# THE COST OF PARTIAL OBSERVABILITY
# IN THE BIVARIATE PROBIT MODEL

By

Chun-Lo Katy Ho

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1982

ABSTRACT

THE COST OF PARTIAL OBSERVABILITY
IN THE BIVARIATE PROBIT MODEL

By

Chun-Lo Katy Ho

Some recent studies have made use of the bivariate probit model in testing various hypotheses, but with only partial observability about the dichotomous dependent variables. The maximum likelihood estimators in these partial observability cases will be inefficient compared to those obtained under full observability. Therefore in this study, we present several cases with different levels of observability for the bivariate probit model and we measure the efficiency loss of maximum likelihood estimators for each case through some experiments.

The example of a two-member committee voting under a unanimity rule can be applied to all of these cases. Case one is the case of full observability in which the dichotomous choices of both voters are always observable. Case two is the case of partial observability in the sense of Poirier, which under the assumption that only the result of the joint choice of the two decision-makers is observed. Case three is called the case of partial partial observability, in which one of the two parties' decision is fully observable. In case four, which is called the case of partial observability with observed veto, when the outcome is "no", we observe one of the two parties casting its "no" vote. Three alternative possibilities are presented for this case, concerning who will use the veto first if both parties wish to vote "no".

The log-likelihood functions are provided for the joint estimation of the parameters for each of the various cases above. Since the inverse of information matrix is the asymptotic variance-covariance matrix of maximum likelihood estimator, the derivation of information matrices for all these cases are presented. The conditions for identification for the partial observability cases are also discussed. Then a large variety of experiments are done to measure the cost (in terms of lost efficiency) of partial observability.

Here are some of our main conclusions. First we notice that the cost of partial observability is quite high, especially for case two. The cost of partial observability decreases markedly if any piece of observability information can be found. The law of diminishing marginal utility of information usually holds: it is the first piece of observability information which is most important. The second conclusion is that specifying $\rho$ (the correlation coefficient of the two probit equations) a priori improves the efficiencies of the estimates of the other parameters a great deal. A third conclusion is that the sample split has a strong influence on the relative efficiencies of the parameter estimates. For a given partial observability case, its efficiency relative to full observability will be higher, the smaller the proportion of observations which fall into the indistinguishable categories. The last conclusion is that the strength of identification matters. The relative efficiency of each partial observability case is very low for parameter values near such points, and it increases rapidly as the parameters move away from such points of singularity.

## ACKNOWLEDGEMENT

I must first express my sincere gratitude to my thesis advisor, Professor Peter Schmidt, who generously bestowed an enormous amount of his time providing detailed guidance on every aspect of this study. Without his continuing assistance and encouragement, I could not finish this thesis and my graduate study at Michigan State University.

I am also very grateful to the other members of my dissertation committee. They are Professors John Goddeeris, Daniel Hamermesh and William Quinn.

I owe thanks to my typist, Miss Kelli Sweet. Her careful and expedient typing is appreciated.

Finally, my deepest appreciation goes to my parents and my husband, Hai-Zui, for their support and encouragement throughout these years. To them, I owe a debt more than I know how to express in words.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

| Table | | Page |
|---|---|---|

# CHAPTER ONE

## INTRODUCTION

The purpose of this study is to consider the bivariate probit model under various levels of observability of the dependent variables, and to measure the loss in efficiency caused by less than full observability.

There have been quite a few studies using the bivariate probit model in a variety of settings. Zellner and Lee (1965) presented the probit model as well as other models to analyze discrete random variables. They showed that a joint estimation approach for a set of equations with dichotomous endogenous variables yields estimators which are asymptotically more efficient than single equation techniques, provided that the variables being analyzed are correlated. They considered the example of a durable good purchase decision (buy or not buy) and a credit decision (use installment credit or not-use such credit), while the exogenous variable is disposable income. In this example, both decisions are observable.

Ashford and Sowden (1970) considered a multivariate probit model and proposed maximum likelihood estimation for its parameters. They applied their techniques to a bivariate probit model, where the two endogenous variables are breathlessness and wheeze of a coal miner, and the exogenous variable is his age. A coal miner may have positive response to neither, to one or the other, or to both of the two symptoms; so there are four possible outcomes. All four possible outcomes are distinguishable; the data gives the number of individuals with each combination of symptoms, within each age group in the sample. Amemiya (1974) proposed two minimum chi-squared estimators for the same model and found that the FIMC (Full Information Minimum Chi-Square) Probit estimator is asymptotically as efficient as the maximum likelihood estimator.

Gunderson (1974) discussed alternative statistical models for estimating the probability that an on-the-job trainee will be retained by the sponsoring company after training. In this situation, the employer must decide whether or not to make a job offer, and the trainee must decide whether or not to seek a job offer. Each individual's (either employer's or trainee's) decision is not observed; only whether or not the trainee continues working after training is known. Gunderson used a single-equation model with the dichotomous dependent variable coded 1 if the trainee stays with the company, and 0 otherwise. Explanatory variables includes the characteristics of both the trainee (age, sex, education, experience, etc.) and the company (company size, area designation, etc.). Poirier (1980) proposed a bivariate probit model under the same assumptions as Gunderson's concerning the amount of available information. His model includes two probit equations each representing the binary choice of a decision-maker, but only the outcome of the joint (unanimous) choice is observable. That is, the only information about the two dichotomies is whether or not both equal unity, and the remaining possible outcomes can't be distinguished from each other whenever there is a negative choice made by either party.

Farber's research (1982) on the demand for unionism shows that the union status of workers is determined by a combination of workers' demand for union representation and the decisions of union employers as to whom to hire. That is, a worker is a union member if and only if he desires a union job and a union employer is willing to hire him. If only the final outcome (union status) is observed, it is impossible to determine whether nonunion workers didn't want a union job, couldn't get a union job, or both, and we have

Poirier's model. In Farber's study, a unique data is employed which can be used to identify the union or non-union preference of non-union workers. So workers' preferences are fully observable, while union employers' decisions are still unknown for those nonunion workers who didn't want to be unionized.

Connolly's study (1982) analyzes the joint decision to arbitrate or negotiate the contracts between employees' unions and municipalities in Michigan. According to law, there will be negotiation if both sides desire so and there will be arbitration otherwise, but one of the two parties has to cast a veto to seek the arbitration. Therefore, besides the observable result that the contract is negotiated or arbitrated, one party (only) is observed to use the veto whenever there is arbitration. However, the decision of the party which didn't use the veto remains unknown.

The examples above can all be analyzed using the bivariate probit model, but under different assumptions concerning what can be observed. The first two cases (Zellner-Lee and Ashford-Sowden) are different from the others in that the two decisions (or symptoms) are all related to one person instead of to two different parties. But they still represent the case in which the two binary dependent variables are both observable, which can be called the case of full observability. All the other cases have less than full observability, in varying degrees. The model established by Poirier (using Gunderson's example) assumes the least observability information among all the cases. As outsiders, we can only tell whether something failed or succeeded. For Farber's or Connally's case, besides the observable joint choice, one of the two individual choices is observed. All of these cases can be called partial observability cases of the bivariate probit model.

With incomplete information, the maximum likelihood estimators obtained in these partial observability cases will be inefficient compared to the estimators obtained in the case of full observability. In other words, there is a cost (in terms of lost efficiency) of partial observability. The point of this research is to measure this cost. The study of the cost of partial observability is important in itself, but it also has some practical implications. For a researcher facing a high price of getting additional information, it is important to know how valuable the information is, so that an intelligent decision can be made about whether additional information is worth obtaining.

This paper is divided into five chapters. In Chapter Two, a formal statement of bivariate probit model is presented. All of the cases considered assume this basic model, but with different levels of observability. Case one is the full observability case, in which both parties' choices are observable, and every possible outcomes can therefore be distinguished. Case two is the model with partial observability in the sense of Poirier, in which the only information is the binary outcome of the joint choice made by both parties. If either one party fails to say "yes", then the remaining outcomes are indistinguishable. Case three is called the partial partial observability case. In this case one of the two parties' decision is fully observable, but if the observable party has a negative response, the other party's decision is not known. Case four is called the partial observability case with observed veto. In this case if both sides do not say "yes", then we can observe one and only one party saying "no" (casting the veto). There are three alternative possibilities which we will consider, under different assumptions about who will cast the veto first if both parties wish to say "no". The appropriate likelihood functions for all cases and possibilities are provided, so that maximum likelihood estimates can be obtained.

Chapter Three contains the derivation of information matrices for all of the cases which were presented in Chapter Two. The conditions for identification for the partial observability cases are also discussed. The information matrix (whose inverse is the asymptotic variance-covariance matrix of maximum likelihood estimator) can help us to measure the efficiency loss with different levels of partial observability. The conditions of identification are relevant for efficiency comparisons, because the closer the information matrix is to being singular, the greater the variances of the estimates will be.

In Chapter Four, a large variety of experiments are done to measure the cost of various levels of partial observability. For the purpose of simplification, we assume there are only two exogenous variables, and one of them is a constant term. For each experiment, specific values of the parameters are picked in order to evaluate the inverse of the information matrix. All of the elements of the inverse of the information matrices for the partial observability cases are divided by the corresponding elements of the inverse of the information matrix for the full observability case. Thus the ratios we present are the ratios of asymptotic variances and covariances of parameters in the partial observability case compared to those of the full observability case. The cost of partial observability increases as these ratios increase. We attempt both to make a rough statement about the cost of partial observability in typical cases, and also to identify what types of changes in the parameters cause this cost to increase or decrease.

The results of these experiments will be interpreted in Chapter Four, and the tables at the end list the main numerical results. The summary and conclusions of this study will be given in Chapter Five.

# CHAPTER TWO

# BIVARIATE PROBIT MODELS WITH FULL AND PARTIAL OBSERVABILITY

## 2.1 Introduction

In this chapter, we will give a formal statement of the bivariate probit model, and consider its estimation under various assumptions about what is observed. Basically, our treatment of the estimation problem is just to provide the appropriate likelihood function to maximize, though in some cases we point out alternative possibilities. The questions of identification and of the relative efficiencies of the various estimators will be deferred until Chapters 3 and 4.

Now we start by reviewing the bivariate probit model. Consider two individuals (j=1,2) each faced with a binary choice, $y_j$=m, m=0,1. The dependent variable $y_j$ takes on the value 1 if an event occurs or 0 if it does not occur. Suppose the two individuals have utility functions of the form

$$U_{1m} = g_{1m}\,(w_{1m},\, y_2^*) + \eta_{1m} \qquad m=0,1$$
$$U_{2m} = g_{2m}\,(w_{2m},\, y_1^*) + \eta_{2m}$$

where for j=1,2, $g_{jm}$ is a non-stochastic scale function, $w_{jm}$ is a fixed vector of characteristics of individual j and choice $y_j$=m, $\eta_{jm}$ is a random disturbance term and $y_j^*$ is the utility differential

$$y_j^* = U_{j1} - U_{j0} \qquad j=1,2.$$

This specification permits interdependency between the utility functions of the two individuals in the sense that the utility of each individual is a function of the sentiment of the other individual.

Further suppose

$$g_{11}\ (w_{11},\ y_2^*) - g_{10}\ (w_{10},\ y_2^*) = \gamma_1 y_2^* + X\delta_1$$

$$g_{21}\ (w_{21},\ y_1^*) - g_{20}\ (w_{20},\ y_1^*) = \gamma_2 y_1^* + X\delta_2$$

$$\eta_{11} - \eta_{10} = V_1$$

$$\eta_{21} - \eta_{20} = V_2$$

where X is a K-dimensional row vector of explanatory variables, $\delta_1$ and $\delta_2$ are K-dimensional column vectors of unknown coefficients, $\gamma_1$ and $\gamma_2$ are unknown parameters, and

$$V \equiv [V_1,\ V_2]' \sim N(0,\ \Omega) \text{ with}$$

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix} .$$

Then it is easy to show that

$$y_1^* = \gamma_1 y_2^* + X\delta_1 + V_1 \text{ ---- } (1)$$

$$y_2^* = \gamma_2 y_1^* + X\delta_2 + V_2 \text{ ---- } (2)$$

and that individual j will select

$$y_j = 1 \quad \text{iff} \quad y_j^* > 0 \quad \text{i.e., } U_{j1} > U_{j0}$$

$$y_j = 0 \quad \text{iff} \quad y_j^* \leq 0 \quad \text{i.e., } U_{j1} \leq U_{j0} .$$

The reduced form equations corresponding to (1) and (2) are

$$y_1^* = X\beta_1 + \varepsilon_1 \text{ ------------- } (3)$$

$$y_2^* = X\beta_2 + \varepsilon_2 \text{ ------------- } (4)$$

where

$$\beta_1 = (\delta_1 - \gamma_1\delta_2)/(1 - \gamma_1\gamma_2)$$

$$\beta_2 = (\delta_2 - \gamma_2\delta_1)/(1 - \gamma_1\gamma_2)$$

$$\varepsilon_1 = (V_1 + \gamma_1 V_2)/(1 - \gamma_1\gamma_2)$$

$$\varepsilon_2 = (V_2 + \gamma_2 V_1)/(1 - \gamma_1\gamma_2)$$

and $\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix}$ has bivariate normal distribution with 0 mean and variance-covariance matrix as $\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$. Here the variances of $\varepsilon_1$ and $\varepsilon_2$ have been normalized to equal unity and $\rho$ is the correlation between $\varepsilon_1$ and $\varepsilon_2$.

The model just presented is common to all of the cases we will consider. However, the cases differ with respect to how much one observes about $y_1$ and $y_2$.

## 2.2 Case One: Full Observability

Here we assume that $y_1$ and $y_2$ are both observed. Among all the cases we are going to discuss here, this is the one which has the most complete observability, and which leads to the most efficient estimates. An example of such a case would be a two member committee voting under a unanimity rule, but with both votes observable. That is, in our random sample of votes, i=1, ---, N, we can not only observe the explanatory variables $X_i$, but also the votes of both voters, i.e., $y_{i1}$ and $y_{i2}$. Therefore, there are four possible outcomes which are all distinguishable:

(1) both vote "yes", i.e., $y_{i1}=1$ and $y_{i2}=1$;

(2) the first party votes "yes" and the second votes "no", i.e., $y_{i1}=1$ and $y_{i2}=0$;

(3) the first party votes "no" and the second votes "yes", i.e., $y_{i1}=0$ and $y_{i2}=1$;

(4) both vote "no", i.e., $y_{i1}=0$ and $y_{i2}=0$.

The distribution of $y_{i1}$ and $y_{i2}$ in this case is

$$P(y_{i1}=1 \text{ and } y_{i2}=1) = F(X_i\beta_1,\ X_i\beta_2;\ \rho) \qquad i=1,---,N$$

$$P(y_{i1}=1 \text{ and } y_{i2}=0) = F(X_i\beta_1,\ -X_i\beta_2;\ -\rho)$$

$$= \Phi\ (X_i\beta_1) - F(X_i\beta_1,\ X_i\beta_2;\ \rho)$$

$$P(y_{i1}=0 \text{ and } y_{i2}=1) = F(-X_i\beta_1,\ X_i\beta_2;\ -\rho)$$

$$= \Phi\ (X_i\beta_2) - F(X_i\beta_1,\ X_i\beta_2;\ \rho)$$

$$P(y_{i1}=0 \text{ and } y_{i2}=0) = F(-X_i\beta_1,\ -X_i\beta_2;\ \rho)$$

$$= 1 - \Phi(X_i\beta_1) - \Phi(X_i\beta_2) + F(X_i\beta_1,\ X_i\beta_2;\ \rho)$$

where $F(\cdot,\cdot;\ \cdot)$ denotes the bivariate standard normal distribution function with correlation coefficient $\rho$, while $\Phi(\cdot)$ is the univariate standard normal distribution function.

We can always estimate the reduced form equations separately. The log-likelihood functions are

$$\ell n\ L = \sum_{i}^{N} \{y_{ij}\ \ell n\ \Phi(X_i\beta_j) + (1-y_{ij})\ \ell n\ \Phi(-X_i\beta_j)\}$$

$$i = 1,\ ---,\ N$$

$$j = 1 \text{ or } 2\ .$$

But this is efficient only when $\rho=0$. When $\rho$ is not equal to zero, it is more efficient to estimate the two probits jointly. Then the log-likelihood function of the sample is

$$\begin{aligned} \ln L(\beta_1, \beta_2, \rho) = \sum_i^N \{ & y_{i1} y_{i2} \ln F(X_i\beta_1, X_i\beta_2; \rho) \\ & + y_{i1}(1 - y_{i2}) \ln [\Phi(X_i\beta_1) - F(X_i\beta_1, X_i\beta_2; \rho)] \\ & + (1 - y_{i1})\, y_{i2} \ln [\Phi(X_i\beta_2) - F(X_i\beta_1, X_i\beta_2; \rho)] \\ & + (1 - y_{i1})(1 - y_{i2}) \ln [1 - \Phi(X_i\beta_1) - \Phi(X_i\beta_2) \\ & + F(X_i\beta_1, X_i\beta_2; \rho)]\} . \end{aligned}$$

2.3 Case Two: Partial Observability in the Sense of Poirier

This is the case treated by Poirier (1980). An example of this would be a two member committee voting under a unanimity rule, but anonymously. As outsiders we can only observe whether a motion passes (i.e. both members vote "yes") or whether it fails (i.e. at least one member votes "no"). So instead of observing $X_i$, $y_{i1}$ and $y_{i2}$, $i=1,2,\ldots,N$, we observe only $X_i$ and $Z_i$, where $Z_i = y_{i1} \cdot y_{i2}$, $i=1,2,\ldots,N$. That is,

$$\begin{aligned} Z_i &= 1 \quad \text{iff } y_{i1}=1 \text{ and } y_{i2}=1 \\ &= 0 \quad \text{otherwise} . \end{aligned}$$

In terms of the four possible outcomes that we listed in section 2.2, the last three are indistinguishable because all we can see for these outcomes is just that the motion does not pass.

This model has been used by Connally (1982) to study the decision to arbitrate or negotiate the contracts between public employees' unions and municipalities in Michigan. By law, binding arbitration occurs if either party so desires. Therefore, a contract is negotiated ($Z_i=1$) if and only if the union desires negotiation ($y_{i1}=1$) and the municipality desires negotiation ($y_{i2}=1$). Otherwise the contract is arbitrated ($Z_i=0$). Poirier's model applies if we observe only the outcome (negotiation or arbitration) and not the desire of either party separately.

The distribution of $Z_i$ is

$$P(Z_i=1) = P(y_{i1}=1 \text{ and } y_{i2}=1)$$
$$= F(X_i\beta_1, X_i\beta_2; \rho)$$
$$P(Z_i=0) = P(y_{i1}=0 \text{ or } y_{i2}=0)$$
$$= 1 - F(X_i\beta_1, X_i\beta_2; \rho)$$

and the log-likelihood function of the sample is

$$\ell n\, L(\beta_1, \beta_2, \rho)$$
$$= \sum_i^N \{Z_i\ \ell n\ F\ (X_i\beta_1, X_i\beta_2; \rho) + (1-Z_i)\ \ell n[1-F(X_i\beta_1, X_i\beta_2; \rho)]\}$$
$$i=1,2,\ldots,N.$$

## 2.4 Case Three: Partial Partial Observability

In this model, we observe more than in Poirier's model, but less than in the full observability case. Specifically, it is assumed that we observe $y_{i1}$, $Z_i = y_{i1} \cdot y_{i2}$, and $X_i$, i=1,2,...,N. Note that the vote of the first voter ($y_{i1}$) is always observed. However, the vote of the second voter is only sometimes observed. Essentially, we observe $y_{i2}$ if and only if $y_{i1}=1$. This is so because if $y_{i1}=1$, then $y_{i2}=Z_i$, and $Z_i$ is observed. However, if $y_{i1}=0$ we have no information about $y_{i2}$. Thus in terms of the four possible outcomes, two ($y_{i1}=0$, $y_{i2}=1$ and $y_{i1}=0$, $y_{i2}=0$) are indistinguishable. This model can also be regarded as a censored probit model, since the sample for which $y_{i2}$ is observed is censored, by the value of $y_{i1}$.

This model has been used by Farber (1982) to study the demand for unionism. Let $y_{i1}$=1 if individual i wishes to be in a union, and $y_{i1}$=0 otherwise; let $y_{i2}$=1 if a union employer is willing to hire individual i, and $y_{i2}$=0 otherwise. Individual i is a union member ($Z_i=y_{i1}\cdot y_{i2}=1$) if both $y_{i1}$=1 and $y_{i2}$=1, and is not a union member ($Z_i$=0) otherwise; $Z_i$ is observed for all i. If nothing more were known, this model would be Poirier's model. However, non-union workers in Farber's sample were asked if they desired union representation, so that $y_{i1}$ is also observed for all i. On the other hand, $y_{i2}$ is observed only if $y_{i1}$=1.

Now give that we observe $y_{i1}$, $Z_i=y_{i1}\cdot y_{i2}$ and $X_i$, the log-likelihood function for the model is

$$\begin{aligned}
&\ln L(\beta_1, \beta_2, \rho) \\
&= \sum_i^N \{y_{i1}\, y_{i2} \ln F(X_i\beta_1, X_i\beta_2; \rho) \\
&\quad + y_{i1}(1 - y_{i2}) \ln[\, \Phi(X_i\beta_1) - F(X_i\beta_1, X_2\beta_2; \rho)] \\
&\quad + (1 - y_{i1}) \ln \Phi(-X_i\beta_1)\} .
\end{aligned}$$

Because it is observable, the first probit equation can always be estimated separately with the log-likelihood function as

$$\ln L(\beta_1) = \sum_i \{y_{i1} \ln \Phi(X_i\beta_1) + (1 - y_{i1}) \ln \Phi(-X_i\beta_1)\}$$

but joint estimation is more efficient for the first equation unless $\rho$=0. Separate estimation for the second equation is possible if and only if $\rho$=0. When $\rho$=0, using the observation of the first party, we can establish the log-likelihood function of the second probit equation as

$$\ln L(\beta_2) = \sum_i^N \{y_{i1} \ln \Phi(X_i\beta_1) + (1 - y_{i1}) \ln \Phi(-X_i\beta_1)\}$$
$$+ \sum_i^N y_{i1} \{y_{i2} \ln \Phi(X_i\beta_2) + (1 - y_{i2}) \ln \Phi(-X_i\beta_2)\} .$$

## 2.5 Case Four: Partial Observability With Observed Veto

In this case, it is assumed that we observe $Z_i=y_{i1} \cdot y_{i2}$ and $X_i$, as in Poirier's case. However, when $Z_i=0$ we observe either $y_{i1}=0$ or $y_{i2}=0$ (but never both). That is, we observe the casting of a veto by one party.

The situation analyzed by Connally (1982) again provides a good example. Negotiation occurs if and only if both the municipality and the union wish to negotiate. Otherwise arbitration occurs. However, one party or the other has to ask for arbitration, so that the veto (of negotiation) by one party is observed. That is, there is negotiation ($Z_i=1$) iff $y_{i1}=1$ and $y_{i2}=1$, and there is arbitration ($Z_i=0$) iff $y_{i1}=0$ or $y_{i2}=0$. Furthermore, when $Z_i=0$ is observed we observe either $y_{i1}$ or $y_{i2}$ as zero.

Among the four possible outcomes, $Z_i=1$ is straightforward. In the case when $Z_i=0$ and $y_{i1}=0$ are observed, however, we don't know if party two votes "yes" or "no". The same is so for the $Z_i=0$ and $y_{i2}=0$ situation. In such a case, the distribution of $Z_i$ is

$$P(Z_i=1) = P(y_{i1}=1 \text{ and } y_{i2}=1) = F(X_i\beta_1, X_i\beta_2; \rho)$$

$$P(Z_i=0 \text{ and } y_{i1}=0) = P(y_{i1}=0 \text{ and } y_{i2}=1) + P(y_{i1}=0 \text{ and } y_{i2}=0)$$
$$\cdot P(\text{1st party requests arbitration}/y_{i1}=0 \text{ and } y_{i2}=0)$$

$$P(Z_i=0 \text{ and } y_{i2}=0) = P(y_{i1}=1 \text{ and } y_{i2}=0) + P(y_{i1}=0 \text{ and } y_{i2}=0)$$
$$\cdot P(\text{2nd party requests arbitration}/y_{i1}=0, y_{i2}=0) .$$

Note that the events ($y_{i1}=0$, $y_{i2}=1$) and ($y_{i1}=0$, $y_{i2}=0$, party one requests arbitration first) are indistinguishable.

We can always ignore the information about the observed veto and convert this back to a standard Poirier partial observability case (Case Two). The cost for doing that is some loss of efficiency of the estimates. Otherwise, to close the model some assumption must be made about who would ask for arbitration first, if both parties desire it. Here we have three possibilities.

Possibility One: Case 4A1, Observed Veto, Given Probability = p

Assume there is some fixed probability for the first party to request arbitration first given that both sides don't want a negotiation. That is,

$$p = P(\text{1st party requests arbitration first}/0,\ 0)$$
$$1-p = P(\text{2nd party requests arbitration first}/0,\ 0)\ .$$

Here p is a given constant. It might come from past experience or just represent a reasonable guess. The log-likelihood function is

$$\ell n\ L(\beta_1, \beta_2, \rho)$$
$$= \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ell n\ F(X_i\beta_1,\ X_i\beta_2;\ \rho)$$
$$+ \sum_{\substack{i \ni y_{i1} = 0 \\ \text{is observed}}} \ell n\ \{F(-X_i\beta_1,\ X_i\beta_2;\ -\rho) + pF(-X_i\beta_1,\ -X_i\beta_2;\rho)\}$$
$$+ \sum_{\substack{i \ni y_{i2} = 0 \\ \text{is observed}}} \ell n\ \{F(X_i\beta_1,\ -X_i\beta_2;\ -\rho) + (1-p)F(-X_i\beta_1,\ -X_i\beta_2;\ \rho)\}$$

$$= \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ln F(X_i\beta_1, X_i\beta_2; \rho)$$

$$+ \sum_{\substack{i \ni y_{i1} = 0 \\ \text{is observed}}} \ln[\{ \Phi(X_i\beta_2) - F(X_i\beta_1, X_i\beta_2; \rho)] + p[1 - \Phi(X_i\beta_1) - \Phi(X_i\beta_2) + F(X_i\beta_1, X_i\beta_2; \rho)]\}$$

$$+ \sum_{\substack{i \ni y_{i2} = 0 \\ \text{is observed}}} \ln \{[\Phi(X_i\beta_1) - F(X_i\beta_1, X_i\beta_2; \rho)] + (1-p)\, [1 - \Phi(X_i\beta_1) - \Phi(X_i\beta_2) + F(X_i\beta_1, X_i\beta_2; \rho)]\}$$

$$= \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ln F(X_i\beta_1, X_i\beta_2; \rho)$$

$$+ \sum_{\substack{i \ni y_{i1} = 0 \\ \text{is observed}}} \ln \{p[1 - \Phi(X_i\beta_1)] + (1-p)\, [\Phi(X_i\beta_2) - F(X_i\beta_1, X_i\beta_2; \rho)]\}$$

$$+ \sum_{\substack{i \ni y_{i2} = 0 \\ \text{is observed}}} \ln \{p\, [\Phi(X_i\beta_1) - F(X_i\beta_1, X_i\beta_2; \rho)] + (1-p)\, [1 - \Phi(X_i\beta_2)]\}.$$

Possibility Two: Case 4A2, Observed Veto, p is Another Parameter

Here instead of having p as a given constant, we let p be an unknown parameter in the model. The log-likelihood function is the same as above except p needs to be estimated too.

$$\ln L(\beta_1, \beta_2, \rho, p)$$

$$= \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ln F(X_i\beta_1, X_i\beta_2; \rho)$$

$$+ \sum_{\substack{i \ni y_{i1}=0 \\ \text{is observed}}} \ell n \{p[1- \Phi(X_i\beta_1)] + (1-p) [\Phi(X_i\beta_2) - F(X_i\beta_1, X_i\beta_2; \rho)]\}$$

$$+ \sum_{\substack{i \ni y_{i2}=0 \\ \text{is observed}}} \ell n \{p[\Phi(X_i\beta_1) - F(X_i\beta_1, X_i\beta_2; \rho)] + (1-p) [1- \Phi(X_i\beta_2)]\} \ .$$

Possibility Three: Case 4B, The First Party to Ask for Arbitration is the One Who Wants Negotiation Least

Recall equations (3) and (4)

$$y_1^* = X\beta_1 + \varepsilon_1$$

$$y_2^* = X\beta_2 + \varepsilon_2$$

$y_1^*$ and $y_2^*$ are the utility differentials between voting "yes" and "no". They represent individual j's (j=1,2) "sentiment" toward $y_j=1$. When $y_{i1}^* < 0$ and $y_{i2}^* < 0$, so that neither party wants negotiation, it may be that the party whose sentiment is more strongly against negotiation will cast the veto first. That is, when

$$y_{1i}^* = X_i\beta_1 + \varepsilon_{i1} < 0$$

$$y_{2i}^* = X_i\beta_2 + \varepsilon_{i2} < 0$$

$$\text{and } X_i\beta_1 + \varepsilon_{i1} < X_i\beta_2 + \varepsilon_{i2} \ ,$$

it is reasonable to conclude that the first party will cast the veto first. So we use $P(\varepsilon_{i2} < - X_i\beta_2, \varepsilon_{i1} - \varepsilon_{i2} < X_i\beta_2 - X_i\beta_1)$ to represent the probability that the first party is observed using the veto and $P(\varepsilon_{i1} < - X_i\beta_1, \varepsilon_{i2} - \varepsilon_{i1} < X_i\beta_1 - X_i\beta_2)$ to represent the probability that the second party uses the veto, given that both of them don't want a negotiation. The log-likelihood function is

$$\ell n\ L\ (\beta_1,\ \beta_2,\ \rho)$$

$$= \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ell n\ F(X_i\beta_1,\ X_i\beta_2;\ \rho)$$

$$+ \sum_{\substack{i \ni y_{i1} = 0 \\ \text{is observed}}} \ell n\ [F(-X_i\beta_1,\ X_i\beta_2;\ -\rho) + P(\varepsilon_{i2} < -X_i\beta_2,\ \varepsilon_{i1} - \varepsilon_{i2} < X_i\beta_2 - X_i\beta_1)]$$

$$+ \sum_{\substack{i \ni y_{i2} = 0 \\ \text{is observed}}} \ell n\ [F(X_i\beta_1,\ -X_i\beta_2;\ -\rho) + P(\varepsilon_{i1} < -X_i\beta_1,\ \varepsilon_{i2} - \varepsilon_{i1} < X_i\beta_1 - X_i\beta_2)]\ .$$

Here $P(\varepsilon_{i2} < -X_i\beta_2,\ \varepsilon_{i1} - \varepsilon_{i2} < X_i\beta_2 - X_i\beta_1)$

$$= P\left(\frac{\varepsilon_{i1} - \varepsilon_{i2}}{\sqrt{2(1-\rho)}} < \frac{X_i\beta_2 - X_i\beta_1}{\sqrt{2(1-\rho)}}\ ,\ \varepsilon_{i2} < -\ X_i\beta_2\right)$$

$$= F\left(\frac{X_i\beta_2 - X_i\beta_1}{\sqrt{2(1-\rho)}}\ ,\ -\ X_i\beta_2;\ -\sqrt{(1-\rho)/2}\right)$$

$$= F\left(\frac{X_i\beta_2 - X_i\beta_1}{-2\rho'}\ ,\ -\ X_i\beta_2;\ \rho'\right)\ ,$$

where $\rho' = -\sqrt{(1-\rho)/2}$

$$P(\varepsilon_{i1} < -X_i\beta_1,\ \varepsilon_{i2} - \varepsilon_{i1} < X_i\beta_1 - X_i\beta_2)$$

$$= F(-X_i\beta_1, -X_i\beta_2; \rho) - F(\frac{X_i\beta_2 - X_i\beta_1}{-2\rho'}, -X_i\beta_2; \rho')$$

$$= 1 - \Phi(X_i\beta_1) - \Phi(X_i\beta_2) + F(X_i\beta_1, X_i\beta_2; \rho)$$

$$- F(\frac{X_i\beta_2 - X_i\beta_1}{-2\rho'}, -X_i\beta_2; \rho') ,$$

then

$$\ell n\ L(\beta_1, \beta_2, \rho) = \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ell n\ F(X_i\beta_1, X_i\beta_2; \rho)$$

$$+ \sum_{\substack{i \ni y_{i1} = 0 \\ \text{is observed}}} \ell n\ [\Phi(X_i\beta_2) - F(X_i\beta_1, X_i\beta_2; \rho) + F(\frac{X_i\beta_2 - X_i\beta_1}{-2\rho'}, -X_i\beta_2; \rho')]$$

$$+ \sum_{\substack{i \ni y_{i2} = 0 \\ \text{is observed}}} \ell n\ [1 - \Phi(X_i\beta_2) - F(\frac{X_i\beta_2 - X_i\beta_1}{-2\rho'}, -X_i\beta_2; \rho')]$$

## 2.6 Summary

In this chapter, six cases are introduced to represent full observability and different types of partial observability for a bivariate probit model. The example of a two member committee voting under a unanimity rule can be applied to all cases. Case One gives full observability about the model, since the dichotomous choices of both voters are always observable. Separate estimation is possible for each probit equation but this would not be efficient unless the correlation coefficient $\rho=0$.

In Case Two, partial observability in the sense of Poirier, only the result of the joint choice of two decision-makers is observed. As long as either party votes "no", the separate votes of the two voters are indistinguishable. This is the case which gives us the least information.

Case Three and Case Four each lie somewhere between the above two cases. One of the two voters' behavior is observed in Case Three. But when this observable party votes "no", the two choices of the other voter are indistinguishable. The observable party's probit equation can always be separately estimated, but only when $\rho=0$ will this be as efficient as joint estimation. Separate estimation for the other party's probit equation is impossible unless $\rho$ is known to be equal to zero.

In Case Four, when either party (or both) votes "no", we observe the casting of a "no" vote. But while the one party is observed casting the veto, the vote of the other party remains unknown. Some assumptions must be made here about who will use the veto first. We can either assume some fixed p to be the probability that the first party does so (Case 4A1), or have p as another unknown parameter in the model (Case 4A2). Another possibility, which is Case 4B, is that the party with the strongest sentiment for a "no" vote will be observed casting the veto.

We have provided likelihood functions for the various cases. In each case, a numerical maximization of the likelihood function provides maximum likelihood estimates. In Chapter 3 we will consider the asymptotic distributions of these estimates, and in Chapter 4 we will compare their relative efficiencies.

# CHAPTER THREE

# DERIVATION OF INFORMATION MATRICES AND CONDITIONS FOR IDENTIFICATION

## 3.1 Introduction

The log-likelihood functions that we presented in the last chapter for full and partial observability cases have prepared us for the derivation of the information matrices here. The information matrix by definition is equal to minus the matrix of the expectation of the second-order derivatives of the log-likelihood function with respect to the parameters. That is, $\mathcal{J} = -E(\frac{\partial^2 \log L}{\partial\Theta\partial\Theta'})$, where $\Theta$ is the vector of unknown parameters. Under certain regularity conditions, it can be shown that the maximum likelihood estimates are consistent and asymptotically normal, with a variance-covariance matrix which is equal to the inverse of information matrix. Therefore, through the information matrices which we derive in this chapter, we can compare the variances and covariances of the parameter estimates in different cases. That is, we can measure the efficiency lost for lack of full observability, which can be called the cost of partial observability.

Note that

$$[-E(\partial^2 \log L/\partial\Theta\partial\Theta')]$$
$$= E\,[(\frac{\partial \log L}{\partial\Theta})(\frac{\partial \log L}{\partial\Theta})']$$

The latter formula will be used for all cases in this chapter.

We also consider the problem of identification of the parameters of the model, under various levels of observability. Since (given certain regularity conditions) a necessary and sufficient condition for local identification is non-singularity of the information matrix, we examine the rank of the information matrices which we present. For all levels of observability which we consider, the parameters are identified except in certain special (perverse) cases, which we point out.

The question of identification is important in its own right, but we are also interested in it because it is relevant for efficiency comparisons. The closer the information matrix is to being singular, the larger are the variances of the estimates. In the next chapter, we will compare efficiencies by evaluating the inverse of the information matrices, for various specific parameter values. Knowledge of the perverse cases which lead to non-identification will help us in picking regions of the parameters space to investigate.

Section 3.2-3.5 contain derivations of the information matrices and discussions of identification for the different cases listed in the last chapter. Section 3.6 gives a summary of this chapter.

## 3.2 Case One: Full Observability

In order to simplify the notation, we let

$$F_i = F(X_i\beta_1,\ X_i\beta_2;\ \rho) \qquad i=1,2,\ldots,N$$

$$a_i = X_i\beta_1$$

$$b_i = X_i\beta_2$$

$$\Theta = (\beta_1',\ \beta_2',\ \rho)' \ .$$

Then the log-likelihood function for the full observability case is

$$\ln L(\Theta) = \sum_i^N \{y_{i1}\ y_{i2}\ \ln F_i + y_{i1}(1-y_{i2})\ \ln[\ \Phi(a_i) - F_i]$$

$$+ (1-y_{i1})y_{i2} \ln [\Phi(b_i) - F_i]$$

$$+ (1-y_{i1})(1-y_{i2}) \ln [1- \Phi(a_i) - \Phi(b_i) + F_i]\},$$

the information matrix of this case is

$$\mathcal{I} = C_1'C_1 + C_2'C_2 + C_3'C_3 + C_4'C_4 \text{ ----(1)}$$

where $C_1'$, $C_2'$, $C_3'$ and $C_4'$ are all $(2K+1)\cdot N$ matrices and:

the ith column of $C_1'$ is $\dfrac{1}{\sqrt{F_i}} \dfrac{\partial F_i}{\partial \Theta}$

" " $C_2'$ is $\dfrac{1}{\sqrt{\Phi(a_i)-F_i}} \dfrac{\partial[\Phi(a_i)-F_i]}{\partial \Theta}$

" " $C_3'$ is $\dfrac{1}{\sqrt{\Phi(b_i)-F_i}} \dfrac{\partial[\Phi(b_i)-F_i]}{\partial \Theta}$

" " $C_4'$ is $\dfrac{1}{\sqrt{1-\Phi(a_i)-\Phi(b_i)+F_i}} \dfrac{\partial[1-\Phi(a_i)-\Phi(b_i)+F_i]}{\partial \Theta}$ .

Some of the derivatives which appear in the information matrix above have been derived by Ashford and Sowden (1970):

$$\frac{\partial F_i}{\partial \beta_1} = \phi(a_i)\Phi(A_i)X_i'$$

$$\frac{\partial F_i}{\partial \beta_2} = \phi(b_i)\Phi(B_i)X_i'$$

$$\frac{\partial F_i}{\partial \rho} = f_i \quad ,$$

and we know that

$$\frac{\partial\Phi(a_i)}{\partial\beta_1} = \phi\,(a_i)\; X_i' \qquad \frac{\partial\Phi(a_i)}{\partial\beta_2} = 0 \qquad \frac{\partial\Phi(a_i)}{\partial\rho} = 0$$

$$\frac{\partial\Phi(b_i)}{\partial\beta_2} = \phi\,(b_i)\; X_i' \qquad \frac{\partial\Phi(b_i)}{\partial\beta_1} = 0 \qquad \frac{\partial\Phi(b_i)}{\partial\rho} = 0$$

where $\phi\,(\cdot)$ denotes the standard normal density function, $f_i = f(X_i\beta_1,\ X_i\beta_2;\ \rho)$ denotes the standard bivariate normal density function and

$$A_i = \frac{1}{\sqrt{1-\rho^2}}\ (b_i - \rho a_i)$$

$$B_i = \frac{1}{\sqrt{1-\rho^2}}\ (a_i - \rho b_i)\ .$$

If $\rho$ is given a priori, the equation (1) is still the same. But $\Theta$ now is $(\beta_1',\ \beta_2')'$ instead of $(\beta_1',\ \beta_2',\ \rho)'$ and the information matrix is $(2K)\cdot(2K)$.

If the two probit equations are separately estimated, the log-likelihood function of the first equation is

$$\ell n\ L\ (\beta_1) = \sum_i^N \{y_{i1}\ \ell n\ \Phi(a_i) + (1-y_{i1})\ \ell n\ \Phi(-a_i)\}$$

and the information matrix of the first probit equation is

$$\mathcal{J} = \sum_i^N M(a_i)\ M(-a_i) X_i' X_i$$

where $M(a_i) = \dfrac{\phi(a_i)}{\Phi(a_i)}$ .

Using the same method, we can get the information matrix of the second equation. Separate estimation will not be as efficient as joint estimation unless $\rho=0$.

Here the information matrix for joint estimation with full observability is the same as given by Amemiya (1974) using the FIMC (Full Information Minimum Chi-Square) Probit method. When the probit equations are estimated separately, the information matrix is the same as his using LIMC (Limited Information Minimum Chi-Square) Probit Method.

Under certain conditions, the information matrices of other (partial observability) cases will be singular. But this is not the case here. For full observability, the parameters are identified, except of course in the case of perfect multicollinearity. Perfect multicollinearity is also a perverse case for all of the levels of observability which we consider, but we will not discuss it further.

## 3.3 Case Two: Partial Observability in the Sense of Poirier

The log-likelihood function of Poirier's partial observability model is $\ln L(\Theta) = \sum_i^N \{Z_i \ln F_i + (1-Z_i) \ln [1-F_i]\}$ and the information matrix is

$$\mathcal{I} = C_5'C_5 \quad \text{----(2)}$$

where $C_5$ is the $N \cdot (2K+1)$ matrix with ith row equalling

$$\frac{1}{\sqrt{F_i(1-F_i)}} [\phi(a_i)\ \Phi(A_i)X_i,\ \phi(b_i)\ \Phi(B_i)\ X_i,\ f_i]\ .$$

When $\rho$ is equal to zero, the ith row of $C_5$ becomes

$$\frac{1}{\sqrt{\Phi(a_i)\Phi(b_i)[1-\Phi(a_i)\Phi(b_i)]}} [\phi(a_i)\Phi(b_i)X_i,\ \phi(b_i)\Phi(a_i)X_i,\ \phi(a_i)\phi(b_i)]\ .$$

If $\rho$ is known, $\Theta$ is just $(\beta_1', \beta_2')'$ and $C_5$ is a $N \cdot 2K$ matrix with ith row equal to

$$\frac{1}{\sqrt{F_i(1-F_i)}} [\phi(a_i)\Phi(A_i)X_i, \ \phi(b_i)\Phi(B_i)X_i] ,$$

which is the same as for $\rho$ unknown except that the last element has been dropped.

In discussing the identification problem, some simplifications are made here. We assume that there are only two independent variables and one of them is a constant term. That is, we assume

$$X_i = [1 \quad X_{i2}]$$

$$\beta_1 = \begin{bmatrix} \beta_{11} \\ \beta_{12} \end{bmatrix}$$

$$\beta_2 = \begin{bmatrix} \beta_{21} \\ \beta_{22} \end{bmatrix} .$$

(The same simplification will be used in discussing identification in the rest of this chapter.) These assumptions are only for the purpose of simplification, and won't change our conclusion.

The first perverse case we consider is the case of $\beta_{12} = \beta_{22} = 0$. That is, all coefficients equal zero except for the constant term.

Then

$$a_i = \beta_{11} = a$$

$$b_i = \beta_{21} = b$$

$$A_i = (\beta_{21} - \rho\beta_{11})/\sqrt{1-\rho^2} = A$$

$$B_i = (\beta_{11} - \rho\beta_{21})/\sqrt{1-\rho^2} = B$$

$$f_i = f(a, b; \rho) = f$$

$$F_i = F(a, b; \rho) = F$$

are all constants for i=1,2,...,N.

Under the assumptions above, the information matrix is

$$\mathcal{I} = \sum_i^N \frac{1}{F(1-F)} (F_\Theta)_i (F_\Theta)_i'$$

where

$$(F_\Theta)_i = \begin{bmatrix} \phi(a)\Phi(A) \\ \phi(a)\Phi(A)X_{i2} \\ \phi(b)\Phi(B) \\ \phi(b)\Phi(B)X_{i2} \\ f \end{bmatrix}.$$

It can be seen that in the information matrix,

the first row $\times \dfrac{\phi(b)\Phi(B)}{\phi(a)\Phi(A)}$ = the third row,

the first row $\times \dfrac{f}{\phi(a)\Phi(A)}$ = the last row,

and the second row $\times \dfrac{\phi(b)\Phi(B)}{\phi(a)\Phi(A)}$ = the fourth row.

Thus the rank of the information matrix is only two; the parameters are not identified. This is so even if $\rho$ is known a priori.

The second perverse case we consider, which was noted by Poirier, is the case in which $\beta_1=\beta_2$. (That is, $\beta_{11}=\beta_{21}$, $\beta_{12}=\beta_{22}$.) Then $b_i=a_i$, $B_i=A_i$ and the information matrix is $\mathcal{J} = \sum_i \frac{1}{F_i(1-F_i)} (F_\Theta)_i \, (F_\Theta)_i'$

where

$$(F_\Theta)_i = \begin{bmatrix} \phi(a_i)\Phi(A_i) \\ \phi(a_i)\Phi(A_i)X_{i2} \\ \phi(a_i)\Phi(A_i) \\ \phi(a_i)\Phi(A_i)X_{i2} \\ f_i \end{bmatrix}$$

The information matrix will have rank equal to three, and the parameters are not identified. Again, this is so whether or not $\rho$ is known a priori.

The third perverse case we consider is the case in which $\beta_{11}=\rho\beta_{21}$ and $\beta_{12}=\rho\beta_{22}$. Then $a_i=\rho b_i$

$$\Phi(B_i) = \Phi\left(\frac{a_i-\rho b_i}{\sqrt{1-\rho^2}}\right) = \Phi(0) = \tfrac{1}{2}$$

$$f_i = \frac{1}{2\Pi\sqrt{1-\rho^2}} \exp \left\{- \frac{(\rho b_i)^2 + b_i^2 - 2\rho(\rho b_i)b_i}{2(1-\rho^2)}\right\}$$

$$= \frac{1}{2\Pi\sqrt{1-\rho^2}} \exp \left\{- \frac{b_i}{2}\right\}$$

$$= s \cdot \phi(b_i)$$

where $s = \frac{1}{\sqrt{2\Pi(1-\rho^2)}}$

and the information matrix is of the same form as in the second perverse case where

$$(F_{\Theta})_i = \begin{bmatrix} \phi(a_i)\Phi(A_i) \\ \phi(a_i)\Phi(A_i)X_{i2} \\ \frac{1}{2}\,\phi(b_i) \\ \frac{1}{2}\,\phi(b_i)X_{i2} \\ s\phi(b_i) \end{bmatrix} .$$

It can be seen that the third row times 2s is equal to the last row, so the information matrix is singular. This will not occur if $\rho$ is known a priori.

The similar situation happens when $\beta_{21}=\rho\beta_{11}$, $\beta_{22}=\rho\beta_{12}$, $\Phi(A_i)=\frac{1}{2}$ and $f_i = s\cdot\phi(a_i)$ where s is the same as above, then the first row times 2s is equal to the last row and the information matrix is singular. Also there is no problem if $\rho$ is known a priori.

These perverse cases can be given an intuitive explanation. In this partial observability case, there are only two distinguishable events (yes, no) and one probability is independently estimatable. When the coefficients of the exogenous variables ($X_i$) except for the constant term are all equal to zero, this probability is unrelated to $X_i$ except for the constant term. Hence there exist only K(the number of exogenous variables) pieces of information, namely one probability and (K-1) things it doesn't correlate with. From this, we can't estimate (2K+1) parameters ($\rho$ is not known) or 2K parameters ($\rho$ is known). In the case when $\beta_1=\beta_2$, the two probit equations are observed to be the same and there are only (K+1) pieces of information. We still don't have enough information to estimate all the parameters. In the case when $\beta_{11}=\rho\beta_{21}$ and $\beta_{12}=\rho\beta_{22}$ (or $\beta_{21}=\rho\beta_{11}$ and $\beta_{22}=\rho\beta_{12}$), there are 2K pieces of information. We can't estimate all the parameters when $\rho$ is not known, but won't have the same problem if $\rho$ is known.

Another situation in which the information matrix will be singular is called "the peculiar case" by Poirier and has been discussed clearly in his paper. This case involves specific exogenous variable configurations and will not be discussed here.

There may be some other situations that the information matrix will be singular; the above cases do not necessarily cover all. It appears that in general one needs to check to see whether parameters are identified according to whether the information matrix has rank equalling 2k+1 for each specific condition. This is also true for all the identification problems of other partial observability cases that we are going to discuss in other sections.

3.4 <u>Case Three</u>: <u>Partial Partial Observability</u>

The log-likelihood function for joint estimates is

$$\ell n\, L(\Theta) = \sum_{i}^{N} \{y_{i1}\, y_{i2}\, \ell n F_i + y_{i1}(1- y_{i2})\, \ell n[\Phi(a_i) - F_i] + (1- y_{i1})\, \ell n[1- \Phi(a_i)]\}$$

$$\Theta = (\beta_1{}', \beta_2{}', \rho)'$$

$$i = 1,2,\ldots,N$$

and the information matrix is

$$\mathcal{I} = c_1{}'c_1 + c_2{}'c_2 + c_6{}'c_6$$

where $c_1$, $c_2$ are the same as before and $c_6{}'$ is a $(2k+1)\cdot N$ matrix with ith column equalling

$$\frac{1}{\sqrt{1- \Phi(a_i)}}\ \frac{\partial[1- \Phi(a_i)]}{\partial\Theta}\ .$$

When $\rho$ is known, $\Theta$ changes to $(\beta_1', \beta_2')'$ and the information matrix is a $2k \cdot 2k$ matrix.

The first probit equation -- the observable one -- can always be estimated separately by maximizing

$$\ell n\, L(\beta_1) = \sum_i \{y_{i1}\, \ell n\, \Phi(a_i) + (1-y_{i1})\, \ell n\, \Phi(-a_i)\}$$

and the information matrix is

$$\mathcal{I} = \sum_i M(a_i)M(-a_i)X_i'X_i$$

where $M(a_i) = \dfrac{\phi(a_i)}{\Phi(a_i)}$ .

Only when $\rho=0$, the second probit can be estimated separately from the observations with $y_{i1}$ by maximizing

$$\ell n\, L(\beta_2) = \sum_i y_{i1}\, \{y_{i2}\, \ell n\, \Phi(b_i) + (1-y_{i2})\, \ell n\, \Phi(-b_i)\}$$

and the information matrix is

$$\mathcal{I} = \sum_i \Phi(a_i)\ [M(b_i)M(-b_i)]X_i'X_i$$

where $M(b_i) = \dfrac{\phi(b_i)}{\Phi(b_i)}$ .

Under the assumption that $X_i = [1\ X_{i2}]$, $\beta_1 = \begin{bmatrix}\beta_{11}\\ \beta_{12}\end{bmatrix}$, $\beta_2 = \begin{bmatrix}\beta_{21}\\ \beta_{22}\end{bmatrix}$ and for the case that $\beta_{12}=\beta_{22}=0$, the information matrix for the joint estimation is

$$\mathcal{I} = \sum_i^N \{\frac{1}{F}(F_\Theta)_i(F_\Theta)_i' + \frac{1}{\Phi(a)-F}[(\Phi_\Theta)_i-(F_\Theta)_i][(\Phi_\Theta)_i - (F_\Theta)_i]' + \frac{1}{1-\Phi(a)}(\Phi_\Theta)_i(\Phi_\Theta)_i'\}$$

where $(F_\Theta)_i = \begin{bmatrix} \phi(a)\Phi(A) \\ \phi(a)\Phi(A)X_{i2} \\ \phi(b)\Phi(B) \\ \phi(b)\Phi(B)X_{i2} \\ f \end{bmatrix}$

and

$$(\Phi_\Theta)_i = \begin{bmatrix} \phi(a) \\ \phi(a)X_{i2} \\ 0 \\ 0 \\ 0 \end{bmatrix} .$$

It can be seen that the third row of the whole information matrix times $\frac{f}{\phi(b)\Phi(B)}$ will be equal to the last row, so the information matrix is singular.

Another case is when $\beta_{11}=\rho\beta_{21}$ and $\beta_{12}=\rho\beta_{22}$, then $a_i=\rho b_i$, $\Phi(B_i)=\frac{1}{2}$ and $f_i=s\cdot\phi(b_i)$ where $s = \frac{1}{\sqrt{2\Pi(1-\rho^2)}}$ . The information matrix is of the form that

$$\mathcal{J} = \sum_i \{\frac{1}{F_i} (F_\Theta)_i(F_\Theta)_i' + \frac{1}{\Phi(a_i)-F_i} [(\Phi_\Theta)_i-(F_\Theta)_i][(\Phi_\Theta)_i-(F_\Theta)_i]'$$

$$+ \frac{1}{1-\Phi(a_i)} (\Phi_\Theta)_i(\Phi_\Theta)_i'\}$$

where

$$(F_\Theta)_i = \begin{bmatrix} \phi(a_i)\Phi(A_i) \\ \phi(a_i)\Phi(A_i)X_{i2} \\ \frac{1}{2}\phi(b_i) \\ \frac{1}{2}\phi(b_i)X_{i2} \\ s\phi(b_i) \end{bmatrix}$$

and

$$(\Phi_{\Theta})_i = \begin{bmatrix} \phi(a_i) \\ \phi(a_i)X_{i2} \\ 0 \\ 0 \\ 0 \end{bmatrix} .$$

It also can be seen that the third row time 2s equals the last row, so the information matrix is singular.

Since the last row of the information matrix is the one corresponding to $\rho$, for the above two perverse cases, the parameters are identified if $\rho$ is known a priori (in which case the last row and column are deleted), but they are not identified when $\rho$ is unknown. However, it should be pointed out that $\beta_1$ is always identified; the lack of identification is only for $\beta_2$ and $\rho$.

Intuitively, since there are three distinguishable events (yes, no but party one voted yes, no and party one voted no), there are two independent probabilities. Thus there are 2K pieces of information in the data for the above two perverse cases, which do identify $\beta_1$ and $\beta_2$ when $\rho$ is known, but which do not identify all of $\beta_1$, $\beta_2$ and $\rho$.

There also may exist some other situations in which identification fails, but these are the only two that we found so far.

3.5 Case Four: Partial Observability with Observed Veto

First we will discuss the case in which there is a given probability for party one to cast the veto first when both parties wish to do so. Let p be this given probability, then the log-likelihood function is

$$\ell n\, L(\theta) = \sum_{\substack{i \ni Z_i = 1 \\ \text{is observed}}} \ell n\, F_i$$

$$+ \sum_{\substack{i \ni y_{i1}=0 \\ \text{is observed}}} \ell n \, [p(1- \Phi(a_i)) + (1-p)(\Phi(b_i)-F_i)]$$

$$+ \sum_{\substack{i \ni y_{i2}=0 \\ \text{is observed}}} \ell n \, [p(\Phi(a_i)-F_i) + (1-p)(1- \Phi(b_i))]$$

where $\Theta = (\beta_1', \beta_2', \rho)'$

$i = 1,2,\ldots,N$

and the information matrix of this case is

$$\mathcal{I} = c_1'c_1 + c_7'c_7 + c_8'c_8$$

where $c_1'$ is as before, $c_7'$ and $c_8'$ are matrices of dimension $(2k+1)\cdot N$ with ith columns given by

$$\frac{1}{\sqrt{p(1-\Phi(a_i))+(1-p)(\Phi(b_i)-F_i)}} \, \frac{\partial[p(1-\Phi(a_i))+(1-p)(\Phi(b_i)-F_i)]}{\partial\Theta}$$

and

$$\frac{1}{\sqrt{p(\Phi(a_i)-F_i)+(1-p)(1-\Phi(b_i))}} \, \frac{\partial[p(\Phi(a_i)-F_i)+(1-p)(1-\Phi(b_i))]}{\partial\Theta}$$

respectively.

If p is another parameter which is unknown and needs to be estimated, so that $\Theta$ now includes p, this does not change any of the above except that we now add another column to $c_1$, $c_7$ and $c_8$. The extra column of $c_1$ is zero, the extra column of $c_7$ has ith element

$$\frac{1}{\sqrt{p(1-\Phi(a_i))+(1-p)(\Phi(b_i)-F_i)}} \, [1- \Phi(a_i)- \Phi(b_i)+F_i]$$

and the extra column of $c_8$ has ith element

$$\frac{-1}{\sqrt{p(\Phi(a_i)-F_i)+(1-p)(1-\Phi(b_i))}} \; [1-\Phi(a_i)-\Phi(b_i)+F_i] \; .$$

The information matrix now is a $(2k+2) \cdot (2k+2)$ matrix.

Now we discuss another possibility of observed veto under the assumption that the first party to cast the veto is the one whose desire for a "no" vote is stronger. That is, given $X_i\beta_1 + \varepsilon_{i1} < 0$ and $X_i\beta_2 + \varepsilon_{i2} < 0$, it is assumed that the party with the lesser value is observed to cast the veto first. Letting $H_i$ be the probability that $X_i\beta_1 + \varepsilon_{i1} < X_i\beta_2 + \varepsilon_{i2}$ given $X_i\beta_1 + \varepsilon_{i1} < 0$ and $X_i\beta_2 + \varepsilon_{i2} < 0$, the log-likelihood function is

$$\begin{aligned} \ln L(\Theta) = & \sum_{\substack{i \ni Z_i=1 \\ \text{is observed}}} \ln F_i \\ & + \sum_{\substack{i \ni y_{i1}=0 \\ \text{is observed}}} \ln [\Phi(b_i)-F_i + H_i] \\ & + \sum_{\substack{i \ni y_{i2}=0 \\ \text{is observed}}} \ln [1-\Phi(b_i)-H_i] \; . \end{aligned}$$

The information matrix then is

$$\mathcal{I} = c_1'c_1 + c_9'c_9 + c_{10}'c_{10}$$

where $c_1'$ is as above, $c_9'$ and $c_{10}'$ are $(2k+1)\cdot N$ matrices and have ith columns respectively as

$$\frac{1}{\sqrt{\Phi(b_i)-F_i+H_i}} \quad \frac{\partial[\Phi(b_i)-F_i+H_i]}{\partial\Theta}$$

and

$$\frac{1}{\sqrt{1-\Phi(b_i)-H_i}} \frac{\partial[1-\Phi(b_i)-H_i]}{\partial\Theta}$$

where $\Theta = (\beta_1', \beta_2', \rho)'$

$$\frac{\partial H_i}{\partial\beta_1} = -\phi(\frac{b_i-a_i}{\sqrt{2(1-\rho)}})\Phi(-\frac{a_i+b_i}{\sqrt{2(1+\rho)}}) \frac{1}{\sqrt{2(1-\rho)}} X_i'$$

$$\frac{\partial H_i}{\partial\beta_2} = \phi(\frac{b_i-a_i}{\sqrt{2(1-\rho)}})\Phi(-\frac{a_i+b_i}{\sqrt{2(1+\rho)}}) \frac{1}{\sqrt{2(1-\rho)}} X_i'$$

$$-\phi(b_i)[1-\Phi(B_i)]X_i'$$

$$\frac{\partial H_i}{\partial\rho} = \frac{1}{2}\frac{1}{\sqrt{2(1-\rho)}} [\phi(\frac{b_i-a_i}{\sqrt{2(1-\rho)}})\Phi(-\frac{a_i+b_i}{\sqrt{2(1+\rho)}})(\frac{b_i-a_i}{1-\rho})$$

$$-f(\frac{b_i-a_i}{\sqrt{2(1-\rho)}}, -b_i; -\sqrt{(1-\rho)/2})]$$

The derivations of these three derivatives are in Appendix A.

Identification problems still occur under the assumptions that $X_i = [1\ X_{i2}]$, $\beta_1 = [\begin{smallmatrix}\beta_{11}\\ \beta_{12}\end{smallmatrix}]$ and $\beta_2 = [\begin{smallmatrix}\beta_{21}\\ \beta_{22}\end{smallmatrix}]$ as before. When $\beta_{12} = \beta_{22} = 0$ and $\rho$ is unknown, the information matrix of the case of observed veto with a given probability p is of the form

$$\mathcal{J} = \sum_{i=1}^{N} \{\frac{1}{F}(F_\Theta)_i(F_\Theta)_i' + \frac{1}{Q_1}(Q_{1\Theta})_i(Q_{1\Theta})_i' + \frac{1}{Q_2}(Q_{2\Theta})_i(Q_{2\Theta})_i'\}$$

where

$$Q_1 = p[1-\Phi(a)] + (1-p)[\Phi(b)-F]$$
$$Q_2 = p[\Phi(a)-F] + (1-p)[1-\Phi(b)]$$

$$(F_\Theta)_i = \begin{bmatrix} \phi(a)\Phi(A) \\ \phi(a)\Phi(A)X_{i2} \\ \phi(b)\Phi(B) \\ \phi(b)\Phi(B)X_i \\ f \end{bmatrix}$$

$$(Q_{1\Theta})_i = \begin{bmatrix} \phi(a)\ [-p-(1-p)\Phi(A)] \\ \phi(a)\ [-p-(1-p)\Phi(A)]X_{i2} \\ (1-p)\phi(b)\ [1-\Phi(B)] \\ (1-p)\phi(b)\ [1-\Phi(B)]X_{i2} \\ -(1-p)f \end{bmatrix}$$

and

$$(Q_{2\Theta})_i = \begin{bmatrix} p\phi(a)\ [1-\Phi(A)] \\ p\phi(a)\ [1-\Phi(A)]X_{i2} \\ \phi(b)\ [-(1-p)-p\ \Phi(B)] \\ \phi(b)\ [-(1-p)-p\ \Phi(B)]X_{i2} \\ -pf \end{bmatrix}.$$

It can be seen that for the whole information matrix

(the first row x $q_1$) + (the third row x $q_2$)
= the last row

where

$$q_1 = \frac{(1-p)f}{\phi(a)\ [(1-p)\Phi(A)+p\Phi(B)]}$$

$$q_2 = \frac{pf}{\phi(b)\ [(1-p)\Phi(A)+p\Phi(B)]} \quad .$$

So the information matrix of the case of observed veto with given probability p is singular, and neither $\beta_1$ nor $\beta_2$ can be identified. However this is true only when $\rho$ is unknown. If $\rho$ is known and $\Theta=(\beta_1', \beta_2',)'$, the information matrix is not singular when $\beta_{12} = \beta_{22} = 0$.

In the case of observed veto with p as a parameter, the information matrix is still of the form as before

$$\mathcal{J} = \sum_{i}^{N} \{\frac{1}{F} (F_\Theta)_i (F_\Theta)_i' + \frac{1}{Q_1} (Q_{1\Theta})_i (Q_{1\Theta})_i' + \frac{1}{Q_2} (Q_{2\Theta})_i (Q_{2\Theta})_i' \},$$

but $\Theta$ now includes p as $(\beta_1', \beta_2', \rho, p)'$ and $(F_\Theta)_i$, $(Q_{1\Theta})_i$ and $(Q_{2\Theta})_i$ are all matrices with dimension $6\cdot1$. The extra rows of $(F_\Theta)_i$, $(Q_{1\Theta})_i$ and $(Q_{2\Theta})_i$ are 0, $\frac{1}{Q_1}$ $[1- \Phi(a)- \Phi(b) + F]$ and $\frac{-1}{Q_2}$ $[1- \Phi(a) - \Phi(b) + F]$ respectively. For this $6\cdot6$ information matrix, besides the relationship that

(the first row x $q_1$) + (the third row x $q_2$)
= the fifth row

where $q_1$, $q_2$ are the same as before, it is also true that

[the first row x $\phi(b)\Phi(B)$ - the third row x $\phi(a)\Phi(A)$]

$$\times \frac{-[1- \Phi(a) - \Phi(b) + F]}{[p\ \phi(a)\ \phi(b)\Phi(B) + (1-p)\ \phi(a)\ \phi(b)\Phi(A)]}$$

= the last row.

Therefore for the case of observed veto with p as another parameter, when $\beta_{12} = \beta_{22} = 0$ the information matrix is singular whether $\rho$ is known a priori or not. Another way to say this is that this model, $\beta_1$ and $\beta_2$ are identified only if both $\rho$ and p are known.

Finally we come to the question of identification for the case of observed veto under the assumption that the first to use the veto is the one who wants it worst. When $\beta_{12} = \beta_{22} = 0$ and $\rho$ is not known, the information matrix is

$$\mathcal{I} = \sum_i^N \{\frac{1}{F}(F_\Theta)_i(F_\Theta)_i' + \frac{1}{R_1}(R_{1\Theta})_i(R_{1\Theta})_i' + \frac{1}{R_2}(R_{2\Theta})_i(R_{2\Theta})_i'\}$$

where $(F_\Theta)_i$ is the same as above,

$$R_1 = \Phi(b) - F + H$$
$$R_2 = 1 - \Phi(b) - H$$

$$(R_{1\Theta})_i = \begin{bmatrix} -\phi(a)\Phi(A) + H_{\beta_1} \\ [-\phi(a)\Phi(A) + H_{\beta_1}]X_{i2} \\ -H_{\beta_1} \\ [-H_{\beta_1}]X_{i2} \\ -f + H_\rho \end{bmatrix}$$

$$(R_{2\Theta})_i = \begin{bmatrix} -H_{\beta_1} \\ (-H_{\beta_1})X_{i2} \\ H_{\beta_1} - \phi(b)\Phi(B) \\ [H_{\beta_1} - \phi(b)\Phi(B)]X_{i2} \\ -H_\rho \end{bmatrix}$$

and where

$$H = F(\frac{b-a}{\sqrt{2(1-\rho)}}, -b; -\sqrt{(1-\rho)/2})$$

$$H_{\beta_1} = - \phi(\frac{b-a}{\sqrt{2(1-\rho)}})\ \Phi(-\frac{a+b}{\sqrt{2(1+\rho)}})\ \frac{1}{\sqrt{2(1-\rho)}}$$

$$H\rho = \frac{1}{2}\ [-\ H_{\beta_1}\ (\frac{b-a}{1-\rho}) + f]$$

$H$, $H_{\beta_1}$ and $H_\rho$ are constant for $i=1,2,\ldots,N$.

It can be seen that for this information matrix

(the first row x $r_1$) + (the third row x $r_2$)

= the last row

where

$$r_1 = \frac{\phi(b)\Phi(B)[f-H_\rho]-f\ H_{\beta_1}}{\phi(a)\phi(b)\Phi(A)\Phi(B)-[\phi(a)\Phi(A)+\phi(b)\Phi(B)]H_{\beta_1}}$$

$$r_2 = \frac{\phi(a)\Phi(A)H_\rho - fH_{\beta_1}}{\phi(a)\phi(b)\Phi(A)\Phi(B)-[\phi(a)\Phi(A)+\phi(b)\Phi(B)]H_{\beta_1}}\ .$$

So for this case, the information matrix is singular when $\beta_{12} = \beta_{22} = 0$ and $\rho$ is not known, but not so when $\rho$ is known.

Intuitively, when the coefficients of all the exogenous variables except the constant term are equal to zero, there are three distinguishable events (yes, no with party one vetoing, no with party two vetoing). Hence two probabilities are independently estimatable and there are 2K pieces of information available. Thus we can identify at most 2K parameters. From this point of view we can see that the information matrices in the first and the third situations are singular because there are (2K+1) parameters when $\rho$ is not known. But all the parameters can be identified if $\rho$ is given. The second case (which is the one with p as another parameter) can't be identified unless both $\rho$ and p are known.

## 3.6 Summary

In this chapter, six different information matrices have been derived corresponding to joint estimation of the bivariate probit model with varying degrees of observability. Some information matrices for separate estimation of the two equations are also derived, but separate estimation won't be as efficient as joint estimation unless $\rho=0$. Because it is the inverse of the information matrix that is the asymptotic variance - covariance matrix of parameters, we also discuss question of identification by analyzing the rank of these matrices. We especially analyze the perverse case in which all the coefficients of exogenous variables are equal to zero. It is found that in the above situation, regardless of the values of the constant terms, all of the information matrices for the partial observability cases are singular if $\rho$ is not known. When $\rho$ is known a priori, only the cases of partial observability in the sense of Poirier and of observed veto with p as another parameter still suffer from a lack of identification. The model with full observability is still identified in this case, whether $\rho$ is known or not.

The second perverse case is the one in which $\beta_1 = \beta_2$; that is, the two probit equations are identical. Then the model with partial observability in the sense of Poirier is not identified, though there are no problems with the other cases.

The third perverse case is when $\beta_{11} = \rho\beta_{21}$ and $\beta_{12} = \rho\beta_{22}$. The model with partial observability in the sense of Poirier and the partial partial observability case are not identified when $\rho$ is unknown. But there are no problems when $\rho$ is known a priori. In the similar situation when $\beta_{21} = \rho\beta_{11}$ and $\beta_{22} = \rho\beta_{12}$, only the model with partial observability in the sense of Poirier is not identified when $\rho$ is unknown.

These results will be useful in picking parameter points at which to evaluate the cost of partial observability. This we will do in the next chapter.

# CHAPTER FOUR

# RESULTS OF EXPERIMENTS MEASURING THE COST OF PARTIAL OBSERVABILITY

## 4.1 Introduction

In this chapter, the results of some experiments measuring the cost of partial observability are presented. We call these "experiments" because various specific values of the parameters have to be picked in order to evaluate the inverse of information matrix. What we are interested in is primarily (1) with the same values of the parameters, a comparison of efficiencies under different levels of partial observability; and (2) the reasons that will cause the change of cost for each individual level of partial observability. With this knowledge, a researcher can compare costs of using and not using a piece of information according to his case and make a better decision.

When measuring the cost of partial observability, we let the elements of the inverse of the information matrices for the various partial observability cases be divided by the corresponding elements of the inverse of the information matrix for case one. The reason for choosing case one as a standard of comparison is because it has the most complete observability and therefore leads to the most efficient estimates. Thus the ratios that we get are the ratios of asymptotic variances and covariances for the various partial observability cases compared to the full observability case, and represent the relative efficiencies of parameter estimates under different levels of observability. The bigger the ratios are, the greater the cost of partial observability.

Some simplifications made in the last chapter are still applied here. Specifically, we assume $X_i=(1\ X_{i2})$, $\beta_1 = \begin{bmatrix}\beta_{11}\\ \beta_{12}\end{bmatrix}$ , $\beta_2 = \begin{bmatrix}\beta_{21}\\ \beta_{22}\end{bmatrix}$ , where $X_{i2} = e^{-X_{i3}}$ and the $X_{i3}$ are random normal deviates with zero mean and unit variance. All the experiments in this chapter have been done with a sample size of 50. We also tried sample sizes of 10, 100 and 500, but the results didn't show any significant difference. This simply indicates that the ratios of asymptotic variances are more or less independent of sample size. We do not address the question of what sample sizes are necessary for the asymptotic distributions to be reliable.

In section 4.2, we first try three arbitrary cases $\beta_1=\beta_2=\begin{bmatrix}1\\ 1\end{bmatrix}$; $\beta_1=\beta_2=\begin{bmatrix}0\\ 1\end{bmatrix}$; $\beta_1=\begin{bmatrix}1\\ -1\end{bmatrix}$ and $\beta_2=\begin{bmatrix}1\\ 1\end{bmatrix}$. We present some general results which we believe are always true for any values of $\beta$. In section 4.3, all the experiments have been done with one common characteristic, namely that $\sum_i X_i\beta_1 = \sum_i X_i\beta_2 = 0$ (ensured by appropriate choice of value of the constant term). The idea of these experiments is to show the effects of changes in the parameters ($\beta_1$, $\beta_2$, $\rho$ and p) when (on average) each party has an equal probability of saying "yes" and "no". These results are not as readily interpreted as we might hope, since changing any parameter (e.g. $\beta_{12}$) changes a number of features of the data and model which are relevant to the relative efficiencies, such as the degree of identification, the split of the sample into the various distinguishable outcomes, and so forth. Therefore in section 4.4 we try some more complicated experiments in which we manipulate the parameters in such a way as to isolate these various influences. For the most part these results are in accord with our a priori expectations. Section 4.5 gives the summary of this chapter.

Twenty-two tables at the end of this chapter present the results of the experiments. Only the ratios of asymptotic variances of the parameter estimates of the non-constant-term exogenous variables ($\beta_{12}$ and $\beta_{22}$), for different levels of partial observability, are listed and discussed. Except for those experiments especially designed to focus on $\rho$ and p, all of the experiments have been done for $\rho=0$ and $\rho=0.5$, and p is fixed at 0.5 for case four. As before, case one is the full observability case; case two is the case of partial observability in the sense of Poirier; case three is the partial partial observability case; case 4A1 is the observed veto case given a known value of p; case 4A2 is the observed veto case with p as another parameter; and case 4B is the observed veto case under the assumption that the party who uses the veto first is the one that wants the veto more strongly. Also we define

$$\overline{P_{11}} = \sum_{i}^{N} F(X_i\beta_1, X_i\beta_2; \rho)/N$$

$$\overline{P_{10}} = \sum_{i}^{N} [\Phi(X_i\beta_1) - F(X_i\beta_1, X_i\beta_2; \rho)]/N$$

$$\overline{P_{01}} = \sum_{i}^{N} [\Phi(X_i\beta_2) - F(X_i\beta_2, X_i\beta_2; \rho)]/N$$

$$\overline{P_{00}} = \sum_{i}^{N} [1- \Phi(X_i\beta_1) - \Phi(X_i\beta_2) + F(X_i\beta_1, X_i\beta_2; \rho)]/N$$

$$\overline{H} = \sum_{i}^{N} F(\frac{X_i\beta_2 - X_i\beta_1}{\sqrt{2(1-\rho)}}, \quad -X_i\beta_2; \; - \sqrt{(1-\rho)/2}\,)/N \, .$$

They are the average probabilities of: both parties say "yes"; party one says "yes" and party two says "no"; party one says "no" and party two says "yes"; both parties say "no"; and party one uses veto first given that both parties desire to do so; respectively. The distribution of the first four probabilities are called "sample split" and all the probabilities are listed in some of the tables. Also in the tables, "$\rho$ is not known" means that $\rho$ is not given but rather is estimated as a parameter, and the information matrices are of dimension $(2k+1)\cdot(2k+1)$ (or $(2k+2)\cdot(2k+2)$ for case 4A2).

## 4.2 General Results of Some Basic Experiments

The experiments presented in this section are called "Experiment 1" in order to be distinguished from the others in section 4.3 and 4.4. Experiment 1 includes three different $\beta$'s. They are (1) $\beta_1 = \beta_2 = [{1 \atop 1}]$; (2) $\beta_1 = \beta_2 = [{0 \atop 1}]$; and (3) $\beta_1 = [{1 \atop -1}]$ and $\beta_2 = [{1 \atop 1}]$. For the first two choices of $\beta$, Poirier's model (case two) is not identified because $X_i\beta_1 = X_i\beta_2$ for all i. The other cases are all identified.

Table 1 lists the (expected) sample splits for each case under both $\rho$=0 and $\rho$=0.5. Table 2 shows the relative efficiencies of $\beta_{12}$ and $\beta_{22}$ for each partial observability case. The results in Tables 1 and 2 are not easy to summarize, but we do note the following.

(1) When $\beta_1 = \beta_2 = [{1 \atop 1}]$, $X_i\beta_1$ and $X_i\beta_2$ are all positive numbers greater than 1, so the average probability of both parties voting "yes" is close to 1, and the average probability of both parties voting "no" is very small. Since $X_i\beta_1 = X_i\beta_2$, $\overline{P_{10}} = \overline{P_{01}}$; the average probability of one party voting "yes" and the other voting "no" is the same as the probability of the opposite situation. Both probabilities are small. This is a fairly extreme sample split.

(2) When $\beta_1 = \beta_2 = [{1 \atop 1}]$ changes to $\beta_1 = \beta_2 = [{0 \atop 1}]$, both the values of $X_i\beta_1$ and $X_i\beta_2$ are decreased, so the average probability that both parties will say "yes" is decreased too. On the other hand, the average probabilities that one or both parties say "no" are all increased. The samples split is still heavily weighted toward "both yes", but less extremely than before.

(3) Knowing $\rho$ results in smaller ratios of asymptotic variances, and hence reduces the cost of partial observability. This agrees with the general principle that we will call the "law of decreasing marginal utility of information" (LDMUI), which is that the more information one has, the less should be the value of another piece of information. Observability information is less valuable when $\rho$ is known a priori than when it is not.

(4) The Poirier case is the worst among all the partial observability cases either when $\rho$ is unknown or known, because it is the one that is based on the least information.

(5) For case three, the ratios of asymptotic variances for estimates of $\beta_{12}$ are equal to one for $\rho$=0 and are still very close to one for $\rho$=0.5, for all values of $\beta$. This is so because party one's behavior is fully observed in this case, and when $\rho$=0 party two's behavior is not informative for party one's parameters. This is not true however when $\rho \neq 0$. For $\beta_2$ (party two's parameters), the ratios are bigger than any of those for case four when $\rho$ is unknown. When $\rho$ is known, they are bigger than those for case 4A1 or case 4B but smaller than those for case 4A2.

(6) When $X_i\beta_1 = X_i\beta_2$ and $p = 0.5$, $\beta_{12}$ and $\beta_{22}$ have the same efficiencies for all three different possibilities of case four.

(7) The estimates for case 4A2 are less efficient than for case 4A1, since in case 4A1 more is known (namely p). In general the gain from knowing p is greater when $\rho$ is known (and conversely). This is a counter-example to the LDMUI.

(8) For case 4B, $\overline{H}$ is the average probability of $X_i\beta_1 + \varepsilon_{i1} < X_i\beta_2 + \varepsilon_{i2}$ given that $X_i\beta_1 + \varepsilon_{i1} < 0$ and $X_i\beta_2 + \varepsilon_{i2} < 0$. So when $X_i\beta_1 = X_i\beta_2$, $\overline{H} = \frac{1}{2}\overline{P_{00}}$. In the case of $\beta_1 = [{1 \atop -1}]$ and $\beta_2 = [{1 \atop 1}]$, $X_i\beta_1 < X_i\beta_2$, so that party one is more likely to use the veto first when both sides vote "no", and thus $\overline{H} > \frac{1}{2}\overline{P_{00}}$. Furthermore, the average probability of the indistinguishable outcome is $\overline{P_{01}} + \overline{H}$ for party two while it is $\overline{P_{10}} + \overline{P_{00}} - \overline{H}$ for party one. Hence $\beta_{22}$ tends to be less efficient than $\beta_{12}$ in case 4B when $\overline{H} > \frac{1}{2}\overline{P_{00}}$, as it is here.

(9) With $X_i\beta_1 = X_i\beta_2$ and $p = 0.5$, the efficiencies of $\beta_{12}$ and $\beta_{22}$ in case 4A1 are very close to those in case 4B. This is so because when $X_i\beta_1 = X_i\beta_2$, the probability of $X_i\beta_1 + \varepsilon_{i1} < X_i\beta_2 + \varepsilon_{i2}$ is indeed 0.5.

The above conclusions will be seen to hold also in the experiments yet to be presented, and thus are fairly general. (Some of them, of course, are perfectly general since they must be true.) We will not discuss them further.

There is some evidence above on the effects of changes in $\rho$. However, this will be discussed in the next section.

## 4.3 Results of Further Experiments

In this section we report the results of more experiments, which vary $\beta_1$, $\beta_2$, $\rho$ and p more widely than in the last section. All of the experiments in this section have in common the feature that $\sum_i X_i\beta_1 = \sum_i X_i\beta_2 = 0$. The point of this is to try to minimize effects of parameter changes on the sample split. For example, consider an experiment which varies $\beta_{12}$ from zero to some high level. As $\beta_{12}$ increases, for some of the partial observability cases the degree of identification increases and we might expect the cost of partial observability to fall. (We might call this an "identification effect".) However, if we change $\beta_{12}$ while holding $\beta_{11}$ constant, the probability of a "yes" vote also changes, and thus the sample split changes. This also may affect relative efficiencies; for example, as $\overline{P_{11}}$ increases there is fuller observability for all cases and the relative efficiencies of the partial observability estimators should increase. (We might call this a "sample split effect".) Therefore in this section when $\beta_{12}$ is changed, $\beta_{11}$ is also changed in such a way that $X_i\beta_1$ is (on average) zero: $\sum_i X_i\beta_1 = 0$, and similarly for $\beta_2$. For $\rho$=0 and small $\beta$'s, this will yield a symmetric sample split: $\overline{P_{00}} = \overline{P_{01}} = \overline{P_{10}} = \overline{P_{11}} = \frac{1}{4}$. For larger $\beta$'s this is unfortunately not so. Because of the non-linearity of the model, the average probabilities (over the sample) are not the same as the probabilities evaluated at the average of $X_i$. The latter will all equal $\frac{1}{4}$ (for $\rho$=0, anyway) but the former will not. Thus we are not entirely successful in separating identification effects from sample split effects. Indeed, it is not clear how well one can hope to do so, but some more successful attempts will be made in the next section.

Several types of experiments are presented in this section:

(1) Experiment 2: $\beta_1 = \beta_2 = [^{-c}_{c}{}^{\overline{X}}]$ for c = 0.3, 0.5, 1, 2 and 3,

where $\overline{X} = \sum_{i=1}^{N} X_{i2} = \sum_{i=1}^{N} e^{-X_{i3}} \simeq 1.525$ ,

$X_{i3}$ , i=1,2,...,N, are random numbers with standard normal distribution having $\mu=0$ and $\sigma=1$. The same X's will be used for all the experiments here, and the sample size is 50. The results of Experiments 2 are in Tables 3, 4 and 5.

(2) Experiment 3: $\beta_1 = [^{c}_{-c}{}^{\overline{X}}]$ and $\beta_2 = [^{-c}_{c}{}^{\overline{X}}]$ for c = 0.3, 0.5, 1, 2 and 3.

The results are in Tables 6, 7 and 8.

(3) Experiment 4A: $\beta_1 = \beta_2 = [^{-\overline{X}}_{1.0}]$;

Experiment 4B: $\beta_1 = [^{\overline{X}}_{-1.0}]$ and $\beta_2 = [^{-\overline{X}}_{1.0}]$ ,

both for $\rho$=-0.5, 0, 0.2, 0.5 and 0.8. The results are in Tables 9, 10, 11 and 12.

(4) Experiment 5A: $\beta_1 = \beta_2 = [^{-\overline{X}}_{1.0}]$;

Experiment 5B: $\beta_1 = [^{\overline{X}}_{-1.0}]$ and $\beta_2 = [^{-\overline{X}}_{1.0}]$,

both for p=0, 0.2, 0.5, 0.8 and 1 (but for cases 4A1 and 4A2 only). The results are in Tables 13 and 14.

It is hard to summarize so many tables briefly. However, we will discuss what we find to be the most interesting results.

For experiment 2, we know that case two is not identified for any value of c because $X_i\beta_1 = X_i\beta_2$, i=1,2,...,N, and case three ($\beta_2$ only) and case four are not identified if c is zero when $\rho$ is not known. If $\rho$ is known only case 4A2 is not identified when c is zero. As c increases, the "identification effect" should be to increase relative efficiencies of the partial observability cases. Meanwhile, we notice that the average probabilities of the indistinguishable outcomes, namely $(\overline{P_{01}} + \overline{P_{00}})$ for $\beta_2$ of case three and $\overline{P_{00}}$ for case four are getting larger when c increases. That is, the identification effect and the sample split effect work against each other when c changes.

From the results when $\rho$ is not known, we can see that when c is smaller (0.3, 0.5) the identification effect is quite strong so that the ratios of asymptotic variances are decreasing as c gets larger for both $\rho$=0 and $\rho$=0.5. But when c increases to a certain level (2 or 3), these two effects seem to cancel each other out and the results are less clear. For the $\rho$=0 case, whether these ratios will increase or decrease after c > 2.0 is uncertain. When $\rho$ = 0.5, although ratios are monotonically decreasing as c increases from 0.3 to 3.0, whether they would keep on increasing or not can't really be predicted.

When $\rho$ is known, the relative efficiencies of case three ($\beta_2$ only) and cases 4A1 and 4B are generally decreasing as c increases, presumably because of the sample split effect. For case 4A2, since it still has an identification problem if c is too close to zero, relative efficiencies are increasing as c increases until c=3.

In Experiment 3, all of the parameters are identified for all cases. However, for c=0 cases 2 and 4A2 would not be identified, while cases 3, 4A1 and 4B would be identified only with $\rho$ known. As for the sample split effect, the average probability of the indistinguishable outcome for case two, which is $(1-\overline{P_{11}})$, is increasing with c but those of case three ($\beta_2$ only) and case four are decreasing as c increases. Thus we are going to discuss the relative efficiencies of the partial observability cases one by one.

For case two, the relative efficiencies of both $\beta_1$ and $\beta_2$ improve as c increases from 0.3 to 3 for either $\rho=0$ or $\rho=0.5$, and whether $\rho$ is known or unknown. This shows that the identification effect is very strong for case two, which is not surprising.

For $\beta_2$ of case three, when $\rho$ is not known, the identification effect plus the sample split effect make the relative efficiency increase as c increases both for $\rho=0$ and $\rho=0.5$. When $\rho$ is known, the sample split effect alone makes the relative efficiency increase for both $\rho=0$ and $\rho=0.5$ after c=0.5.

For case 4A1, the effect of changing c is ambiguous. Note that case 4A1 yields much more efficient estimates than case 4A2, especially when c is small. The same is true in comparing case 4A1 to case 4B when $\rho$ is unknown. When $\rho$ is known, cases 4A1 and 4B yield rather similar efficiencies.

For case 4A2, for either $\rho=0$ or $\rho=0.5$ and whether $\rho$ is known or not, the identification effect and the sample split effect both make $\beta_1$ and $\beta_2$ relatively more efficient as c gets larger. Therefore, the efficiency relative to full observability is monotonically increasing all the way from c=0.3 to 3.

Since case 4B has identification problems only when $\rho$ is not known, its relative efficiency improves dramatically when c increases, when $\rho$ is unknown. When $\rho$ is known, the effect of increasing c is ambiguous.

Experiment 4 shows the effect of the correlation coefficient $\rho$. For Experiment 4A with $\rho$ unknown, the relative efficiencies of all cases (only $\beta_2$ for case three) get worse as $\rho$ increases. The changes are considerable especially for bigger $\rho$. When $\rho$ is known, the result is mostly the opposite. Most ratios get smaller as $\rho$ increases, but these are mixed results for case 4A2. However all the changes are much smaller; this is, $\rho$ doesn't matter much if it is known.

For Experiment 4B with $\rho$ unknown, relative efficiencies of all cases (only $\beta_2$ for case three) except case 4A1 improve as $\rho$ increases. The changes are especially big when $\rho$ increases from -0.5 to 0. The ratios for case 4A1 are very small compared to other cases, but they increase as $\rho$ increases. When $\rho$ is known, the relative efficiencies of case two and case 4A2 improve but cases three ($\beta_2$ only), 4A1 and 4B get worse as $\rho$ increases. All the changes are also smaller when $\rho$ is known.

For both Experiment 4A and 4B, the relative efficiency of $\beta_1$ in case three, being affected by the correlation with another unobservable party, gets away from 1 as the absolute value of $\rho$ increases.

Experiment 5 shows the effect of p on cases 4A1 and 4A2. For case 4A1 and 4A2, the average probability of the indistinguishable outcome for party one is $\overline{P_{10}} + (1-p)\ \overline{P_{00}}$ and is $\overline{P_{01}} + p\ \overline{P_{00}}$ for party two. So as expected, increasing p decreases the ratios for $\beta_1$ but increases them for $\beta_2$, and the effects are very strong.

Both experiments 5A and 5B have the same results for cases 4A1 and 4A2 whether $\rho$ is known or not, but the results are mixed for case 4A2 in Experiment 7B when $\rho$ is known.

## 4.4 The Results of Experiments with Either the Identification Effect or the Sample Split Effect Constant

From the results of the last section, we can see that because of the mixture of the identification effect and the sample split effect, sometimes we can not really tell the direction of change of the relative efficiency when a parameter changes. Therefore, in this section, we try some other experiments designed to change one effect while holding the other constant. All the experiments are done by adjusting the values of the constant terms to manipulate the sample split. Since different cases depend on different features of the sample split, we do a different experiment for each partial observability case.

There are three types of experiments in this section:

(1) Experiment 6A: $\beta_1 = [{d \atop -c}]$ and $\beta_2 = [{d \atop c}]$ for c = 0.3, 0.5, 1, 2 and 3;

d is adjusted so that $\overline{P_{11}}$ is fixed at 0.25.

Experiment 6B: $\beta_1 = [{d \atop -c}]$ , $\beta_2 = [{d \atop c}]$ and c=1; d is adjusted so that $\overline{P_{11}}$ varies from 0.15, 0.25, 0.35, 0.45, 0.55 to 0.65.

These are experiments designed for case two and results are in Tables 15 and 16.

(2) Experiment 7A: $\beta_1 = \beta_2 = [{-d \atop c}]$ for c = 0.3,0.5, 1, 2 and 3;

d is adjusted so that $\overline{P_{01}} + \overline{P_{00}} = 0.50$.

Experiment 7B: $\beta_1 = [{-d_1 \atop c}]$ and $\beta_2 = [{-d_2 \atop c}]$ for c = 0.3, 0.5, 1, 2 and 3;

$d_1$ and $d_2$ are adjusted so that $\overline{P_{01}} = \overline{P_{00}}$ and both are fixed at 0.25.

Experiment 7C: $\beta_1 = \beta_2 = [{}^{-d}_{c}]$ and $c = 1.0$;

d is adjusted so that $\overline{P_{01}} + \overline{P_{00}}$ varies from 0.20, 0.30, 0.40, 0.50, 0.60 to 0.70.

Experiment 7D: $\beta_1 = [{}^{-d_1}_{c}]$, $\beta_2 = [{}^{-d_2}_{c}]$ and $c = 1.0$; $d_1$ and $d_2$

are adjusted so that $\overline{P_{01}} = \overline{P_{00}}$ and both vary from 0.10, 0.15, 0.20, 0.30 and 0.35.

These four experiments are designed for case three. Since $\beta_1$ has all the ratios very close to one, only the results for $\beta_2$ are listed and they are in Tables 17, 18, 19 and 20.

(3) Experiment 8A: $\beta_1 = \beta_2 = [{}^{-d}_{c}]$ for c = 0.3, 0.5, 1, 2 and 3;

d is adjusted so that $\overline{P_{00}} = 0.25$.

Experiment 8B: $\beta_1 = \beta_2 = [{}^{-d}_{c}]$ and $c = 1.0$;

d is adjusted so that $\overline{P_{00}}$ varies from 0.15, 0.25, 0.35, 0.45, 0.55 to 0.65.

These are experiments designed for case four and results are in Tables 21 and 22.

We get the following conclusions from the results of these experiments:

From the results of Experiment 6B, 7C, 7D and 8B, it can be seen that sample split effect does affect the relative efficiencies of partial observability cases. The higher the average probability of the indistinguishable outcome for each case, the worse the relative efficiency and higher the cost of partial observability. This result holds under different situations concerning $\rho$. Thus, in Experiment 6B (Table 16), increasing $\overline{P_{11}}$ increases the efficiency of case two relative to case one. In Experiments 7C and 7D (Tables 19 and 20), increasing $\overline{P_{01}} + \overline{P_{00}}$ decreases the efficiency of case three relative to case one. In Experiment 8B (Table 22), increasing $\overline{P_{00}}$ decreases the efficiency of case four relative to case one. All of this is as expected.

Experiments 6A, 7A, 7B and 8A attempt to investigate the identification effect while holding the sample split effect constant. This leads to results that are less clear-cut than those just reported. Basically, identification effects are strong and predictable near points of singularity, but less so far from points of singularity.

Experiment 6A investigates the identification effect for case two. The probability of the indistiguishable event for case two are $1-\overline{P_{11}}$, so we hold the sample split effect constant by holding constant $\overline{P_{11}} = 0.25$. Lack of identification occurs if c=0 regardless of whether $\rho$ is known or not. Therefore the efficiency of case two relative to case one is expected to increase (and the entires in Table 15 to fall) as c increases. The results in Table 15 show that mostly they do. The exceptions occur when c is big (c=3) and $\rho$ is known, which are far from points of singularity.

Experiments 7A and 7B investigate the identification effect for case three. Since the probability of the indistinguishable event for case three is $(\overline{P_{01}} + \overline{P_{00}})$, we attempt to hold the sample split effect constant by holding constant $\overline{P_{01}} + \overline{P_{00}} = \frac{1}{2}$ (Experiment 7A) or $\overline{P_{01}} = \overline{P_{00}} = \frac{1}{4}$ (Experiment 7B). Lack of identification for case three occurs when c=0 and $\rho$ is unknown. Therefore when $\rho$ is unknown we would expect the efficiency of case three relative to case one to rise (and the entires in Table 17 and 18 to fall) as c increases. This does occur as c increases, when c is small, but for c > 1 the opposite occurs. In other words, the identification effect shows up only when the model is close to non-identification. The same phenomenon occurs when $\rho$ is known. Then the parameters are identified for all c, and the relative efficiency for case three falls monotonically except for very small c.

Experiment 8A investigates the identification effect for case four. The probability of the indistinguishable event here is $\overline{P_{00}}$, so the relevant portion of the sample split is $\overline{P_{00}}$ , which we hold constant as c changes. Lack of identification occurs when c=0; for case 4A2 this is so regardless of whether $\rho$ is known, while for case 4A1 and 4B this is so only if $\rho$ is unknown. The results in Table 2.1 are fairly predictable. Wherever identification is relevant (all cases when $\rho$ is unknown, but only case 4A2 when $\rho$ is known) the efficiency of case four relative to case one rises (entires in the table fall) as c increases. For cases where identification is not relevant (cases 4A1 and 4B when $\rho$ is known) relative efficiency first rises and then falls as c increases.

## 4.5 Summary

In this chapter, we have conducted a large variety of experiments to measure the cost of various levels of partial observability. The results have been given in some detail in the preceeding sections. Here we will give a brief summary of the most important conclusions.

The cost of partial observability is quite high. The estimates from Poirier's model (our case two) typically have variances tens or hundreds of times as large as do the estimates from the model with full observability (our case one). This cost decreases markedly if any piece of observability information can be found; for example, observability for either party (our case three) or observed veto (our case four). The law of diminishing marginal utility of information usually holds: the gain in moving from case two to case three or four usually exceeds the gain in moving from case three or four to full observability (our case one). It is the first piece of observability information which is most important.

A second clear conclusion is that specifying $\rho$ a priori improves the efficiency of the estimates of the other parameters a great deal. Furthermore the improvement from knowing $\rho$ is largest when it is needed most; that is, when the relative efficiency is lowest.

A third conclusion is that the sample split has a strong influence on the relative efficiencies of the estimates. For a given partial observability case, its efficiency relative to full observability will be higher, the smaller the proportion of observations which fall into the indistinguishable categories. Thus, for example, for Poirier's model relative efficiency will be high only when most observations are of the "yes, yes" variety. The fraction of such observations is observable. Similarly, in our case three the observations which reduce relative efficiency are the ones for which the observable party votes "no", and the proportion of such observations is also observable. On the other hand, for the observed veto cases the relevant proportion of observations is not directly observable.

Our last main conclusion is that the strength of identification matters. All of the partial observability cases are unidentified for some perverse parameter points (as described in Chapter 3). Their relative efficiency is naturally low for parameter values near such points, and it increases rapidly as the parameters move away from such points of singularity. However, these effects are not strong except in the immediate neighborhood of points of non-identification. Furthermore, this last conclusion depends on unobserved parameters, and therefore is less likely to be informative, in practical applications, than the other three conclusions listed above.

# TABLE 1

Sample Splits for Different $\beta$

| $\beta$ | $\beta_1=\beta_2=\begin{bmatrix}1\\1\end{bmatrix}$ | | $\beta_1=\beta_2=\begin{bmatrix}0\\1\end{bmatrix}$ | | $\beta_1=\begin{bmatrix}1\\-1\end{bmatrix}, \beta_1=\begin{bmatrix}1\\1\end{bmatrix}$ | |
|---|---|---|---|---|---|---|
| $\rho$ | 0.0 | 0.5 | 0.0 | 0.5 | 0.0 | 0.5 |
| $\overline{P_{11}}$ | 0.9238 | 0.9320 | 0.6846 | 0.7243 | 0.4279 | 0.4422 |
| $\overline{P_{10}}$ | 0.0365 | 0.0283 | 0.1297 | 0.0900 | 0.0276 | 0.0133 |
| $\overline{P_{01}}$ | 0.0365 | 0.0283 | 0.1297 | 0.0900 | 0.5325 | 0.5182 |
| $\overline{P_{00}}$ | 0.0031 | 0.0113 | 0.0560 | 0.0957 | 0.0120 | 0.0263 |
| $\overline{H}$ | 0.0016 | 0.0057 | 0.0280 | 0.0478 | 0.0076 | 0.0192 |

TABLE 2

Ratios of Asymptotic Variances (Cost of Partial Observability)

| $\beta$ | | $\beta_1=\beta_2=\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ | | $\beta_1=\beta_2=\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ | | $\beta_1=\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$, $\beta_2=\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$ | |
|---|---|---|---|---|---|---|---|
| $\rho$ | | 0.0 | 0.5 | 0.0 | 0.5 | 0.0 | 0.5 |
| | | | | $\rho$ is not known | | | |
| $\beta_{12}$ | Case 2 | * | * | * | * | 17.6810 | 6.2666 |
| | Case 3 | 1.0000 | 1.0117 | 1.0000 | 1.0119 | 1.0000 | 1.0007 |
| | Case 4A1 | 5.1284 | 28.0120 | 8.7086 | 41.8127 | 4.5844 | 2.8136 |
| | Case 4A2 | 7.0715 | 33.2000 | 10.9572 | 45.7816 | 4.7963 | 2.9075 |
| | Case 4B | 5.1225 | 27.9721 | 8.6645 | 41.6975 | 2.6225 | 1.8931 |
| $\beta_{22}$ | Case 2 | * | * | * | * | 8857.6250 | 35637.07 |
| | Case 3 | 9.3734 | 52.2389 | 17.2976 | 78.4054 | 178.6166 | 864.2910 |
| | Case 4A1 | 5.1284 | 28.0120 | 8.7086 | 41.8127 | 2.4228 | 1.9527 |
| | Case 4A2 | 7.0715 | 33.2000 | 10.9572 | 45.7816 | 46.7191 | 3.8118 |
| | Case 4B | 5.1225 | 27.9706 | 8.6644 | 41.6975 | 19.8626 | 47.1023 |
| | | | | $\rho$ is known | | | |
| $\beta_{12}$ | Case 2 | * | * | * | * | 7.6118 | 3.7920 |
| | Case 3 | 1.0000 | 1.0088 | 1.0000 | 1.0073 | 1.0000 | 1.0008 |
| | Case 4A1 | 1.0255 | 1.0731 | 1.1135 | 1.1755 | 1.0189 | 1.0273 |
| | Case 4A2 | 2.9686 | 6.2613 | 3.3621 | 5.1455 | 4.3046 | 2.0516 |
| | Case 4B | 1.0196 | 1.0332 | 1.0694 | 1.0604 | 1.0103 | 1.0067 |
| $\beta_{22}$ | Case 2 | * | * | * | * | 105.3535 | 191.2173 |
| | Case 3 | 1.0500 | 1.1219 | 1.2118 | 1.2796 | 2.1731 | 5.0474 |
| | Case 4A1 | 1.0255 | 1.0731 | 1.1135 | 1.1755 | 1.3622 | 1.6642 |
| | Case 4A2 | 2.9686 | 6.2613 | 3.3621 | 5.1455 | 10.1104 | 3.7010 |
| | Case 4B | 1.0196 | 1.0331 | 1.0694 | 1.0604 | 1.6767 | 3.0236 |

*The information matrix of case two is singular (parameters are not identified).

TABLE 3

Sample Splits for Different c

When $\beta_1 = \beta_2 = [-c \; c \; \bar{X}]$

$\rho$=0.0

| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|
| $\overline{P_{11}}$ | 0.2491 | 0.2486 | 0.2542 | 0.2629 | 0.2811 |
| $\overline{P_{10}}$ | 0.2297 | 0.2076 | 0.1464 | 0.0993 | 0.0450 |
| $\overline{P_{01}}$ | 0.2297 | 0.2076 | 0.1464 | 0.0993 | 0.0450 |
| $\overline{P_{00}}$ | 0.2916 | 0.3362 | 0.4529 | 0.5385 | 0.6288 |
| $\overline{P_{01}}+\overline{P_{00}}$ | 0.5213 | 0.5438 | 0.5993 | 0.6378 | 0.6838 |
| $\overline{H}$ | 0.1458 | 0.1681 | 0.2265 | 0.2693 | 0.3144 |

$\rho$=0.5

| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|
| $\overline{P_{11}}$ | 0.3244 | 0.3151 | 0.2982 | 0.2913 | 0.2944 |
| $\overline{P_{10}}$ | 0.1544 | 0.1411 | 0.1024 | 0.0709 | 0.0317 |
| $\overline{P_{01}}$ | 0.1544 | 0.1411 | 0.1024 | 0.0709 | 0.0317 |
| $\overline{P_{00}}$ | 0.3669 | 0.4027 | 0.4970 | 0.5670 | 0.6422 |
| $\overline{P_{01}}+\overline{P_{00}}$ | 0.5213 | 0.5438 | 0.5994 | 0.6379 | 0.6739 |
| $\overline{H}$ | 0.1834 | 0.2013 | 0.2485 | 0.2835 | 0.3211 |

## TABLE 4

Ratios of Asymptotic Variances (Cost of Partial Observability)
When $\beta_1=\beta_2=[^{-c}{}_{c}{}^{\bar{X}}]$ and $\rho=0$

| | c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|---|
| | | | $\rho$ is not known | | | |
| $\beta_{12}$ | Case 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Case 4A1 | 91.2821 | 50.9838 | 19.9076 | 23.0626 | 18.5827 |
| | Case 4A2 | 103.0591 | 56.0029 | 22.1733 | 23.8038 | 18.8309 |
| | Case 4B | 91.1068 | 50.7766 | 19.4382 | 21.7407 | 16.8380 |
| $\beta_{22}$ | Case 3 | 205.0555 | 121.5853 | 42.5952 | 52.0272 | 37.4875 |
| | Case 4A1 | 91.2821 | 50.9838 | 19.9076 | 23.0626 | 18.5827 |
| | Case 4A2 | 103.0591 | 56.0029 | 22.1733 | 23.8038 | 18.8309 |
| | Case 4B | 91.1063 | 50.7766 | 19.4382 | 21.7407 | 16.8380 |
| | | | $\rho$ is known | | | |
| $\beta_{12}$ | Case 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | Case 4A1 | 1.3747 | 1.4050 | 1.7985 | 3.0932 | 3.6511 |
| | Case 4A2 | 13.1517 | 6.4240 | 4.0643 | 3.8345 | 3.8994 |
| | Case 4B | 1.1991 | 1.1977 | 1.3291 | 1.7713 | 1.9064 |
| $\beta_{22}$ | Case 3 | 1.7174 | 1.7369 | 2.3280 | 4.0159 | 4.6467 |
| | Case 4A1 | 1.3747 | 1.4050 | 1.7985 | 3.0932 | 3.6511 |
| | Case 4A2 | 13.1517 | 6.4240 | 4.0643 | 3.8345 | 3.8994 |
| | Case 4B | 1.1991 | 1.1977 | 1.3291 | 1.7713 | 1.9064 |

*The information matrix of case two is singular in this experiment.

## TABLE 5

Ratios of Asymptotic Variances (Cost of Partial Observability)
When $\beta_1=\beta_2= [^{-c}\ _{c}^{\overline{X}}]$ and $\rho=0.5$

| | c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|---|
| | | | | $\rho$ is not known | | |
| $\beta_{12}$ | Case 3 | 1.0008 | 1.0038 | 1.0080 | 1.0095 | 1.0119 |
| | Case 4A1 | 267.7677 | 207.5012 | 63.1785 | 61.4412 | 43.9869 |
| | Case 4A2 | 281.0156 | 213.1307 | 65.5496 | 62.1682 | 44.2313 |
| | Case 4B | 267.4762 | 207.1903 | 62.6215 | 60.1184 | 42.3013 |
| $\beta_{22}$ | Case 3 | 527.4460 | 402.6952 | 125.7184 | 129.2031 | 91.1600 |
| | Case 4A1 | 267.7677 | 207.5012 | 63.1785 | 61.4412 | 43.9869 |
| | Case 4A2 | 281.0156 | 213.1307 | 65.5496 | 62.1682 | 44.2313 |
| | Case 4B | 267.4762 | 207.1903 | 62.6215 | 60.1184 | 42.3013 |
| | | | | $\rho$ is known | | |
| $\beta_{12}$ | Case 3 | 1.0016 | 1.0019 | 1.0046 | 1.0064 | 1.0074 |
| | Case 4A1 | 1.4345 | 1.4460 | 1.7343 | 2.6848 | 3.1727 |
| | Case 4A2 | 14.6908 | 7.0827 | 4.1108 | 3.4136 | 3.4176 |
| | Case 4B | 1.1428 | 1.1351 | 1.1760 | 1.3589 | 1.4840 |
| $\beta_{22}$ | Case 3 | 1.6862 | 1.6701 | 1.9542 | 2.8662 | 3.3950 |
| | Case 4A1 | 1.4345 | 1.4460 | 1.7343 | 2.6848 | 3.1727 |
| | Case 4A2 | 14.6908 | 7.0827 | 4.1108 | 3.4136 | 3.4176 |
| | Case 4B | 1.1428 | 1.1351 | 1.1760 | 1.3589 | 1.4840 |

*The information matrix of case two is singular in this experiment.

TABLE 6

Sample Splits for Different c When $\beta_1 = [c\ \bar{X}_{-c}]$ and $\beta_2 = [-c\ \bar{X}_{c}]$

| $\rho=0.0$ | | | | | |
|---|---|---|---|---|---|
| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| $\overline{P_{11}}$ | 0.2297 | 0.2075 | 0.1464 | 0.0704 | 0.0450 |
| $\overline{P_{10}}$ | 0.2916 | 0.3362 | 0.4529 | 0.5881 | 0.6288 |
| $\overline{P_{01}}$ | 0.2491 | 0.2486 | 0.2542 | 0.2712 | 0.2811 |
| $\overline{P_{00}}$ | 0.2297 | 0.2076 | 0.1464 | 0.0704 | 0.0450 |
| $\overline{P_{01}}+\overline{P_{00}}$ | 0.4788 | 0.4562 | 0.4006 | 0.3416 | 0.3261 |
| $\overline{H}$ | 0.1079 | 0.0914 | 0.0566 | 0.0282 | 0.0205 |
| $\rho=0.5$ | | | | | |
| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| $\overline{P_{11}}$ | 0.3011 | 0.2667 | 0.1800 | 0.0856 | 0.0558 |
| $\overline{P_{10}}$ | 0.2202 | 0.2771 | 0.4194 | 0.5729 | 0.6181 |
| $\overline{P_{01}}$ | 0.1777 | 0.1895 | 0.2206 | 0.2560 | 0.2704 |
| $\overline{P_{00}}$ | 0.3011 | 0.2667 | 0.1800 | 0.0856 | 0.0558 |
| $\overline{P_{01}}+\overline{P_{00}}$ | 0.4788 | 0.4562 | 0.4006 | 0.3416 | 0.3262 |
| $\overline{H}$ | 0.1334 | 0.1130 | 0.0605 | 0.0332 | 0.0255 |

## TABLE 7

Ratios of Asymptotic Variances (Cost of Partial Observability)
When $\beta_1 = \begin{bmatrix} c & \bar{X} \\ & -c \end{bmatrix}$, $\beta_2 = \begin{bmatrix} -c & \bar{X} \\ & c \end{bmatrix}$ and $\rho=0$

| | c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|---|
| | | | $\rho$ is not known | | | |
| $\beta_{12}$ | Case 2 | 605.5864 | 518.8082 | 151.5306 | 221.9935 | 75.3350 |
| | Case 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0002 |
| | Case 4A1 | 1.7312 | 1.5502 | 1.3647 | 1.5126 | 1.6559 |
| | Case 4A2 | 104.7621 | 57.8441 | 10.8805 | 4.8029 | 3.8208 |
| | Case 4B | 48.9987 | 22.8783 | 13.1748 | 11.8071 | 10.9172 |
| $\beta_{22}$ | Case 2 | 21751.885 | 5910.3139 | 659.7719 | 78.4286 | 42.9161 |
| | Case 3 | 890.1655 | 411.7187 | 119.8679 | 30.1673 | 19.4758 |
| | Case 4A1 | 1.8019 | 1.6830 | 1.6583 | 1.7880 | 1.8296 |
| | Case 4A2 | 91.1751 | 48.8784 | 12.1145 | 7.1413 | 5.9189 |
| | Case 4B | 64.2327 | 30.7263 | 11.6161 | 6.0176 | 5.0118 |
| | | | $\rho$ is known | | | |
| $\beta_{12}$ | Case 2 | 365.3644 | 313.1222 | 44.8365 | 32.2472 | 22.9143 |
| | Case 3 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0002 |
| | Case 4A1 | 1.3767 | 1.3311 | 1.3634 | 1.4480 | 1.5318 |
| | Case 4A2 | 81.4734 | 36.9280 | 10.7038 | 4.5243 | 3.4053 |
| | Case 4B | 1.1951 | 1.2293 | 1.5222 | 2.5273 | 3.0136 |
| $\beta_{22}$ | Case 2 | 589.6485 | 529.1590 | 59.2528 | 22.1721 | 12.8180 |
| | Case 3 | 2.7547 | 3.2830 | 3.1431 | 2.5740 | 2.1090 |
| | Case 4A1 | 1.6010 | 1.6209 | 1.5473 | 1.4188 | 1.3331 |
| | Case 4A2 | 71.8316 | 32.5675 | 12.0924 | 6.2325 | 4.7055 |
| | Case 4B | 1.5562 | 1.8215 | 1.9029 | 1.7929 | 1.5520 |

TABLE 8

Ratios of Asymptotic Variances (Cost of Partial Observability) When $\beta_1 = [c\ \bar{X}\ -c]$, $\beta_2 = [-c\ \bar{X}\ c]$ and $\rho=0.5$

| | c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
|---|---|---|---|---|---|---|
| | **$\rho$ is not known** | | | | | |
| $\beta_{12}$ | Case 2 | 314.6830 | 312.4838 | 35.6600 | 48.1634 | 27.8839 |
| | Case 3 | 1.0042 | 1.0078 | 1.0089 | 1.0081 | 1.0075 |
| | Case 4A1 | 1.7337 | 1.4753 | 1.3893 | 1.5222 | 1.6044 |
| | Case 4A2 | 54.1470 | 18.4067 | 5.1199 | 2.6099 | 2.5202 |
| | Case 4B | 19.8579 | 10.2136 | 8.1579 | 8.5622 | 8.4649 |
| $\beta_{22}$ | Case 2 | 4389.5386 | 2022.1393 | 82.0819 | 18.1384 | 14.0965 |
| | Case 3 | 344.3801 | 176.7509 | 36.3043 | 11.6794 | 7.9632 |
| | Case 4A1 | 1.8750 | 1.7257 | 1.8023 | 1.7998 | 1.7341 |
| | Case 4A2 | 45.6778 | 15.6901 | 6.3376 | 3.9413 | 3.7014 |
| | Case 4B | 30.8417 | 15.6040 | 6.6949 | 3.9136 | 3.2999 |
| | **$\rho$ is known** | | | | | |
| $\beta_{12}$ | Case 2 | 154.3449 | 47.5010 | 18.1745 | 15.2464 | 14.2868 |
| | Case 3 | 1.0049 | 1.0086 | 1.0076 | 1.0047 | 1.0034 |
| | Case 4A1 | 1.3572 | 1.3378 | 1.3793 | 1.4578 | 1.5356 |
| | Case 4A2 | 35.7942 | 12.6180 | 5.1139 | 2.5252 | 2.3125 |
| | Case 4B | 1.1593 | 1.2532 | 1.8462 | 3.6369 | 3.8029 |
| $\beta_{22}$ | Case 2 | 297.7500 | 105.2685 | 20.4917 | 8.7703 | 5.8623 |
| | Case 3 | 3.1706 | 4.0090 | 3.6484 | 2.7647 | 2.3078 |
| | Case 4A1 | 1.7264 | 1.7192 | 1.6060 | 1.4609 | 1.3779 |
| | Case 4A2 | 31.4848 | 11.8300 | 6.1416 | 3.5386 | 2.9358 |
| | Case 4B | 1.6245 | 2.0532 | 2.1276 | 1.9355 | 1.6636 |

## TABLE 9

Sample Splits for Different $\rho$
When $\beta_1=\beta_2= [{}^{-c}{}_{c}{}^{\bar{X}}]$ and c=1.0

| $\rho$ | -0.5 | 0.0 | 0.2 | 0.5 | 0.8 |
|---|---|---|---|---|---|
| $\overline{P_{11}}$ | 0.2206 | 0.2542 | 0.2701 | 0.2982 | 0.3363 |
| $\overline{P_{10}}$ | 0.1800 | 0.1464 | 0.1305 | 0.1024 | 0.0643 |
| $\overline{P_{01}}$ | 0.1800 | 0.1464 | 0.1305 | 0.1024 | 0.0643 |
| $\overline{P_{00}}$ | 0.4194 | 0.4970 | 0.4689 | 0.4970 | 0.5351 |
| $\overline{H}$ | 0.2097 | 0.2485 | 0.2345 | 0.2485 | 0.2676 |

## TABLE 10

Sample Splits for Different $\rho$
When $\beta_1= [{}^{c}{}_{-c}{}^{\bar{X}}]$, $\beta_2= [{}^{-c}{}_{c}{}^{\bar{X}}]$ and c=1.0

| $\rho$ | -0.5 | 0.0 | 0.2 | 0.5 | 0.8 |
|---|---|---|---|---|---|
| $\overline{P_{11}}$ | 0.1024 | 0.1464 | 0.1609 | 0.1800 | 0.1961 |
| $\overline{P_{10}}$ | 0.4970 | 0.4529 | 0.4385 | 0.4194 | 0.4032 |
| $\overline{P_{01}}$ | 0.2982 | 0.2542 | 0.2398 | 0.2206 | 0.2045 |
| $\overline{P_{00}}$ | 0.1024 | 0.1464 | 0.1609 | 0.1800 | 0.1961 |
| $\overline{H}$ | 0.0441 | 0.0566 | 0.0590 | 0.0605 | 0.0603 |

## TABLE 11

Ratios of Asymptotic Variances (Cost of Partial Observability)
When $\beta_1=\beta_2 \left[\begin{smallmatrix} -c & \bar{X} \\ & c \end{smallmatrix}\right]$ and c=1.0

| | ρ | -0.5 | 0.0 | 0.2 | 0.5 | 0.8 |
|---|---|---|---|---|---|---|
| | **ρ is not known** | | | | | |
| $\beta_{12}$ | Case 3 | 1.0088 | 1.0000 | 1.0014 | 1.0080 | 1.0166 |
| | Case 4A1 | 9.5192 | 19.9076 | 29.4137 | 63.1785 | 225.4031 |
| | Case 4A2 | 11.5298 | 22.1733 | 31.7488 | 65.5496 | 227.6452 |
| | Case 4B | 9.1134 | 19.4382 | 28.9126 | 62.6215 | 224.7830 |
| $\beta_{22}$ | Case 3 | 20.0543 | 42.5952 | 62.4095 | 125.7184 | 351.5529 |
| | Case 4A1 | 9.5192 | 19.9076 | 29.4137 | 63.1785 | 225.4031 |
| | Case 4A2 | 11.5298 | 22.1733 | 31.7488 | 65.5496 | 227.6452 |
| | Case 4B | 9.1134 | 19.4382 | 28.9126 | 62.6215 | 224.7830 |
| | **ρ is known** | | | | | |
| $\beta_{12}$ | Case 3 | 1.0042 | 1.0000 | 1.0007 | 1.0046 | 1.0106 |
| | Case 4A1 | 2.0485 | 1.7985 | 1.7632 | 1.7343 | 1.7078 |
| | Case 4A2 | 4.0619 | 4.0643 | 4.0991 | 4.1108 | 3.9623 |
| | Case 4B | 1.6420 | 1.3291 | 1.2619 | 1.1760 | 1.0843 |
| $\beta_{22}$ | Case 3 | 3.1039 | 2.3280 | 2.1651 | 1.9542 | 1.6801 |
| | Case 4A1 | 2.0485 | 1.7985 | 1.7632 | 1.7343 | 1.7078 |
| | Case 4A2 | 4.0619 | 4.0643 | 4.0991 | 4.1108 | 3.9623 |
| | Case 4B | 1.6420 | 1.3291 | 1.2619 | 1.1760 | 1.0843 |

*The information matrix of case two is singular in this experiment.

## TABLE 12

Ratios of Asymptotic Variances (Cost of Partial Observability)
When $\beta_1 = \begin{bmatrix} c & \bar{X} \\ & -c \end{bmatrix}$, $\beta_2 = \begin{bmatrix} -c & \bar{X} \\ & c \end{bmatrix}$ and c=1.0

| | $\rho$ | -0.5 | 0.0 | 0.2 | 0.5 | 0.8 |
|---|---|---|---|---|---|---|
| | | | $\rho$ is not known | | | |
| $\beta_{12}$ | Case 2 | 2358.6990 | 151.5306 | 74.2684 | 35.6600 | 20.9822 |
| | Case 3 | 1.0079 | 1.0000 | 1.0014 | 1.0089 | 1.0232 |
| | Case 4A1 | 1.3347 | 1.3647 | 1.3735 | 1.3893 | 1.4071 |
| | Case 4A2 | 34.7615 | 10.8805 | 7.8550 | 5.1199 | 3.2795 |
| | Case 4B | 25.1054 | 13.1748 | 10.7202 | 8.1579 | 6.3586 |
| $\beta_{22}$ | Case 2 | 7560.6707 | 659.7719 | 288.2825 | 82.0819 | 21.5540 |
| | Case 3 | 323.2625 | 119.8679 | 77.3109 | 36.3043 | 14.5077 |
| | Case 4A1 | 1.5246 | 1.6583 | 1.7142 | 1.8023 | 1.8533 |
| | Case 4A2 | 35.6887 | 12.1145 | 9.1256 | 6.3376 | 4.2237 |
| | Case 4B | 23.2502 | 11.6161 | 9.2417 | 6.6949 | 4.8725 |
| | | | $\rho$ is known | | | |
| $\beta_{12}$ | Case 2 | 193.8613 | 44.8365 | 30.0131 | 18.1745 | 11.3890 |
| | Case 3 | 1.0085 | 1.0000 | 1.0014 | 1.0076 | 1.0168 |
| | Case 4A1 | 1.3369 | 1.3634 | 1.3692 | 1.3793 | 1.3910 |
| | Case 4A2 | 31.9380 | 10.7038 | 7.8430 | 5.1139 | 3.2787 |
| | Case 4B | 1.3655 | 1.5222 | 1.6122 | 1.8462 | 2.6237 |
| $\beta_{22}$ | Case 2 | 252.0110 | 59.2528 | 38.0065 | 22.1721 | 11.0502 |
| | Case 3 | 2.5817 | 3.1431 | 3.3396 | 3.6484 | 3.8425 |
| | Case 4A1 | 1.4760 | 1.5473 | 1.5679 | 1.6060 | 1.6596 |
| | Case 4A2 | 33.5898 | 12.0924 | 9.0963 | 6.1416 | 4.0939 |
| | Case 4B | 1.6567 | 1.9029 | 1.9851 | 2.1276 | 2.4247 |

TABLE 13

Cost of Partial Observability for Case Four

When $\beta_1=\beta_2=\begin{bmatrix}-\bar{X}\\1.0\end{bmatrix}$

| p | | 0.0 | 0.2 | 0.5 | 0.8 | 1.0 |
|---|---|---|---|---|---|---|
| ρ=0.0 but isn't known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 42.5952 | 37.9504 | 19.9076 | 3.9124 | 1.0000 |
| | Case 4A2 | 44.5322 | 38.0962 | 22.1733 | 8.3578 | 3.5517 |
| $\beta_{22}$ | Case 4A1 | 1.0000 | 3.9124 | 19.9076 | 37.9504 | 42.5952 |
| | Case 4A2 | 3.5517 | 8.3578 | 22.1733 | 38.0962 | 44.5322 |
| ρ=0.0 and is known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 2.3280 | 2.1301 | 1.7985 | 1.3913 | 1.0000 |
| | Case 4A2 | 3.3388 | 3.8546 | 4.0643 | 3.8814 | 3.2158 |
| $\beta_{22}$ | Case 4A1 | 1.0000 | 1.3913 | 1.7985 | 2.1301 | 2.3280 |
| | Case 4A2 | 3.2158 | 3.8814 | 4.0643 | 3.8546 | 3.3388 |
| ρ=0.5 but isn't known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 125.7184 | 120.5843 | 63.1785 | 9.2338 | 1.0080 |
| | Case 4A2 | 147.9888 | 127.0938 | 65.5496 | 16.4543 | 3.3842 |
| $\beta_{22}$ | Case 4A1 | 1.0080 | 9.2338 | 63.1785 | 120.5843 | 125.7184 |
| | Case 4A2 | 3.3842 | 16.4543 | 65.5496 | 127.0938 | 147.9888 |
| ρ=0.5 and is known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 1.9542 | 1.9082 | 1.7343 | 1.4090 | 1.0046 |
| | Case 4A2 | 3.0465 | 3.8067 | 4.1108 | 3.8258 | 2.9611 |
| $\beta_{22}$ | Case 4A1 | 1.0046 | 1.4090 | 1.7343 | 1.9082 | 1.9542 |
| | Case 4A2 | 2.9611 | 3.8258 | 4.1108 | 3.8067 | 3.0465 |

## TABLE 14

Cost of Partial Observability for Case Four When $\beta_1 = [\overset{\bar{X}}{-1.0}]$, $\beta_2 = [\overset{\bar{X}}{1.0}]$

| | p | 0.0 | 0.2 | 0.5 | 0.8 | 1.0 |
|---|---|---|---|---|---|---|
| ρ=0.0 but isn't known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 42.5952 | 5.3042 | 1.3647 | 1.3192 | 1.0000 |
| | Case 4A2 | 71.7093 | 37.5250 | 10.8805 | 3.2471 | 1.0002 |
| $\beta_{22}$ | Case 4A1 | 1.0000 | 1.3031 | 1.6583 | 7.9127 | 119.8679 |
| | Case 4A2 | 1.0008 | 3.6176 | 12.1145 | 40.9526 | 154.8729 |
| ρ=0.0 and is known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 2.3280 | 1.7456 | 1.3634 | 1.1418 | 1.0000 |
| | Case 4A2 | 7.6646 | 14.1330 | 10.7038 | 2.2691 | 1.0001 |
| $\beta_{22}$ | Case 4A1 | 1.0000 | 1.1850 | 1.5473 | 2.2172 | 3.1431 |
| | Case 4A2 | 1.0002 | 2.2614 | 12.0924 | 19.1311 | 5.3093 |
| ρ=0.5 but isn't known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 20.0543 | 2.4665 | 1.3893 | 1.1930 | 1.0089 |
| | Case 4A2 | 23.2379 | 14.3879 | 5.1199 | 1.8668 | 1.0089 |
| $\beta_{22}$ | Case 4A1 | 1.0088 | 1.1927 | 1.8023 | 5.8469 | 36.3043 |
| | Case 4A2 | 1.0089 | 2.0992 | 6.3376 | 18.5893 | 36.4342 |
| ρ=0.5 and is known a priori | | | | | | |
| $\beta_{12}$ | Case 4A1 | 3.1039 | 1.9304 | 1.3793 | 1.1413 | 1.0076 |
| | Case 4A2 | 4.5041 | 10.5578 | 5.1139 | 1.6483 | 1.0076 |
| $\beta_{22}$ | Case 4A1 | 1.0025 | 1.1938 | 1.6060 | 2.4451 | 3.6484 |
| | Case 4A2 | 1.0061 | 1.9620 | 6.1416 | 10.5167 | 3.6696 |

TABLE 15

Cost of Partial Observability for Case Two
When $\beta_1 = [{d \atop -c}]$, $\beta_2 = [{d \atop c}]$ and $\overline{P_{11}}=0.25$

| | | | | | |
|---|---|---|---|---|---|
| $\rho=0.0$ | | | | | |
| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| d | 0.0957 | 0.1963 | 0.4508 | 0.9458 | 1.4487 |
| | | | $\rho$ is not known | | |
| $\beta_{12}$ | 1486.7109 | 618.7743 | 48.6514 | 13.8296 | 6.1085 |
| $\beta_{22}$ | 121893.56 | 52071.731 | 36717.786 | 27168.080 | 18131.267 |
| | | | $\rho$ is known | | |
| $\beta_{12}$ | 552.7849 | 75.3081 | 17.4288 | 5.6840 | 3.7545 |
| $\beta_{22}$ | 2524.1335 | 728.7705 | 256.3893 | 146.7752 | 270.6402 |
| $\rho=0.5$ | | | | | |
| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| d | -0.0857 | 0.0447 | 0.3535 | 0.9138 | 1.4409 |
| | | | $\rho$ is not known | | |
| $\beta_{12}$ | 680.0607 | 35.2165 | 11.3308 | 4.3019 | 3.2452 |
| $\beta_{22}$ | 327797.83 | 159416.21 | 117079.78 | 97378.301 | 62839.36 |
| | | | $\rho$ is known | | |
| $\beta_{12}$ | 103.4635 | 15.2357 | 6.2589 | 3.1468 | 2.1256 |
| $\beta_{22}$ | 1536.1228 | 641.0490 | 448.0437 | 459.6155 | 1106.0470 |

TABLE 16

Cost of Partial Observability for Case Two
When $\beta_1 = \begin{bmatrix} d \\ -c \end{bmatrix}$, $\beta_2 = \begin{bmatrix} d \\ c \end{bmatrix}$ and c=1.0

| | | | | | |
|---|---|---|---|---|---|
| $\rho$=0.0 | | | | | |
| $\overline{P_{11}}$ | 0.15 | 0.35 | 0.45 | 0.55 | 0.65 |
| d | 0.0877 | 0.7636 | 1.0680 | 1.3908 | 1.7623 |
| | | | $\rho$ is not known | | |
| $\beta_{12}$ | 108.2441 | 26.6034 | 15.8334 | 9.7708 | 6.0623 |
| $\beta_{22}$ | 111910.395 | 15725.367 | 7589.4556 | 3886.9336 | 2047.7613 |
| | | | $\rho$ is known | | |
| $\beta_{12}$ | 33.9630 | 10.6121 | 6.8617 | 4.7238 | 3.2380 |
| $\beta_{22}$ | 570.5558 | 148.0751 | 96.5562 | 67.6946 | 49.9944 |
| $\rho$=0.5 | | | | | |
| $\overline{P_{11}}$ | 0.15 | 0.35 | 0.45 | 0.55 | 0.65 |
| d | -0.0543 | 0.6977 | 1.0258 | 1.3666 | 1.7510 |
| | | | $\rho$ is not known | | |
| $\beta_{12}$ | 17.7333 | 8.1456 | 6.1336 | 4.6763 | 3.8507 |
| $\beta_{22}$ | 273997.19 | 60851.628 | 34097.547 | 19487.394 | 10932.74 |
| | | | $\rho$ is known | | |
| $\beta_{12}$ | 9.2281 | 4.7224 | 3.7254 | 2.9860 | 2.3859 |
| $\beta_{22}$ | 864.2748 | 277.7323 | 185.5475 | 127.9618 | 89.0390 |

TABLE 17

Cost of Partial Observability for Case Three
When $\beta_1=\beta_2=\begin{bmatrix}-d\\c\end{bmatrix}$ and $\overline{P_{01}}+\overline{P_{00}}=0.50$

| | | | | | |
|---|---|---|---|---|---|
| $\rho=0.0$ | | | | | |
| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| d | 0.4003 | 0.6357 | 1.1557 | 2.0572 | 2.9064 |
| | | | $\rho$ is not known | | |
| $\beta_{22}$ | 206.2943 | 94.5461 | 37.6200 | 43.0889 | 48.6845 |
| | | | $\rho$ is known | | |
| $\beta_{22}$ | 1.6581 | 1.6153 | 1.8644 | 2.5978 | 2.9655 |
| $\rho=0.5$ | | | | | |
| c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| d | 0.4005 | 0.6356 | 1.1557 | 2.0578 | 2.9064 |
| | | | $\rho$ is not known | | |
| $\beta_{22}$ | 554.9503 | 340.8961 | 117.2571 | 115.8412 | 125.9888 |
| | | | $\rho$ is known | | |
| $\beta_{22}$ | 1.6471 | 1.5918 | 1.7113 | 2.1069 | 2.3266 |

TABLE 18

Cost of Partial Observability for Case Three
When $\beta_1 = \begin{bmatrix} -d_1 \\ c \end{bmatrix}$, $\beta_2 = \begin{bmatrix} -d_2 \\ c \end{bmatrix}$ and $\overline{P_{01}} = \overline{P_{00}} = 0.25$

| | | | | | |
|---|---|---|---|---|---|
| $\rho=0.0$ | | | | | |
| $c$ | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| $d_1$ | 0.4003 | 0.6357 | 1.1557 | 2.0578 | 2.9064 |
| $d_2$ | 0.3008 | 0.4323 | 0.6820 | 1.1212 | 1.5651 |
| $\rho$ is not known | | | | | |
| $\beta_{22}$ | 234.5919 | 96.6320 | 58.0045 | 119.1044 | 217.5330 |
| $\rho$ is known | | | | | |
| $\beta_{22}$ | 1.6596 | 1.6206 | 1.9260 | 3.1437 | 6.2018 |
| $\rho=0.5$ | | | | | |
| $c$ | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| $d_1$ | 0.4005 | 0.6357 | 1.1557 | 2.0578 | 2.9064 |
| $d_2$ | -0.0622 | 0.1009 | 0.4250 | 0.9644 | 1.4679 |
| $\rho$ is not known | | | | | |
| $\beta_{22}$ | 1623.2131 | 706.0171 | 503.8493 | 1265.8145 | 3283.1157 |
| $\rho$ is known | | | | | |
| $\beta_{22}$ | 1.9167 | 1.9058 | 2.3351 | 4.1795 | 9.2418 |

TABLE 19

Cost of Partial Observability for Case Three
When $\beta_1=\beta_2=\begin{bmatrix}-d\\c\end{bmatrix}$ and c=1.0

| $\rho=0.0$ | | | | | |
|---|---|---|---|---|---|
| $\overline{P_{01}}+\overline{P_{00}}$ | 0.20 | 0.30 | 0.40 | 0.60 | 0.70 |
| d | 0.0630 | 0.4564 | 0.8084 | 1.5275 | 1.9623 |
| | | | $\rho$ is not known | | |
| $\beta_{22}$ | 18.1194 | 24.4290 | 31.3962 | 42.6311 | 54.2505 |
| | | | $\rho$ is known | | |
| $\beta_{22}$ | 1.2294 | 1.3728 | 1.5702 | 2.3317 | 3.0872 |
| $\rho=0.5$ | | | | | |
| $\overline{P_{01}}+\overline{P_{00}}$ | 0.20 | 0.30 | 0.40 | 0.60 | 0.70 |
| d | 0.0630 | 0.4564 | 0.8084 | 1.5275 | 1.9623 |
| | | | $\rho$ is not known | | |
| $\beta_{22}$ | 80.7077 | 95.9903 | 108.5038 | 125.7936 | 149.8488 |
| | | | $\rho$ is known | | |
| $\beta_{22}$ | 1.2941 | 1.4034 | 1.5364 | 1.9560 | 2.3189 |

TABLE 20

Cost of Partial Observability for Case Three
When $\beta_1 = [{-d_1 \atop c}]$, $\beta_2 = [{-d_2 \atop c}]$ and c=1.0

| | | | | | |
|---|---|---|---|---|---|
| $\rho$=0.0 | | | | | |
| $\overline{P_{01}}=\overline{P_{00}}$ | 0.10 | 0.15 | 0.20 | 0.30 | 0.35 |
| $d_1$ | 0.0630 | 0.4564 | 0.8084 | 1.5275 | 1.9623 |
| $d_2$ | 0.5546 | 0.5962 | 0.6376 | 0.7329 | 0.7951 |
| | | | $\rho$ is not known | | |
| $\beta_{22}$ | 14.1692 | 22.4785 | 35.5819 | 100.6174 | 196.7489 |
| | | | $\rho$ is known | | |
| $\beta_{22}$ | 1.2093 | 1.3642 | 1.5850 | 2.5186 | 3.7620 |
| $\rho$=0.5 | | | | | |
| $\overline{P_{01}}=\overline{P_{00}}$ | 0.10 | 0.15 | 0.20 | 0.30 | 0.35 |
| $d_1$ | 0.0630 | 0.4564 | 0.8084 | 1.5275 | 1.9623 |
| $d_2$ | -0.0085 | 0.1564 | 0.2961 | 0.5513 | 0.6817 |
| | | | $\rho$ is not known | | |
| $\beta_{22}$ | 89.9929 | 158.1305 | 276.4605 | 1014.1426 | 2513.5569 |
| | | | $\rho$ is known | | |
| $\beta_{22}$ | 1.3155 | 1.5315 | 1.8423 | 3.2329 | 5.2786 |

<u>TABLE 21</u>

Cost of Partial Observability for Case Four
When $\beta_1=\beta_2=\begin{bmatrix}-d\\c\end{bmatrix}$, p=0.5 and $\overline{P_{00}}$=0.25

| | | | | | | |
|---|---|---|---|---|---|---|
| ρ=0.0 | | | | | | |
| | c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| | d | 0.3497 | 0.5294 | 0.8933 | 1.4822 | 2.0127 |
| ρ is not known | | | | | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 89.5407 | 34.2244 | 15.5636 | 14.8929 | 13.2041 |
| | Case 4A2 | 100.9828 | 38.8355 | 18.0756 | 16.4319 | 14.3767 |
| | Case 4B | 89.3978 | 34.0865 | 15.3714 | 14.5476 | 12.6529 |
| ρ is known | | | | | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 1.3176 | 1.2881 | 1.3605 | 1.5742 | 1.8609 |
| | Case 4A2 | 12.7598 | 5.8992 | 3.8724 | 3.1132 | 3.0335 |
| | Case 4B | 1.1745 | 1.1502 | 1.1682 | 1.2289 | 1.3097 |
| ρ=0.5 | | | | | | |
| | c | 0.3 | 0.5 | 1.0 | 2.0 | 3.0 |
| | d | 0.1338 | 0.3202 | 0.6993 | 1.3050 | 1.8420 |
| ρ is not known | | | | | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 312.4136 | 125.3176 | 54.4721 | 46.5124 | 40.1396 |
| | Case 4A2 | 326.3303 | 131.2031 | 57.7210 | 48.3723 | 41.5624 |
| | Case 4B | 312.2180 | 125.1313 | 54.2374 | 46.1354 | 39.5681 |
| ρ is known | | | | | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 1.3013 | 1.2777 | 1.3297 | 1.4946 | 1.7272 |
| | Case 4A2 | 15.2242 | 7.1676 | 4.5824 | 3.3578 | 3.1530 |
| | Case 4B | 1.1058 | 1.0913 | 1.0947 | 1.1170 | 1.1559 |

TABLE 22

Cost of Partial Observability for Case Four

When $\beta_1=\beta_2=\begin{bmatrix}-d\\c\end{bmatrix}$, p=0.5 and c=1.0

| | | | | | | |
|---|---|---|---|---|---|---|
| $\rho$=0.0 | | | | | | |
| | $\overline{P_{00}}$ | 0.15 | 0.35 | 0.45 | 0.55 | 0.65 |
| | d | 0.5452 | 1.2098 | 1.5158 | 1.8394 | 2.2113 |
| | | | | $\rho$ is not known | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 12.3757 | 17.9619 | 19.8483 | 22.7009 | 31.2475 |
| | Case 4A2 | 14.8237 | 20.4170 | 22.1214 | 24.6609 | 32.8082 |
| | Case 4B | 12.2678 | 17.6562 | 19.3845 | 22.0101 | 30.2299 |
| | | | | $\rho$ is known | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 1.2250 | 1.5389 | 1.7896 | 2.1554 | 2.6895 |
| | Case 4A2 | 3.6730 | 3.9940 | 4.0627 | 4.1154 | 4.2228 |
| | Case 4B | 1.1171 | 1.2333 | 1.3258 | 1.4646 | 1.6718 |
| $\rho$=0.5 | | | | | | |
| | $\overline{P_{00}}$ | 0.15 | 0.35 | 0.45 | 0.55 | 0.65 |
| | d | 0.2919 | 1.0422 | 1.2098 | 1.7083 | 2.0917 |
| | | | | $\rho$ is not known | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 47.3035 | 58.5931 | 60.0426 | 66.2930 | 80.0371 |
| | Case 4A2 | 50.9623 | 61.4985 | 62.7722 | 68.4357 | 81.6770 |
| | Case 4B | 47.1487 | 58.2580 | 59.6427 | 65.6211 | 79.0767 |
| | | | | $\rho$ is known | | |
| $\beta_{12}$ & $\beta_{22}$ | Case 4A1 | 1.2272 | 1.4570 | 1.5383 | 1.8772 | 2.2411 |
| | Case 4A2 | 4.8880 | 4.3675 | 4.2734 | 4.0250 | 3.8850 |
| | Case 4B | 1.0723 | 1.1213 | 1.1377 | 1.2038 | 1.2785 |

# CHAPTER FIVE

## SUMMARY AND CONCLUSIONS

Some recent studies have made use of the bivariate probit model in testing various hypotheses, but with only partial observability about the dichotomous dependent variables. These studies include Poirier's bivariate probit model using Gunderson's example of the retention of trainees, Farber's research on the demand for union representation, and Connolly's study concerning the decisions to arbitrate or negotiate the contracts between employees' unions and municipalities in Michigan. The maximum likelihood estimators in these partial observability cases will be inefficient compared to those obtained under full observability. But the degree of efficiency loss caused by partial observability is not yet known. Therefore in this study, we present several cases with different levels of observability for the bivariate probit model and we measure the efficiency loss of maximum likelihood estimators for each case through some experiments. The results that we get give us some idea about the cost of partial observability, and have practical relevance in studies like those above.

In Chapter Two, a formal statement of the bivariate probit model is presented. A general form of the model would be

$$y_{i1}^* = X_i\beta_1 + \varepsilon_{i1} \qquad i=1, 2, \ldots, N.$$
$$y_{i2}^* = X_i\beta_2 + \varepsilon_{i2}$$

and

$$y_{ij} = 1 \quad \text{iff} \quad y_{ij}^* > 0 \qquad j=1 \text{ or } 2.$$
$$y_{ij} = 0 \quad \text{iff} \quad y_{ij}^* \leq 0$$

where $X_i$ is a k-dimensional row vector of explanatory variables, $\beta_1$ and $\beta_2$ are k-dimensional column vector of unknown parameters and disturbance term $\begin{bmatrix}\varepsilon_1 \\ \varepsilon_2\end{bmatrix}$ has bivariate normal distribution with zero mean and variance-covariance matrix as $\begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}$. The variables $y_1^*$ and $y_2^*$ are always unobserved; different assumptions about the observability of $y_1$ and $y_2$ are considered. Six cases are introduced to represent full observability and different types of partial observability for the model. The example of a two-member committee voting under a unanimity rule can be applied to all of these cases. Case one is the case of full observability in which the dichotomous choices of both voters are always observable. Case two is the case of partial observability in the sense of Poirier, under the assumption that only the result of the joint choice of two decision-makers is observed. Case three is called the case of partial partial observability, in which one of the two parties' decision is fully observable. The other party's decision can be known only when the observable party votes "yes". In case four, which is called the case of partial observability with observed veto, when the outcome is "no", we observe one of the two parties casting its "no" vote. There are three alternative possibilities here concerning who will use the veto first if both parties wish to vote "no". The first possibility is that we assume some fixed and known probability p that the first party does so (case 4A1). The second possibility is having p as another parameter which needs to be estimated (case 4A2). Another possibility is that the party with the strongest sentiment toward a "no" vote will be observed casting the veto (case 4B).

We have provided likelihood functions for the joint estimation of the parameters for each of the various cases. Separate estimation (one equation at a time) is always possible for case one and for the first probit equation (the observable one) of case three. The separate estimation of the second probit equation of case three is possible only when the correlation coefficient $\rho$ is equal to zero. However, joint estimation is always more efficient than separate estimation unless the correlation coefficient ($\rho$) between the two probit equations is equal to zero.

In Chapter Three, six different information matrices have been derived corresponding to the joint estimation of the parameters in our six cases (with varying degrees of observability). The question of identification is also discussed by analyzing the rank of these matrices. We especially analyze the perverse case in which all of the coefficients of the exogenous variables except the constant terms are equal to zero. It has been found that in the above situation, regardless of the values of the constant terms, all of the information matrices for the partial observability cases are singular if $\rho$ is not known. When $\rho$ is known a priori, only the case of partial observability in the sense of Poirier and of observed veto with p as another parameter still can't be identified.

Another perverse case is when the two probit equations are identical. Then the case of partial observability in the sense of Poirier is not identified. But there are no problems with the other cases.

There are also other situations that will cause identification problems for some cases, which we have discussed in Chapter Three. The perverse cases that we mentioned there do not necessarily cover all that will make the information matrices of various partial observability cases singular. In general one needs to check the rank of the information matrix in each specific situation to make sure that the parameters are identified.

In Chapter Four, a large variety of experiments have been done to measure the cost of partial observability. We first try three arbitrary experiments and illustrate some general results. Then we try a second set of experiments by varying the values of parameters from low to high levels while holding the sample average values of $X_i\beta_1$ and $X_i\beta_2$ equal to zero. Two important effects have been observed from these experiments. One effect is that the degree of identification changes as the values of parameters change, and we call it the "identification effect". The other effect is that the probability of a "yes" vote of either (or both) party changes when the values of parameters change, and thus the sample split between the four possible outcomes changes. Hence we call this the "sample split effect". Both of these two effects change the cost of partial observability. If they work against each other, sometimes we can't tell the direction of the change of efficiency when the values of parameters change. Therefore, some more complicated experiments have been done with either the identification effect or the sample split effect held constant while the other changes. We then are more certain about the change of the cost under only one effect.

Among all the conclusions that we obtain from the results of these experiments, here we report some rather general and important ones. First we notice that the cost of partial observability is quite high, especially for the case of partial observability in the sense of Poirier (our case two). The cost of partial observability decreases markedly if any piece of observability information can be found. The law of diminishing marginal utility of information usually holds: the gain in moving from case two to case three (partial partial observability) or case four (observed veto) usually exceeds the gain in moving from case three or four to full observability (our case one). It is the first piece of observability information which is most important. From this, our suggestion for Poirier's model using Gunderson's retention of trainees or other similar examples is that if any information can be obtained, for example, observability for either party's decision or an observed veto, then the efficiencies of the estimated parameters can be greatly improved. This is relevant, for example, to Connolly's research on the arbitration or negotiation of the contracts between employees' unions and municipalities. In this case there is an observed veto. If this information is not used, this would be just a case of partial observability in the sense of Poirier. The high cost of partial observability for Poirier's model should make one reconsider the possibility of using the observed veto information.

The second conclusion is that specifying $\rho$ a priori improves the efficiencies of the estimates of the other parameters a great deal. Also the improvement from knowing $\rho$ is largest when the relative efficiency is lowest. This can be applied to all the partial observability cases. For Farber's case as an example, if the observability of the union employers' selection decision can't be obtained or the cost of getting the information is too high, then specifying $\rho$ in the model is another way to improve the efficiency.

A third conclusion is that the sample split has a strong influence on the relative efficiencies of the parameter estimates. For a given partial observability case, its efficiency relative to full observability will be higher, the smaller the proportion of observations which fall into the indistinguishable categories. For Poirier's model, the more observations that are of the "yes, yes" variety, the higher the relative efficiency. The fraction of such observations is observable. In case three, the higher the proportion of observations having the observable party voting "no", the lower the relative efficiency will be. The proportion of such observations is also observable. For example, 62.8% of Farber's sample are non-union workers and 37.6% of these nonunion workers expressed a preference for union representation. That is, 39.2% of the whole sample belongs to the indistinguishable categories (not in a union and would not vote for a union). The relative efficiency of the estimated parameters in the probit equation for the union employers' selection will be lower as this percentage increases. For the observed veto case, it is the proportion of observations having both parties voting "no" that is relevant, but this proportion is not directly observable.

The last conclusion is that the strength of identification matters. All of the partial observability cases are unidentified for some perverse values of the parameters, as we mention above. Their relative efficiency is very low for parameter values near such points, and it increases rapidly as the parameters move away from such points of singularity. However, these effects are not strong except in the immediate neighborhood of points of nonidentification.

# APPENDIX A

$$H_i = P(\varepsilon_{i2} < -X_iB_2,\ \varepsilon_{i1} - \varepsilon_{i2} < X_iB_2 - X_iB_1)$$

$$= P\left(\frac{\varepsilon_{i1}-\varepsilon_{i2}}{\sqrt{2(1-\rho)}} < \frac{X_iB_2-X_iB_1}{\sqrt{2(1-\rho)}},\ \varepsilon_{i2} < -X_iB_2\right)$$

$$= F\left(\frac{X_iB_2-X_iB_1}{\sqrt{2(1-\rho)}}\ -X_iB_2;\ -\sqrt{(1-\rho)/2}\right.$$

$$= F\left(\frac{X_iB_2-X_iB_1}{-2\rho'}\ -X_iB_2;\ \rho'\right)$$

where $\rho' = -\sqrt{(1-\rho)/2}$ .

Recall that $\frac{\partial F(a,b;\rho)}{\partial a} = \phi(a)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right)$, here define $a = \frac{X_iB_2-X_iB_1}{-2\rho'}$ , $b= -X_iB_2$

$$\frac{\partial H_i}{\partial \beta_1} = \phi(a)\Phi\left(\frac{b-\rho' a}{\sqrt{1-\rho'^2}}\right)\frac{1}{2\rho'}\ X_i'$$

$$= -\phi\left(\frac{X_iB_2-X_iB_1}{\sqrt{2(1-\rho)}}\right)\Phi\left(-\frac{X_iB_2+X_iB_1}{\sqrt{2(1+\rho)}}\right)\frac{1}{\sqrt{2(1-\rho)}}\ X_i' \ .$$

Let $G_i = P(\varepsilon_{i1} < -X_iB_1,\ \varepsilon_{i2} - \varepsilon_{i1} < X_iB_1 - X_iB_2)$ using the same method as above we can get

$$\frac{\partial G_i}{\partial \beta_2} = -\ \phi\left(\frac{X_iB_1-X_iB_2}{\sqrt{2(1-\rho)}}\right)\Phi\left(-\frac{X_iB_1+X_iB_2}{\sqrt{2(1+\rho)}}\right)\frac{1}{\sqrt{2(1-\rho)}}\ X_i'$$

$$= \frac{\partial H_i}{\partial \beta_1} \quad .$$

But $G_i = F(-X_iB_1, -X_iB_2; \rho) - H_i$

$$= 1 - \Phi(X_iB_1) - \Phi(X_iB_2) + F_i - H_i$$

and $$\frac{\partial G_i}{\partial \beta_2} = -\frac{\partial \Phi(X_iB_2)}{\partial \beta_2} + \frac{\partial F_i}{\partial \beta_2} - \frac{\partial H_i}{\partial \beta_2} ,$$

so $$\frac{\partial H_i}{\partial \beta_2} = [-\frac{\partial G_i}{\partial \beta_2} - \phi(X_iB_2) + \phi(X_iB_2)\Phi(\frac{X_iB_1-\rho X_iB_2}{\sqrt{1-\rho^2}})]X_i'$$

$$= \{-\frac{\partial H_i}{\partial \beta_1} - \phi(X_iB_2)\ [1-\Phi(\frac{X_iB_1-\rho X_iB_2}{\sqrt{1-\rho^2}})]\}\ X_i' \ .$$

For $\frac{\partial H_i}{\partial \rho}$ ,

note that $$\rho' = -\frac{1}{\sqrt{2}}(1-\rho)^{\frac{1}{2}}$$

$$\frac{\partial \rho'}{\partial \rho} = \frac{1}{2\sqrt{2(1-\rho)}}$$

and $$\frac{\partial[(X_iB_2-X_iB_1)/-2\rho']}{\partial \rho'}$$

$$= \frac{X_iB_2-X_iB_1}{2\rho'^2} = \frac{X_iB_2-X_iB_1}{1-\rho} \quad ,$$

$$\frac{\partial H_i}{\partial \rho} = \frac{\partial \rho'}{\partial \rho}\frac{\partial H_i}{\partial \rho'}$$

$$= \frac{\partial\rho'}{\partial\rho}\left[\frac{\partial H_i}{\partial\beta_1}\cdot\frac{\partial\beta_1}{\partial\left(\frac{X_iB_2-X_iB_1}{-2\rho'}\right)}\cdot\frac{\partial\left(\frac{X_iB_2-X_iB_1}{-2\rho'}\right)}{\partial\rho'}\right.$$

$$\left. + f\left(\frac{X_iB_2-X_iB_1}{\sqrt{2(1-\rho)}}, - X_iB_2; - \sqrt{(1-\rho)/2}\right)\right]$$

$$= \frac{1}{2\sqrt{2(1-\rho)}}\left\{\phi\left(\frac{X_iB_2-X_iB_1}{\sqrt{2(1-\rho)}}\right)\Phi\left(-\frac{X_iB_2+X_iB_1}{\sqrt{2(1+\rho)}}\right)\left(\frac{X_iB_2-X_iB_1}{1-\rho}\right)\right.$$

$$\left. + f\left(\frac{X_iB_2-X_iB_1}{\sqrt{2(1-\rho)}}, - X_iB_2; - \sqrt{(1-\rho)/2}\right)\right\}$$

## REFERENCES

Abowd, J. M. and Farber, H.S., 1982, "Job Quenes and the Union Status of Workers," Industrial and Labor Relation Review 35, 354-367.

Amemiya, T., 1974, "Bivariate Probit Analysis: Minimum Chi-Squared Methods," Journal of American Statistical Association 69, 940-944.

__________, 1975, "Qualitative Response Models," Annals of Economic and Social Measurement 4, 363-372.

__________, 1978, "The Estimation of a Simultaneous Equation Generalized Probit Model," Econometrica 46, 1193-1205.

Ashford, J. R. and Sowden, R. R., 1970, "Multivariate Probit Analysis," Biometrica 26, 535-546.

Connally, M., 1982, "The Effect of Compulsory Arbitration on the Bargaining Process and Wage Outcomes," unpublished Ph.D. dissertation, Michigan State University.

Farber, H. S., 1982, "Worker Preference for Union Representation," Research in Labor Economics, forthcoming.

__________, 1982, "The Demand for Union Representation," Working Paper No. 295, Massachusetts Institute of Technology.

Gunderson, M., 1974, "Retention of Trainees: A Study With Dichotomous Dependent Variables," Journal of Econometrics 2, 79-93.

Heckman, J. J., 1978, "Dummy Endogenous Variables in a Simultaneous Equation System," Econometrica 46, 931-959.

______________, 1979, "Sample Bias as a Specification Error," Econometrica 47, 153-161.

Kmenta, J., 1971, Elements of Econometrics, New York: Macmillan Publishing Co., Inc.

Poirier, D. J., 1980, "Partial Observability in Bivariate Probit Models," Journal of Econometrics 12, 210-217.

Rothenberg, T. J., 1971, "Identification in Parametric Models," Econometrica 39, 577-591.

Theil, H., 1971, Principle of Econometrics, New York: John Wiley and Sons, Inc.

Zellner, A. and Lee, T. H., 1965, "Joint Estimation of Relationships Involving Discrete Random Variables," Econometrica 33, 382-394.