients That Have The Wrong Sign

Sample Size	Number of Wrong Sign (Out of 489 Iterations)	In Percentage
500	127	25.97
1,000	89	18.20
1,500	61	12.47
2,000	51	10.43
2,500	30	6.13
3,000	31	6.34
3,500	26	5.32
4,000	17	3.48
4,500	15	3.07
5,000	10	2.04
5,500	6	1.23
6,000	2	0.41
6,500	7	1.43
7,000	6	1.23
7,500	7	1.43
8,000	3	0.61
8,500	6	1.22

Table 5.7: Percentage Of Estimated Coefficients That Have The Wrong Sign (continued)

Sample Size	Number of Wrong Sign (Out of 489 Iterations)	In Percentage
9,000	2	0.41
9,500	0	0
10,000	0	0
10,500	1	0.21
11,000	1	0.21
11,500	0	0
12,000	0	0
12,500	0	0
12,680	1	0.21

Figure 5.11: Percentage Of Coefficients With The Wrong Sign



5.5 Concluding Remarks

The constant rejection of the null hypothesis in Table 5.2 can be explained by the large variances of the estimated coefficients at the smaller sample sizes. The variances of the estimated coefficients at small sample sizes are so large that rejecting the null hypothesis is almost impossible. Similarly, in Tables 5.3 and 5.4, the large variances of the smaller sample sizes have caused the rejection rates on the null hypothesis to be similar to those at the larger sample sizes. The large variances have caused the test to fail to reject the null hypotheses at small sample sizes.

Table 5.6 has shown that many of the estimated coefficients at small sample sizes are, in fact, statistically indistinguishable from zero. This means that the wage variable has no effect on labour supply. This is contrary to theoretical expectations. In the literature of labour economics, the coefficient of ln(GW) is expected to be statistically different from zero. Moreover, Table 5.7 has shown that there are relatively more estimated coefficients with the wrong sign at small sample sizes than at larger sample sizes. Figure 5.1 also shows that there is a high volatility in the estimated coefficient at small sample sizes. Figures of the standard deviations of the estimated coefficient imply stable behaviour in large sample sizes. Given all these results, it is evident that at any sample sizes smaller than 6,000 observations, there is inconsistency in the behaviour of the estimated coefficient. The large variances at small sample sizes have resulted in imprecise estimation of the coefficient of ln(GW).

CHAPTER 6: SUMMARY AND CONCLUSIONS

This chapter is divided into four sections. Section 6.1 provides a review of the application of the Monte Carlo techniques in this study. Section 6.2 summarizes the highlights of the major findings of this study, and section 6.3 provides the implications of the results. Section 6.4 indicates the limitations of this thesis and makes recommendations for future research.

6.1 Review

Monte Carlo techniques are applied to investigate the stability of the estimated coefficients for different sample sizes because it allows the freedom to choose the true values of the parameters, and to compare the estimated coefficients to the true value. The model employed in this study is the semi-log labour supply function which is commonly used by labour economists. There are a total of 20,160 observations in the data, including both workers and nonworkers. The data identify those who are not satisfied with their total weeks of work. Individuals who are not constrained in weeks of work can be considered as a selective sample from the labour force. This form of sample selection bias is corrected by introducing an inverse Mills ratio in the labour supply function. To calculate the inverse Mills ratio, a probit model is used. The probit model calculates the probability that an individual is quantity constrained, that is, would prefer additional weeks of work. Thus, the dependent variable (hours of work) in the probit model is a dummy variable which determines if an individual is underemployed or not.

Once the inverse Mills ratio is calculated from the probit model, it is substituted into the semi-log labour supply equation as an explanatory variable. Only seven out of the twenty explanatory variables in the probit model have been substituted back into the semi-log labour supply equation as part of the explanatory variables. This is to prevent a serious collinearity problem and to satisfy the condition for identification in Heckman's model (Maddala, 1983: 233).

The semi-log labour supply equation, corrected for sample selectivity, is estimated using the ordinary least squares regression (OLS) with a total sample size of 12,680 for male workers who are satisfied with their total weeks of work. Estimates of the parameters are obtained and are assumed to be the true parameters of the model. Using the random number generating process in *Shazam* (1993), a disturbance term for each observation in the data is generated. From the estimated model and the randomly generated error term, the predicted value of the dependent variable is generated.

Using different sample sizes, and combining all values of the explanatory variables and generated dependent variable, estimates of the parameters of the labour supply equation can be obtained. In this experiment, the sample sizes are increments of 500 observations until all observations are exhausted. Numerous experiments are performed for each subsample to avoid any bias constraints. Subsamples are randomly drawn using a random number generating process.

The effect of sample size is revealed in a number ways:

1. a *t*-test is conducted to see if the estimated coefficients at different sample sizes are statistically different from the true value of the parameter;
2. a pairwise test is devised to test if each estimated coefficient is statistically different from one another at each sample size;
3. confidence intervals are constructed around the true value and its standard error to investigate the percentage of the estimated coefficients that lie within these boundaries for the various sample sizes;
4. the percentage of the estimated coefficients which are statistically indistinguishable from zero; and
5. the percentage of the estimated coefficients which have the wrong sign.

6.2 Summary of Findings

Visual inspection of the behaviour of the estimated coefficients across different sample sizes shows that there exist large variability in these estimators for small sample sizes. The graph appears to show the unstable behaviour of the estimated coefficients for sample sizes smaller than 6,000 observations. Graphical presentation of the standard deviations also shows that, at small sample sizes, the estimated coefficients have relatively larger variances. However, the results from the *t*-test show that, at every sample size, the null hypothesis (that is, the estimated coefficients at each sample are statistically indistinguishable from the true value) is rejected at an average rate of 23.5%. This indicates that, for any sample size chosen, only about 23.5% of the estimated coefficients are statistically different from the true value. Similarly, the results of pairwise testing (that is, the null hypothesis is that each estimated coefficient is not statistically different in pairwise testing at each sample size) show that the null hypothesis is rejected at an average rate of 23.3%. This result implies that, for a same sample size, only 23.3% of the estimated coefficients are statistically different from the other estimated coefficients. The large variances of the estimators at the smaller sample sizes have made it impossible to reject the null hypotheses. The test on the mean of 1,000 estimated coefficients at small sample sizes also failed to reject the null hypothesis (that is, the null hypothesis is that the mean of the estimated coefficients at each sample size is not statistically different from the true value). Again, this failure of not rejecting the null hypothesis is attributed to the large variances at small sample sizes.

When confidence intervals are constructed, the percentage of the estimated coefficients that lies outside these boundaries is so much higher at small sample sizes. Although the test on the mean of 1,000 estimated coefficients fails to reject the null hypothesis for all sample sizes, constructing the confidence intervals has shown that, at 99% confidence interval, more than 20% of the 1,000 estimated coefficients lie outside the allowed boundaries for sample sizes smaller than 6,000 observations. This indicates that the variances of the estimated coefficients at small sample sizes are large, and thus, results in inconsistent estimates of the estimated coefficients.

The coefficient of ln(GW) is expected to be statistically different from zero on the basis of the underlying economic theory. Unfortunately, this is not the case for small sample sizes. Results have shown that, for sample size smaller than 1,500 observations, less than half of the estimated coefficient are statistically different from zero. As the sample size increases, however, the percentage of the estimated coefficients which are statistically different from zero increases. This result indicates that sample size does have an impact on the estimated coefficients. Decentralization, which is a measure for the number of times the estimated coefficients have the wrong sign, is also high for small sample sizes. This result indicates that there is a higher proportion of the estimated coefficients which have the wrong sign in small sample sizes. This explains that previous results of uncompensated wage varies from positive to negative.

6.3 Implications of Results

The results from both the pairwise and *t*-test have shown an identical and constant rejection rate of the null hypotheses. This unusual behaviour of the estimated coefficients is attributed to the large variances at small sample sizes. This is especially true for sample size less than 6,000 observations. The relatively large variances of the estimated coefficients at small sample sizes imply that there is huge discrepancy in these estimated coefficients. This causes the estimated coefficients at small sample sizes to vary over a wide range. Graphical presentation of the variances of the estimated coefficients at each sample size indicates that the variances approach a constant value as the sample size surpass 6,000 observations.

The evidence from constructing the confidence intervals shows that a higher percentage of the estimated coefficients lie outside these boundaries for small sample sizes. This implies that the large variances at the small sample sizes have caused these large variations in the estimated coefficients. As the sample size increases, however, the percentage of the estimated coefficients that lie outside the boundaries decreases. Therefore, at larger sample sizes, the results show that the estimated coefficients are more consistent, and have relatively smaller variances too.

The result from the test of significance on the estimated coefficient of the ln(GW) have shown that at small sample sizes, there exists relatively lesser estimated coefficients which are statistically different from zero. This suggests that, for small sample sizes, there is a higher probability that the estimated coefficients are statistically indistinguishable from zero. This result implies inconsistency in the estimated coefficient at small sample sizes. The result from decentralization also shows that the estimated coefficients at small sample sizes have higher tendency to produce estimates with the wrong sign, and are thus inconsistent.

It is concluded that, given a semi-log labour supply model, the estimated coefficients exhibit unstable characteristics and large variances for small sample sizes. As the sample size increases, however, the asymptotic properties of the estimated coefficient begins to show. There is relatively much lesser variability in the estimated coefficients for sample sizes larger than 6,000 observations, and the variances are also relatively smaller. This empirical evidence has led to question the validity of the results of labour supply studies using small sample size.

6.4 Areas of Future Research

Although this study is one of the few to analyze the stability of the estimated coefficients in a labour supply model using Monte Carlo techniques, there exists several areas which should be clarified in order to direct future research. There exist two major limitations of the present study: data limitations, functional specification and methodology.

For any empirical research, the quality of and quantity of the data is crucially important since the statistical results are only as good as the data. The present study employed data originating from the Labour Market Activity Survey. Although it was perhaps the best data available at the time, the industry unemployment rate is not available. Other sources, like the Canadian Socio-economic Information Management and the Statistics Canada catalogue, are used to obtain the figures. There are 52 industry and occupation codes found in the LMAS. However, there are only 14 industry codes found in the aforementioned sources. Assigning 14 industry unemployment rates to 52 others can be a difficult task because many definitions are close to each other. Moreover, there are discrepancies between these two sources. This may lead to statistical bias in estimating the inverse Mills ratio, and thus, the coefficients in the linear regression equation.

The heterogeneous nature of the data itself has caused the coefficient of determination (R^2) to be very small. This is not, however, uncommon in the labour supply literature. This implies that only a small proportion of the variation in the dependent variable (hours of work) is explained by the independent variables, and a large proportion of the variation in the dependent variable is, hence, attributed to random forces. Therefore, due to the nature of the data, the low value of the coefficient of determination should not be a surprising issue in estimating labour supply.

The semi-log labour supply function estimated in this research groups all employment categories into one aggregate index. As such, the wage coefficient and elasticity estimates represent this aggregate grouping. This can be generalized to allow different elasticity estimates over different employment groupings by using dummy slope variables to identify the different employment categories.

The Monte Carlo technique has assisted in investigating the consistency of the estimated coefficient. However, the *t*-test and pairwise test have shown constant rejection of the null hypotheses in all sample sizes. These results are undesirable. Apart from using the Monte Carlo methodology, one can employ the Efron (1979) Bootstrapping Method. The idea of bootstrapping is to use the single available data set to design a sort of Monte Carlo experiment in which the data themselves are used to approximate the distribution of the error terms or other random quantities in the model. These error terms or random quantities are usually not drawn from an assumed distribution, such as the normal, but rather from the empirical distribution function of their sample counterparts. However, the bootstrap method can be a very expensive in terms of computer time, and is an inaccurate way of getting OLS standard errors.[21] As the costs of computers decrease

[21] Davidson, R, and J. G. MacKinnon (1993), pp. 763.

and the speed becomes faster, it will be economical to use the bootstrap method in such applied work.

Realizing the limitations of this thesis, one has to remember this study is one of the few to investigate the effect of sample size on the stability of the estimated coefficient of labour supply elasticity using Monte Carlo techniques. The refinement in the methodology employed and functional specification with the increase in the quality of the data will allow future researcher to provide a more concrete evidence on the effect of sample size on the estimated coefficient in the labour supply literature.

REFERENCES

Ashenfelter, O. (1980). "Unemployment as Disequilibrium in a Model of Aggregate Labour Supply," *Econometrica*, **48**, 547-64.

Ashraf, J. (1992). "Union Wage Premiums in an Instrumental Variables Framework," *Journal of Labor Research*, **13**, 231-36.

Berndt, E. R. (1990). *The Practice of Econometrics: Classic and Contemporary*. "Whether and How Much Women Work for Pay: Applications of Limited Dependent Variable Procedures," Massachusetts:Addison-Wesley Publishing Company, Inc., 593-65.

Blomquist, S. (1996). "Estimation Methods for Male Labor Supply Functions: How to Take Account of Nonlinear Taxes," *Journal of Econometrics*, **70**:2, 383-405.

Blundell, R., A. Duncan, and C. Meghir (1992). "Taxation in Empirical Labour Supply Models: Lone Mothers in the UK," *Economic Journal*, **102**, 265-78.

Borjas, G. (1980). "The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias," *Journal of Human Resources*, **15**, 424-34.

Bourguignon, F., and T. Magnac (1990). "Labour Supply and Taxation in France," *Journal of Human Resources*, **25**:3 358-89.

Brown, C.V., ed. (1981). *Taxation and Labour Supply*, London: George Allen and Unwin.

Brown, L., and Ulph (1981). "The Basic Model," in Browns (eds.) *Taxation and Labour Supply*, London: George Allen and Unwin, 35-52.

Burtless, G., and J. A. Hausman (1978). "The Effect of Taxation on Labour Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy*, **86**:2, 1103-30.

Buslenko, N. P., D. I. Golenko, Y. A. Shreider, I. M. Sobol', and V. G. Sragovich, (1966). *The Monte Carlo Method: The Method of Statistical Trials*, translated by G. J. Tee, ed. by Y. A. Shreider, Vol 87. Oxford: Pergamon Press Ltd.

Cragg, J. G. (1967). "On The Relative Small-Sample Properties Of Several Structural-Equation Estimators," *Econometrica*, **35**:1, 89-110.

Davidson, R., and J. G. MacKinnon (1993). *Estimation and Interence in Econometrics*, New York: Oxford University Press, 763-768.

Dickens, W. T., and S. J. Lundberg (1985). "Hours Restrictions and Labour Supply," *NBER Working Papers*, 1638.

Doan, T. A. (1995). *RATS User's Manual Version 4*. Estima.

Donner, A. (1984). "Approaches to Sample Size Estimation in the Design of Clinical Trials -- A Review," *Statistics in Medicine*, **3**, 199-214.

Dougherty, C. (1992). *Introduction to Econometrics*, New York: Oxford University Press, 76-80.

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jacknife," *Annals of Statistics*, Vol. 7.

Goldberger, A. S. (1981). "Linear Regression After Selection," *Journal of Econometrics*, **15**, 357-66.

Gordon, D. V., L. Osberg,and S. Phipps (1991). "How Big is Big Enough? The Problem of Adequate Samle Size in Empirical Studies of Labour Supply," *Working Paper No. 91-93*, Department of Economics, Dalhousie University.

Gosset, W. S. (1908). *Student Probable Error of a Correlation Coefficient*, Cambridge: Cambridge University Press, 6, 302.

Greene, W. H. (1981). "On the Asymptotic Bias of the Ordinary Least Squares Estimator," *Econometrica*, **49**, 505-13.

——— (1993). *Econometric Analysis*, New York: Macmillan Publishing Company.

Griffiths, W.E., R. Carter Hill, and G. G. Judge (1993). *Learning and Practicing Econometrics*, New York: John Wiley & Sons, Inc.

Gujarati, D. N. (1995). *Basic Econometrics*, New York: McGraw Hill Inc.

Ham, J. (1982). "Estimation of Labour Supply with Censoring Due to Unemployment and Underemployment," *Review of Economic Studies*, **49**, 335-54.

Harris, R. (1992). "Ethnicity, Gender and Labour Supply in New Zealand," *New Zealand Economic Papers*, **26**, 199-218.

Hausman, J. A. (1978) "Specification Tests in Econometrics," *Econometrica*, **46**, 1251-71.

——— (1979). "The Econometric of Labour Supply on Convex Budget Sets," *Economics Letters*, **3**, 171-74.

——— (1980). "The Effects of Wages, Taxes, and Fixed Costs on Women's Labour Force Participation," *Journal of Public Economics*, **14**, 161-94.

——— (1981a). "Labour Supply," in Aaron, H. J. and Pechman, J. A., eds, *How Taxes Affect Economic Behaviour*, Washington, D.C.: The Brookings Institution, 27-72.

——— (1981b). "Income and Payroll Tax Policy and Labour Supply", in Meyer, L. H., ed., *The Supply-Side Effects of Economic Policy*, St. Louis, Mo.: Center for the Study of American Business, Washington University, 173-202.

——— (1985). "Taxes and Labour Supply," in Auerbacj, A. and Feldstein, M., eds., *Handbook of Public Economics*, New York: Elseiver Science Publishers BV, 213-63.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error," *Econometrica*, **47**, 153-61.

——— (1993). "What Has Been Learned About Labor Supply in the Past 20 Years," *American Economic Review*, **83**:2, 116-21.

Heckman, J. J., and T. E. MaCurdy (1980). "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies*, **47**, 47-74.

Hendry, D. F., and R. W. Harrison (1974). "Monte Carlo Methodology And The Small Sample Behaviour Of Ordinary And Two-Stage Least Squares," *Journal of Econometrics*, **2**:2, 151-74.

Hiemstra, C., and Kelejian, H. H. (1991). "A Rare Events Model: Monte Carlo Results on Sample Design and Large Sample Guidance," *Economics Letters*, **37**:3, 255-63.

Hum, D. and W. Simpson (1989). *Income Transfers, Work Effort and the Canadian Experiment*, Economic Council of Ottawa.

Johnston, J. (1963). *Econometric Methods*, New York: McGraw Hill.

Kahn, S., and K. Lang (1988). "The Effects of Hours Constraints on Labour Supply Estimates," *NBER Working Paper*, 2647.

Kalos, M. H., and P.A. Whitlock (1986). *Monte Carlo Methods*, New York: John Wiley & Sons.

Killingsworth, M. (1983). *Labour Supply*, Cambridge: Cambridge University Press

Kleijnen, J. P. C. (1974). *Statistical Techniques in Simulation*, Part I New York: Marcel Dekker, Inc.

Kmenta, J. (1986). *Elements of Econometrics*, New York: Macmillan Publishing Company.

Kosters, M. (1967). "Effects of an Income Tax on Labor Supply" in A. Harberger and M. Baily, eds., *The Taxation of Income From Capital*, Washington, DC: Brookings Institution, 301-21.

Koutsoyiannis, A. (1977). *Theory of Econometrics*, 2nd ed., London: Macmillan Press.

Lachin, J. M. (1981). "Introduction to Sample Size Determination and Power Analysis for Clinical Trials." *Controlled Clinical Trials*, **2**, 93-113.

Lee, L. F., G. S. Maddala and R. P. Trost (1980). "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity." *Econometrica*, **48**, 491-503.

Leuthold, J. H. (1978). "The Effect of Taxation on the Hours Worked by Married Women," *Industrial and Labor Relations Review*, **31**, 520-26.

Leung, S. F., and S. Yu (1996). "On the Choice Between Sample Selection and Two-Part Models," *Journal of Econometrics*, **72**:1-2, 197-229.

Lo, A. W., and A. C. MacKinlay (1988). "The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation," *NBER Technical Papers*, 66.

Mace, A. E. (1974). *Sample-Size Determination*, New York: Robert E. Krieger Publishing Company.

MaCurdy, T., D. Greene, and H. Paarsch (1990). "Assessing Empirical Approaches for Analyzing Taxes an Labor Supply," *Journal of Human Resources*, **25**, 415-90.

Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

Manning, W. G., N. Duan, and W. H. Rogers (1987). "Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models," *Journal of Econometrics*, **35**:1, 59-82.

McCracken, D. D. (1955). "The Monte Carlo Method," *Scientific American*, **192**, 90-97.

Morgenthaler, G. W. (1961). "The Theory and Application of Simulation in Operations Research," in R. L. Ackoff, ed., *Progress in Operations Research*, New York: John Wiley & Sons.

Mroz, T. A. (1987) "The Sensitivity of Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, **55**, 765-99.

Mullen, J. D., M. K. Wohlgenant, and D. E. Farris (1988). "Input Substitution and the Distribution if Surplus Gains from Lower U.S. Beef Processing Costs," *American Journal of Agricultural Economics*, **70**, 245-54.

Nakamura, A., and M. Nakamura (1981). "A Comparison of Labor Force Behavior of Married Women in the United States and Canada, with Special Attention to the Impact of Income Taxes," *Econometrica*, **49**:2, 451-90.

——— (1983). "Part-time and Full-time Work Behaviour of Married Women: A Model with a Doubly Truncated Dependent Variable," *The Canadian Journal of Economics*, **16**, 229-57.

Nakamura, A, M. Nakamura, and D. Cullen (1979). "Job Opportunities, the Offered Wage, and the Labour Supply of Married Women," *American Economic Review*, **69**:5, 787-805.

Nawata, K. (1994). "Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator," *Economics Letters*, **45**:1, 33-40.

Nawata, K., and N. Nagase (1996). "Estimation of Sample Selection Bias Models," *Econometric Reviews*, **15**:4, 387-400.

Parker, E. (1994). "The Accuracy of Generalized Cost Function Estimation: A Monte Carlo Approach," *Southern Economic Journal*, **60**:4, 907-26.

Rosen, H. S. (1976). "Taxes in a Labour Supply Model with Joint Wage-Hours Determination," *Econometrica*, **44**:3 485-507.

Rubinstein, R. Y. (1981). *Simulation and The Monte Carlo Method*, New York: John Wiley & Sons.

Sawyer, R. (1982). "Sample Size and the Accuracy of Predictions Made From Multiple Regression Equation," *Journal of Educational Statistics*, **7**:2, 91-104.

Schork, M. A., and R. D. Remington (1967). "The Determination of Sample Size in Treatment-Control Comparisons for Chronic Disease Studies in which Dropout or Non-Adherence Is A Problem," *Journal of Chronic Diseases*, **20**, 233-39.

SHAZAM User's Reference Manual Version 7.0, McGraw-Hill, 1993.

Sheykhet, I. I., and B.Y. Simkin (1990). "Monte Carlo Method in The Theory of Solutions," *Computer Physics Reports*, **12**:3, 67-133.

Smith, V. K. (1971). "Economic Anonymity and Monte Carlo Studies," *Applied Economics*, **3**, 35-46.

Sowey, E. R. (1973). "A Classification Bibilography of Monte Carlo Studies in Econometrics," *Journal of Econometrics*, **1**, 377-95.

Wales, T. J., and A. D. Wodland (1979). "Labour Supply and Progressive Taxes," *Review of Economic Studies*, **46**, 83-95.

White, H. (1980). "A Heteroscadesticity -- Consistence Covariance Matrix Estimation and a Direct Test for Heteroscadesticity," *Econometrica*, **46**, 817-38.

IMAGE EVALUATION TEST TARGET (QA-3)

150mm

6"

APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989