A Parametric Test for the Distinction between Unemployed and Out of the Labor Force Statuses

Seung Chan Ahn and Stuart Low *

Whether or not to empirically consider two (employed *versus* not employed) or three (employed, unemployed, and out-of-labor-force) classifications in labor supply studies is a controversial issue. We develop a generalized censored probit likelihood function that nests both possibilities. A novelty of this likelihood function is that it allows researchers to test which representation of the labor market is appropriate as well as to estimate the degree to which classification errors may cloud inferences. Our empirical results demonstrate that classifying the three groups is useful to identify individuals' labor force and employment decisions separately. However, failure to incorporate classification ambiguities may result in unemployed rates that are understated and out-of-labor-force rates that are overstated.

Keywords: Bivariate probit, Misclassification, Unemployment, Out of labor force

JEL Classification: C3, C5, J6

I. Introduction

Labor supply decisions can be considered a joint outcome of two

* Associate Professor, Arizona State University, Tempe, AZ 85287-3806, USA, (E-mail) miniahn@asu.edu; Professor, Arizona State University, Tempe, AZ 85287-3806, USA, (E-mail) Stuart.Low@asu.edu, respectively. The authors gratefully acknowledge research support from Arizona State University's Business Dean's Council of 100. We also appreciate comments from Paul Burgess on an earlier draft of this paper. This paper was presented at the 14*th Seoul Journal of Economics* International Symposium on Sustainable Growth in East Asia, November 2006 and has benefited from suggestions made by two reviewers at that symposium.
[**Seoul Journal of Economics** 2007, Vol. 20, No. 1]

distinct choices. The initial choice may be characterized as an individual's preference to work or the labor force participation decision. Given entry into the labor force, the second choice reflects the ability to find a job prospect with wage offer exceeding the reservation wage. Identification and estimation of these two processes are important for the correct measurement of labor force participation and unemployment rates. Labor statistics estimate these two measures by categorizing individuals into three distinct labor market statuses: Out of the labor force (OLF) individuals, who choose not to enter the labor force; unemployed (UN) individuals, who enter the labor force but are unsuccessful in obtaining a satisfactory offer; and employed (EMP) individuals, who enter the labor force and receive a satisfactory offer. Although neither OLF nor UN individuals are working, labor statistics distinguish between them by including UN (as well as EMP) in measures of the labor force. An implicit assumption behind this distinction is that UN individuals will work if jobs paying prevalent market wages (and requiring acceptable working hours) are offered, while OLF individuals prefer not to work since their reservation wages are higher than their market wages. Indeed, under this assumption, unemployment rate is commonly used as a measure of general economic hardship or frictions in the labor market.

The distinction between OLF and UN is important not only for accurately measuring the unemployment rate, but also for modeling labor force decisions and employment outcomes. In early studies of labor supply, unemployment is considered a voluntary phenomenon. Only the labor force entry decision is relevant for employment, as an individual's employment outcome is not constrained by the ability to find a job. In this formulation OLF and UN are treated as behaviorally equivalent statuses (Heckman 1974; Hausman 1980). In contrast, recent studies explicitly distinguish between the two states by treating the labor force participation decision and the ability to find a job as distinct choices. This alternative formulation is consistent with the classification of three separate labor force statuses. Separation of the three groups leads to identification of the differential effects of demographic or economic variables on labor force participation and employment probabilities (Blundell and Meghir 1987; Blundell, Ham, and Meghir 1987, 1998).[1]

[1] Ahn (1990) and Sundt (1990) also estimate the labor force participation

Although the theoretical distinction between OLF and UN is intuitively straightforward, whether or not to empirically consider two (EMP, not EMP) or three (EMP, UN, and OLF) classifications in labor supply studies is still a controversial issue. Empirical relevance of the distinction between OLF and UN depends on whether or not nonworking individuals' observed OLF or UN statuses reveal their *true* willingness to work at prevalent market wages. In this paper, we address this issue by estimating separate index functions for the labor force entry decision and the ability to receive an acceptable job offer outcomes. Our particular concern is the potential classification errors which may exist between OLF and UN individuals which may cloud inferences based on labor force classifications. If sizeable proportions of UN or OLF workers are misclassified, these errors will result in biased estimates of the two index functions and incorrect estimates of labor force participation and unemployment rates as well as inappropriate inferences relative to labor supply issues.

Search theory stipulates that the major difference between OLF and UN states relates to job search activity, with OLF individuals engaging in zero quantity and UN individuals pursuing a positive amount (Burdett and Mortensen 1980; Devine and Kiefer 1991). From this foundation, the Bureau of Labor Statistics (BLS) has quite detailed specifications for classification, defining the unemployed as those who are available for a job during the reference week and have actively looked for a job during the preceding four weeks using at least one of a specified list of methods.[2] However, the BLS classification of OLF and UN cohorts may fail to correctly reveal nonworking individuals' preferences to work for several reasons. First, job search activity alone may not be a sufficient criterion by which nonworking individuals who prefer to work can be differentiated from those who prefer not to work. In the U.S., the average monthly flow to EMP from OLF is greater than the average flow to EMP from UN (Ehrenberg and Smith 1987, Chapter 15). This may indicate that a non-negligible portion of OLF individuals are in fact available for employment, but are classified as OLF because of their low search intensity.[3] Second, all search information is

and employment equations by distinguishing the three labor market statuses. See also Bowlus (1995).

[2] Additionally included in UN are temporarily laid off workers and those waiting to report to a new job within a month.

[3] A nonworking individual preferring employment may not engage in job

self-reported and not independently verified. Thus responses are likely to be influenced by the form of the question.[4] Third, some individuals, particularly those seeking to qualify for or continue to receive unemployment insurance benefits may have an incentive to over-report their search activity (Burgess 1992). Finally, even if there were no reporting errors, the official BLS classification criteria lack concrete thresholds for the quantity of and intensity of the minimally required search activity needed for an individual to be classified as UN, leaving this determination to the discretion of the interviewer.[5] For these reasons, one might expect the distinction between UN and OLF to be imprecise, with some individuals observationally equivalent to those who are UN (OLF) classified as OLF (UN). Thus use of BLS criteria may produce only an arbitrary distinction between UN and OLF (Clark and Summers 1982).

Previous empirical consideration of whether OLF and UN are observationally identical or distinct has relied on an examination of labor market outcomes of the groups. Clark and Summers (1982) conclude that there is no distinction between the states based on identical mean durations of UN.[6] Flinn and Heckman (1982) and Gönül (1992) view transition probabilities to employment, with Flinn and Heckman concluding that OLF and UN are distinct for young men, while Gönül finds a distinction for young women but not for young men.[7] These tests thus do not provide uniform inferences. Further, by assuming that OLF and UN states are correctly identified, these studies may yield misleading inferences relative to unemployment rate estimates if there exists imprecision in the classification process.

search if the job arrival rate for non-job-searchers is nonzero and search costs are high.

[4] An example of the potential impact of the form of survey questions is found in Filer, Hamermesh and Rees (1996, p. 7). In 1994 the Current Population Survey officially changed some questions related to female job search. Estimated unemployment rates based on the old *vs.* new questions differed by 0.8 percent.

[5] Estimated transitions from OLF to UN or *vice versa* are non-negligible and may exceed transitions to employment (Gönül 1992; Flaim and Hogue 1985). These results may suggest that many nonworking individuals are erroneously classified based on misreported search activity or interviewer error.

[6] Clark and Summers utilize Current Population Survey data for teenagers.

[7] Both of these studies use National Longitudinal Survey data for young people.

In this paper, we consider an empirical specification (likelihood function) by which researchers can identify and estimate both the labor force participation and employment index functions using cross sectional data, even in the presence of potential classification errors between OLF and UN individuals. The contribution of our general specification is twofold. First, it provides a simple parametric test for the empirical distinction between the two nonworking groups in terms of their demographic and economic attributes. Second, the specification framework allows researchers to estimate the empirical lack of distinctness between individuals classified as UN or OLF. At one extreme, we may find no ambiguity of classification. Alternately, we may find classification imprecision in the UN classification (some UN individuals may have attributes more closely associated with OLF), in the OLF classification or in both. As the degree of estimated empirical ambiguity of classification rises, our ability to determine the distinctness of reported UN and OLF statuses is diminished. Further, unemployment rates estimated assuming all classifications are accurate may be well off the mark. Our estimates will allow us to determine the degree of classification ambiguity and will provide us with revised estimates of unemployment and OLF rates accounting for the classification uncertainty.

It has been well known in the literature that misclassification of dependent variables in univariate binary choice models (*e.g.*, probit or logit models) leads to inconsistent coefficient estimates. As a treatment to this problem, Hausman, Abrevaya, and Scoot-Morton (1998) have developed a modified maximum likelihood estimator that can control for the biases due to misclassification. Using the modified method, they found that classifying as job-changers the respondents (in popular labor survey data such as Current Population Survey or Panel Study of Income Dynamics) who report tenure as 12 months or fewer could overestimate the true probability of individuals changing their jobs within a year. Using the similar method, Caudill and Mixon (2005) recently found that (true) incidence of cheating in undergraduate classrooms could be much higher than the value of incidence estimated from students' self reports in survey data. Our approach can be viewed as an extension of the modified approach to bivariate binary-choice models (two separate decisions of LF participation and employment).[8]

[8] Hausman, Abrevaya, and Scoot-Morton (1998) also develop a

In order to demonstrate empirical importance of the method we develop, we apply it to a sample of married women obtained from the 1988 Panel Study of Income Dynamics. We find that the sample separation of OLF and UN individuals is useful to identify labor force participation and employment success decisions, although our results are consistent with the presence of classification errors between OLF and UN. We also find that estimates that ignore the possible classification errors are potentially biased and underpredict both labor force entry and unemployment probabilities.

This paper is organized as follows. Section II presents our basic model and discusses estimation procedures. Section III explains the sample used in our empirical study, and Section IV discusses our empirical results. Concluding remarks follow in Section V.

II. Model

In this section, we introduce a simple three-state model in which individuals' labor market statuses are distinguished based on two separate decisions. For this model, we derive a likelihood function which is designed to control for potential classification errors among OLF and UN individuals. We also discuss the hypotheses of interest and model specification tests.

A. Basic Model

The foundation of our approach is a simple three-state model based on search theory, which is also considered by Blundell, Ham, and Meghir (1998).[9] We begin by assuming that jobs are not always available for individuals considering employment. Each worker is assumed to be aware of the probability that she can receive an acceptable job offer as well as the wage offer distribution and search costs, and then compares the expected value of job search to the value of her leisure and home production before she begins to look

semiparametric estimation method that does not require parametric assumptions about the probability of the dependent variable being one. Lewbel (2000) provides general conditions under which the modified binary-choice models with potential misclassification can be semiparametrically identified. It would be an interesting future research agenda to develop a semiparametric estimation method for the models with two choices.

[9] See also Ahn (1990) and Sundt (1990).

for a job. A woman becomes available for work and spends non-zero time on job search only if the former value exceeds the latter.[10] With these assumptions, we define a married woman's disposition to be in the labor force by the index function:

$$y^*_{lf} = X_{lf}\beta_{lf} + e_{lf}. \tag{1}$$

Here X_{lf} contains explanatory variables relevant for labor force participation decisions, β_{lf} is a vector of their coefficients and e_{lf} is the random error term.[11] The latent variable y^*_{lf} can be viewed as the difference between the expected value of job search and the value of OLF activity. Given an individual's participation, the likelihood of her being employed depends both upon job-search intensity and effectiveness as well as on labor demand. In order to capture this probability, we define the employment index function by:

$$y^*_{emp} = X_{emp}\beta_{emp} + e_{emp}, \tag{2}$$

where all the terms are defined similarly to those in Equation (1). Here the latent variable y^*_{emp} measures job availability. We assume that given her participation decision, a woman is employed whenever $y^*_{emp} > 0$, and otherwise remains unemployed.

Two points made by Blundell, Ham, and Meghir (1998) are worth noting for the proper interpretation of the employment index function. First, the probability of positive employment, $\Pr(y^*_{emp} > 0)$, does not simply coincide with the job arrival rate. Since laid-off workers are also included as unemployed in the data, this probability can be interpreted as the sum of the arrival rate for job searchers and the job-retention rate for employed workers. Second, since employment probability affects labor force decisions through the value of search, and since all the variables relevant for labor force decisions would also likely affect the employment probability through the reservation wage and search intensity, it is unlikely that different variables influence labor force participation and employment probabilities. Accordingly, we specify $X_{lf} = X_{emp} \equiv X$. In addition, we

[10] For a rigorous theoretical derivation of this labor force participation rule, see Blundell, Ham, and Meghir (1998). The value of search also depends on the probability of separation from jobs (*e.g.*, lay offs).

[11] Here and later, we drop subscript "*i*" indexing individuals for notational convenience.

assume that the employment function (2) is an unconditional one defined for all individuals regardless of their participation decisions. Therefore, the positive sign of y^*_{emp} for an OLF individual should be interpreted as meaning that an acceptable job would be available to her if she decided to participate in the labor force.

A woman's latent *true* EMP, UN, and OLF states, which we denote by TEMP, TUN, and TOLF, respectively, depend on the signs of the latent variables y^*_{lf} and y^*_{emp}. Specifically, if we assume that the error terms e_{lf} and e_{emp} follow a bivariate standard normal distribution, the probability of being in one of the three states is given by:

$$\Pr(i \in TEMP) = \Pr(y^*_{lf} > 0 \text{ and } y^*_{emp} > 0) = F(X\beta_{emp}, X\beta_{emp}, \rho); \quad (3.1)$$

$$\Pr(i \in TUN) = \Pr(y^*_{lf} > 0 \text{ and } y^*_{emp} < 0) = \Phi(X\beta_{lf}) - F(X\beta_{lf}, X\beta_{emp}, \rho); \quad (3.2)$$

$$\Pr(i \in TOLF) = \Pr(y^*_{lf} < 0) = 1 - \Phi(X\beta_{lf}), \quad (3.3)$$

where $F(\cdot, \cdot, \cdot)$ and $\Phi(\cdot)$ represent bivariate and single standard normal cumulative density functions, respectively, and ρ is the correlation coefficient between e_{lf} and e_{emp}. Therefore, given an individual's demographic and economic attributes X, these three probabilities can be explained by the parameter vector $\theta \equiv (\beta'_{lf}, \beta'_{emp}, \rho)'$.[12]

B. Model with Classification Errors

Identification and estimation of the model given by Equations (1) and (2) require a sample classification of individuals into EMP, UN, and OLF groups. We denote the classified labor market states of the women in our sample by CEMP, CUN, and COLF, respectively. In cases where these observed states coincide with true states, Equations (1) and (2) can be viewed as a bivariate probit model with partial observability (see Meng and Schmidt 1985). In particular, since employment outcomes are observable only for labor force participants, the model corresponds to the censored probit case (Farber 1983), which leads to the log-likelihood function:[13]

[12] If we restrict $\rho = 0$, the employment Equation (2) may be regarded as a conditional one defined over LF participants only. In this case, the parameters in Equations (1) and (2) can be estimated by two separate probits.

$$
\begin{aligned}
l_c(\theta) &= \sum_{i\in CEMP} \ln[\Pr(i\in CEMP)] + \sum_{i\in CUN} \ln[\Pr(i\in CUN)] \\
&\quad + \sum_{i\in COLF} \ln[\Pr(i\in COLF)] \\
&= \sum_{i\in CEMP} \ln[F(X\beta_{lf}, X\beta_{emp}, \rho)] + \sum_{i\in CUN} \ln[\Phi(X\beta_{lf}) - F(X\beta_{lf}, X\beta_{emp}, \rho)] \\
&\quad + \sum_{i\in COLF} \ln[1-\Phi(X\beta_{lf})].
\end{aligned} \tag{4}
$$

Consistency of the censored probit (maximum likelihood) estimates crucially depends on whether the sample distinction between CUN and COLF is relevant. When this distinction is questioned, for the reasons mentioned in the previous section, one may wish to estimate Equations (1) and (2) without distinguishing the two states. This scenario leads to an alternative estimation procedure that is considered by Poirier (1980). Using Poirier's method we need only distinguish employed and nonemployed (both CUN and COLF) women. Under this formulation the relevant log-likelihood function is given by:

$$
\begin{aligned}
l_p(\theta) &= \sum_{i\in CEMP} \ln[\Pr(i\in CEMP)] + \sum_{i\notin CEMP} \ln[\Pr(i\notin CEMP)] \\
&= \sum_{i\in CEMP} \ln[F(X\beta_{lf}, X\beta_{emp}, \rho)] + \sum_{i\notin CEMP} \ln[1-F(X\beta_{lf}, X\beta_{emp}, \rho)].
\end{aligned} \tag{5}
$$

Given that observed employment and nonemployment statuses do not contain classification errors, maximizing the log-likelihood function (5) can yield a consistent estimator of the true values of θ.

However, a serious limitation in the Poirier method is that the parameter vectors β_{lf} and β_{emp} are not identified because of their interchangability in Equation (5). That is, although it is possible to

[13] As classified and true states are here assumed at this point to be identical, $\Pr(i\in TOLF)=\Pr(i\in COLF)$ with comparable equivalencies for UN and EMP. For later clarity, we express the likelihood function in terms of CEMP, CUN, and COLF.

estimate the two parameter vectors by maximizing $l_p(\theta)$, it is not possible to determine which estimates are for which equation unless some prior information is available on different effects a variable may have (in terms of sign or size) on participation decisions and employment outcomes, or unless X_{lf} and X_{emp} are distinct, which is a restriction that may be difficult to justify in practice.

A method we adopt to circumvent this identification problem is to generalize the censored probit model in (4) by parameterizing the probabilities of discrepancies between observed and true non-employment statuses UN and OLF. Specifically, we define:

$$P_1 \equiv \Pr(i \in CUN \mid i \in TUN) = \Phi(Z_1 \gamma_1); \tag{6.1}$$

$$P_2 \equiv \Pr(i \in CUN \mid i \in TOLF) = \Phi(Z_2 \gamma_2), \tag{6.2}$$

where Z_1 and Z_2 denote vectors of explanatory variables, and γ_1 and γ_2 are corresponding coefficients. Here P_1 represents the conditional probability that an individual's reported UN status (CUN) coincides with her true UN status (TUN), while $(1-P_1)$ represents the probability that an UN individual is misclassified as OLF. The conditional probability P_2 represents the probability that an OLF individual is misclassified as UN, while $(1-P_2)$ is the probability that her reported OLF status is correctly reported. Since both P_1 and P_2 are defined as conditional on true unemployed or true OLF status, they are likely to be related to all the explanatory variables for the labor-force-participation-decision and ability-to-find-a-job-outcome equations. Therefore, we simply specify $Z_1 = Z_2 = X$.[14]

The conditional probability P_1 can be also interpreted as a measure of the unambiguity of reported UN individuals' true status in terms of their demographic and economic attributes X. For example, if P_1 equals one for all nonworkers, this implies that all nonworkers with characteristics consistent with UN (TUN) are classified as UN (CUN). If P_1 is less than one, it indicates that some nonworkers with characteristics consistent with UN (TUN) are potentially misclassified as OLF (COLF).[15] In contrast to P_1, the conditional probability P_2

[14] Since this specification is somewhat arbitrary, its validity is subject to some justifying specification tests, which we discuss below.

[15] An example of such individuals is discouraged workers who desire to work but quit job search.

measures the degree of ambiguity of observed OLF individuals. That is, if P_2 equals zero, all nonworkers with OLF attributes (TOLF) are classified as OLF (COLF). However, when P_2 is greater than zero, it indicates that some nonworkers having attributes consistent with OLF (TOLF) are potentially misclassified as UN (CUN).[16]

We may now specify a generalized censored model, which is used for our empirical study. Introducing the two conditional probabilities P_1 and P_2, we can define the unconditional probabilities of being in CUN and COLF as:

$$\begin{aligned} \Pr(i \in CUN) &= \Pr(i \in CUN \mid i \in TUN)\Pr(i \in TUN) \\ &\quad + \Pr(i \in CUN \mid i \in TOLF)\Pr(i \in TOLF) \\ &= \Phi(X\gamma_1)[\Phi(X\beta_{lf}) - F(X\beta_{lf}, X\beta_{emp}, \rho)] + \Phi(X\gamma_2)[1 - \Phi(X\beta_{lf})]; \end{aligned} \tag{7}$$

$$\begin{aligned} \Pr(i \in COLF) &= \Pr(i \in COLF \mid i \in TUN)\Pr(i \in TUN) \\ &\quad + \Pr(i \in COLF \mid i \in TOLF)\Pr(i \in TOLF) \\ &= [1 - \Phi(X\gamma_1)][\Phi(X\beta_{lf}) - F(X\beta_{lf}, X\beta_{emp}, \rho)] \\ &\quad + [1 - \Phi(X\gamma_2)][1 - \Phi(X\beta_{lf})]. \end{aligned} \tag{8}$$

If we insert Equations (7) and (8) into Equation (4), we obtain the following log-likelihood function for the generalized censored probit specification:

$$\begin{aligned} l_g(\theta, \gamma) &= \sum_{i \in CEMP} \ln[F(X\beta_{lf}, X\beta_{emp}, \rho)] \\ &\quad + \sum_{i \in CUN} \ln[\Phi(X\gamma_1)\{\Phi(X\beta_{lf}) - F(X\beta_{lf}, X\beta_{emp}, \rho)\} + \Phi(X\gamma_2)\{1 - \Phi(X\beta_{lf})\}] \\ &\quad + \sum_{i \in COLF} \ln[\{1 - \Phi(X\gamma_1)\}\{\Phi(X\beta_{lf}) - F(X\beta_{lf}, X\beta_{emp}, \rho)\} \\ &\qquad + \{1 - \Phi(X\gamma_2)\}\{1 - \Phi(X\beta_{lf})\}], \end{aligned} \tag{9}$$

where $\theta = (\beta'_{lf}, \beta'_{emp}, \rho)'$ and $\gamma = (\gamma'_1, \gamma'_2)'$. It can be easily shown that all the parameters in Equation (9) can be identified unless $\gamma_1 = \gamma_2$.[17]

[16] An example of such individuals is people who do not have serious intention to work, but report job search activities to solicit unemployment benefits.

[17] It would also be possible to specify the likelihood function to include the probability of classification errors in reported EMP status. However, as this is an observable event, we presume CEMP=TEMP.

The generalized censored probit model (9) directly nests both the censored and Poirier probit models (4) and (5) and will thus permit specification tests of the appropriateness of either specification. Specifically, the censored probit model (4) is obtained if $\Phi(X\gamma_1)=1$ and $\Phi(X\gamma_2)=0$ for all nonworking individuals. This occurs when there are no classification errors for either unemployed or OLF individuals, and implies that CUN and COLF coincide with TUN and TOLF, respectively. Accordingly, the presence of misclassified UN and OLF statuses in our sample can be easily checked by conventional likelihood-ratio (LR), Lagrangean-Multiplier (LM) or Wald tests of the hypothesis that $\Phi(X\gamma_1)=1-\Phi(X\gamma_2)=1$.

On the other hand, if $\Phi(X\gamma_1)=\Phi(X\gamma_2)$ for all nonworkers ($\gamma_1=\gamma_2$, or equivalently $P_1=P_2$), the likelihood function (9) reduces to:

$$l_p(\theta)+\{\sum_{i\in CUN}\ln[\Phi(X\gamma_1)]+\sum_{i\in COLF}\ln[1-\Phi(X\gamma_1)]\} \tag{10}$$

and provides estimates of θ that are equivalent to the Poirier probit estimates of θ from Equation (5).[18] Testing the Poirier specification (5) against the generalized censored probit model (9) is equivalent to testing the information content of the distinction between reported UN and OLF (CUN and COLF). When $\gamma_1=\gamma_2$, the general censored probit and Poirier models are informationally equivalent in terms of estimation of β_{lf} and β_{emp} and the distinction between CUN and COLF provides no information for the separate identification of labor force and employment decisions.[19] In contrast, if $\gamma_1\neq\gamma_2$, the parameters β_{lf} and β_{emp} are no longer interchangeable and labor force and employment decisions can be separately identified. This implies that whether or not the distinction between CUN and COLF is informative for individuals' labor force and employment decisions can be easily checked by parametric tests of the relevance of the restriction $\gamma_1=\gamma_2$.

Several intermediate outcomes warrant discussion. If P_1 is less than one and P_2 equals zero, the sole ambiguity of classification arises as some individuals observationally equivalent to those who

[18] This occurs because the second term of Equation (10) is irrelevant for the estimation of θ as it contains only $\gamma_1(=\gamma_2)$.

[19] This occurs because β_{lf} and β_{emp} are interchangeable in Equation (9) if $\gamma_1=\gamma_2$.

are TUN (in terms of demographic and economic attributes) are classified as OLF. Conversely, if P_2 is greater than zero and P_1 equals one, some individuals observationally equivalent to those who are TOLF are classified as UN, with no ambiguity of classification for those individuals with UN characteristics. Finally, if P_1 is less than one and P_2 is greater than zero (and are unequal) we would have dual classification ambiguity. In each of these intermediate cases, we might reject *both* the censored probit and Poirier probit representations of the labor force. Our results would indicate that distinguishing three labor force statuses (OLF, UN, and EMP) is appropriate for the estimation of the labor force and employment decisions, but failure to consider the ambiguity of classification may provide misleading inferences. In addition, any of the intermediate cases has implications for estimated UN and OLF proportions. In the first case, unemployment rate estimates would be overstated; in the second case, unemployment rate estimates would be understated; while in the third case, unemployment estimates could be over or understated depending on the magnitude of the classification overlap.

C. Specification Tests

The reliability of statistical inferences based on the generalized censored probit model (9) is critically dependent on the correct specification of the model. Some specification tests can be utilized to test the null hypothesis that the general model is correctly specified. We utilize two different statistics. The first is a Hausman (1978) test statistic,

$$HT_g \equiv (\hat{\theta}_p - \hat{\theta}_g)'[V(\hat{\theta}_p - \hat{\theta}_g)]^{-1}(\hat{\theta}_p - \hat{\theta}_g), \tag{11}$$

which is asymptotically χ^2-distributed under the null hypothesis that the general model is correctly specified.[20] In Equation (11), $\hat{\theta}_p$ and $\hat{\theta}_g$

[20] The motivation of this Hausman test is as follows. Whenever the generalized censored probit model is correctly specified, both the Poirier and generalized censored probit estimators are consistent. In contrast, suppose that the generalized model is misspecified; that is, some important regressors are omitted, and/or, the error terms in the labor-force and employment index functions are not normal. In this case, the two estimators are inconsistent with different probability limits. Thus, the Hausman test would have power to detect possible misspecification of the generalized model. One may consider

denote parameter estimates from the Poirier and generalized censored probit models, respectively; and $V(\bullet)$ captures the relevant variance-covariance matrix. The Hausman statistic has the degrees of freedom equal to the number of parameters in θ(say, q). The second statistic we use is a Hausman score test (following Peters and Smith 1991),

$$HST_g \equiv s_p(\hat{\theta}_g)'\{V[s_p(\hat{\theta}_g)]\}^{-1}s_p(\hat{\theta}_g) \tag{12}$$

where $s_p(\theta) = \partial l_p(\theta)/\partial\theta$ represents a score vector for the Poirier probit. This statistic is also χ^2-distributed with q degrees of freedom. Appendix A provides the motivations of the HT_g and HST_g test statistics as well as a description of how the variance-covariance matrices may be consistently estimated.[21]

III. Data and Variables

We estimate the generalized censored probit model (9) using a sample of married women from the 1988 Panel Study of Income Dynamics (PSID).[22] The initial potential sample of 4,048 women is reduced to 2,706 observations by several data exclusions.[23]

an alternative Hausman test based on the censored and generalized censored probits. But this alternative test would be inappropriate, because the censored probit estimator can be inconsistent while the generalized probit estimator is consistent. This happens, for example, if TUN (or TOLF) is different from CUN (or COLF). Thus, rejection by the alternative test could indicate either misspecification of the generalized model, or the discrepancy between true and classified unemployment (or OLF).

[21] As such, the Hausman and Hausman score tests may not be omnipotently powerful. Newey (1985) shows that the Hausman test can be interpreted as a Generalized Method of Moments (GMM) overidentifying restriction test, and that the GMM tests could have little power in some directions of model misspecification although they do in other directions.

[22] The empirical results are intended to illustrate the generalized censored probit model. They are thus illustrative of the inferences that might be drawn if the generalized model is estimated. However a different selection of groups (unmarried women, married men, unmarried men) or different time period (when the labor market may be more or less tight) could certainly lead to somewhat different inferences.

[23] Exclusions include: Ethnicities other than black or white; households with female heads (as no information on husbands is available); women who are retired, disabled, students, prisoners, or employed in agriculture (or whose husbands are employed in agriculture); women residing outside North

TABLE 1
VARIABLE DEFINITION AND DESCRIPTIVE STATISTICS

Variable	Definition	Mean	S.D.
LF	=1 if in LF (EMP or UN); =0 otherwise (OLF)	0.738	0.440
EMP	=1 if employed; =0 otherwise	0.695	0.461
HSGRAD	=1 if high school (not college) graduate; =0 otherwise	0.588	0.492
COGRAD	=1 if college graduate; =0 otherwise	0.220	0.414
AGE	years of age	36.63	10.81
EXP	years of actual work experience	10.191	7.709
HLINC	husband's labor income (in $10,000s)	2.646	2.558
HNLINC	husband's nonlabor income (in $10,000s)	0.216	0.723
WNLINC	wife's nonlabor income (in $10,000s)	0.036	0.285
WPHLIM	=1 if physical handicap limits some types of job; =0 otherwise	0.101	0.302
BLACK	=1 if black; =0 if white	0.237	0.425
KIDS5	number of children of age 5 in household	0.508	0.774
KIDS17	number of children of age 6-17 in household	0.799	1.029
SMSA	=1 if living in SMSA; =0 otherwise	0.565	0.496
REGNC	=1 if living in North Central region; =0 otherwise	0.213	0.409
REGS	=1 if living in Southern region; =0 otherwise	0.412	0.492
REGW	=1 if living in Western region; =0 otherwise	0.170	0.376
UNEMPR	unemployment rate in county of residence	0.055	0.025

Definitions of the variables used in our estimation along with sample means and standard deviations are presented in Table 1. 73.8% of our sample is in the labor force, with the remaining 26.2% classified as OLF, with many of these individuals reporting their status as housewives. For the full sample, 69.5% are employed, implying that 4.3% are UN.[24] These UN women are so classified because they have been looking for jobs during the last four weeks or are temporarily laid-off.

Other variables in Table 1 are explanatory variables included to capture the woman's disposition to enter the labor force and ability to find an acceptable job. As noted above, provided the two conditional probabilities P_1 and P_2 are not equal, our generalized

America; women older than 64; and women with missing or unreliable data (such as experience greater than age).

[24] Equivalently, among those in the labor force, 5.9% are classified as UN.

censored probit model permits identification of both LF and EMP equations (along with P_1 and P_2) with identical sets of covariates. We thus include all explanatory variables in both equations and avoid a difficulty to theoretically justify distinction between regressors included in each equation.[25] Demographic effects on both labor force and employment decisions are captured by using variables such as dummy variables for high school and college diplomas (*HSGRAD* and *COGRAD*), the numbers of children below the ages of 6 and 18 years (*KIDS5* and *KIDS17*), a dummy variable for black women (*BLACK*) and age (*AGE*). Prior work experience could affect both labor force decisions and job opportunities. The actual number of years worked since the age of 18 (*EXP*) is used to capture this effect. Regional effects are captured by city size and area of residence. The dummy variable *SMSA* represents residency in a SMSA, and the three dummy variables *REGNC*, *REGS*, and *REGW* represent residency in North Central, Southern and Western areas of the U.S. continent, respectively. In order to capture income effects on a woman's labor force and employment decisions, we use her nonlabor income (*WNLINC*) and her husband's labor and nonlabor incomes (*HLINC* and *HNLINC*). The potential health effect on labor force and employment statuses is controlled for by using a dummy variable indicating physical condition limiting some types of work (*WPHLIM*). Finally, we include the local unemployment rate (*UNEMPR*) in order to capture differing demand conditions across areas.

IV. Results

Table 2 reports maximum likelihood estimates of the generalized censored probit log-likelihood function specified in Equation (9). For the most part, the direction of variable impacts conforms to our prior expectations of their effect on LF, in column 2, and EMP in column 3.

[25] Identification of the equations is provided by the generalized probit model. The two index functions and the two conditional probabilities in the generalized probit model may depend on different regressors. If there is some theoretical guidance on what variables should appear in one equation, but not in others, that is, if we have some exclusive restrictions on the coefficients of regressors, that information could be used to obtain more efficient estimates. The impact of such information on the power of the test is left to future research.

TABLE 2

GENERALIZED CENSORED PROBIT ESTIMATES

Regressors	Labor Force	Employment	Prob. classified as UN given UN attributes (P_1)	Prob. classified as UN given OLF attributes (P_2)
Constant	1.543* (5.614)[1)]	3.305* (7.727)	6.727** (2.338)	2.411* (3.669)
HSGRAD	0.368* (3.606)	-0.015 (0.113)	-0.393 (0.643)	-0.318 (1.300)
COGRAD	0.465* (3.686)	0.696* (3.465)	0.981 (1.154)	0.301 (0.872)
KIDS5	-0.417* (7.803)	-0.464* (4.384)	-0.687 (1.620)	-0.765* (4.599)
KIDS17	-0.089** (1.982)	-0.040 (0.670)	-0.422 (1.514)	0.004 (0.043)
BLACK	0.283** (1.771)	-0.469* (3.103)	0.365 (0.627)	0.709** (2.290)
AGE	-0.042* (5.503)	-0.055* (6.621)	-0.160* (2.841)	-0.069* (4.303)
EXP	0.151* (11.08)	0.034* (2.800)	0.016 (0.326)	0.088** (2.122)
SMSA	-0.174** (1.909)	0.144 (1.188)	-1.707** (2.560)	0.253 (0.949)
REGNC	0.135 (1.018)	-0.013 (0.071)	0.241 (0.326)	-0.275 (0.912)
REGS	0.183 (1.590)	0.045 (0.292)	0.205 (0.314)	-0.865* (2.804)
REGW	0.198 (1.291)	-0.268 (1.541)	-0.522 (0.719)	-1.034** (2.460)
WNLINC	0.512 (1.637)	-0.090 (0.873)	-0.883 (0.560)	-2.295 (0.512)
HLINC	-0.101* (5.294)	0.010 (0.329)	0.126 (0.706)	-0.202** (2.537)
HNLINC	0.194 (1.517)	-0.153* (2.579)	-0.850** (1.909)	-0.741 (1.253)
WPHLIM	-0.221 (0.911)	-0.689* (4.460)	-0.172 (0.246)	0.014 (0.041)
UNEMPR	-6.299** (3.610)	-3.709 (1.633)	7.227 (0.827)	-10.68** (1.768)
ρ	0.661* (2.745)			
Log of likelihood	-1581.38			
# of observations	2,706			
Specification Tests	Hausman Test (HT_g, $df=35$) 18.7 [$p=0.99$][2)]		Hausman Score Test (HST_g, $df=35$) 27.4 [$p=0.82$][2)]	

Notes: 1) Absolute value of t-statistic in parentheses.
2) p-values.
3) * Significant at $\alpha=.01$(two tail test).
** Significant at $\alpha=.10$(two tail test).

Individuals with higher educational levels (relative to the excluded less than high school degree group) are significantly more likely to be EMP. Women with more experience are significantly more likely, and older women significantly less likely, to be both in the LF and EMP. The existence of children in the household is associated with lower probabilities of both LF entry and EMP outcome, with the impact far more pronounced in both significance and magnitude for households with children younger than 6. Black women are more likely to be in the LF, but less likely to be EMP, implying a higher unemployment rate relative to white women which is often observed in economy-wide data.

Viewing income effects on women's labor force and employment decisions, our results show that higher labor income from the husband is associated with lower LF participation with an insignificant impact on EMP. Higher HLINC may be associated with more hours of work by the husband and given the need to provide household services (among wife and husband) may make it less likely the female is in the labor force. The direction of impact of wife and husband non labor income on both labor force participation and employment is identical although the only significant impact shows nonlabor income reducing the EMP likelihood as HNLINC rises. Regional effects are generally insignificant determinants of either LF or EMP, with the exceptions being that women living in a SMSA or in an area with a higher local unemployment rate are less likely to be in the LF. Our results also indicate that the correlation between the LF and EMP process is significant, with a point estimate of .661.

The final two columns in Table 2 represent estimates of P_1, the conditional probability that reported UN status corresponds to true UN status, and P_2, the conditional probability that an OLF woman is misclassified as UN.[26] Our results indicate that the only significant determinants of P_1 are the woman's age, her husband's non-labor income and the indicator for living in an SMSA, each of which is associated with a lower conditional probability that the observed and true unemployed statuses coincide. With respect to P_2, women who are older or have young children in the household, women whose husbands have greater labor income, and women from areas with

[26] Equation (6) defines P_1 and P_2 in greater detail. In the estimation of Equation (9), both P_1 and P_2 are parameterized as cumulative normal density functions of all of the independent variables in the model.

higher local unemployment rates or from the south or west have significantly reduced likelihoods of having characteristics comparable to an OLF individual but being misclassified as being unemployed. In contrast, black women and those with greater labor market experience have a significantly larger P_2 probability.

In sum, our generalized censored probit results in Table 2 allow us to determine not only a covariate's impact on labor force participation decisions and employment outcomes but also its effect on the likelihood that an individual with characteristics comparable to OLF or UN individuals is misclassified as UN or OLF. For example, women with 1 or more child under 6 years old in the household are less likely to be in labor force, less likely to be employed, and, as their status is often reported as "housewife," less likely to be misclassified as UN when they are truly OLF.[27] Alternatively, women with greater actual labor market experience are more likely to be both in the labor force and employed, and are more likely to be misclassified as UN when, in fact, they are truly OLF. This latter impact might be due to a desire to maintain unemployment insurance eligibility by reporting search activity when, in reality, the status is observationally indistinguishable from OLF. Finally, black women are more likely to be in the labor force and less likely to be employed, with a higher likelihood that their OLF status is misclassified as UN. The initial impacts on LF and EMP imply a higher unemployment rate for black women (relative to white females), while the latter effect indicates that reported unemployment rate for black females may be too high due to misclassification of OLF as UN.

For inferences from our generalized censored probit models to provide reliable estimates of the LF, EMP, P_1 and P_2 processes, the underlying model must be correctly specified. We use both the Hausman (*HT*) and Hausman score (*HST*) tests that are introduced in section II-*C* and Appendix A. Results of these tests reported in Table 2 demonstrate that neither test rejects the null hypothesis that our empirical specification (9) is satisfactory. We thus conclude that our representation of the generalized censored probit model is not inappropriate for the analysis of the labor force and employment decisions of married women.

[27] A comparable example is older women, who may have never entered the labor force or who may have retired.

A primary objective for developing the generalized censored probit likelihood function in Equation (9) is that it nests both the censored probit model used when UN and OLF are considered distinct states (specified in Equation (4)) as well as the Poirier probit specification used when UN and OLF are unnecessary to distinguish (given in Equation (5)). Our generalized censored probit function (9) thus permits us to determine if the censored probit model is appropriate, which occurs when $P_1 = 1 - P_2 = 1$; if the Poirier probit model is supported, which implies that $P_1 = P_2$; or if neither is confirmed due to the failure to consider possible ambiguity of UN and OLF classifications, which would arise if $P_1 < 1$ and/or $P_2 > 0$. Table 3 contains likelihood ratio, Wald and LM tests of these restrictions based on our estimated generalized censored probit model. Based on each of our three tests, in all cases our model rejects the restrictions implied by the censored probit model as well as those implied by the Poirier probit model. We can thus conclude that the Poirier probit specification, which treats repeated UN and OLF statuses as indistinguishable, is not the best representation of the labor market environment represented by our data. Indeed, our results suggest that correct inferences regarding individuals' true labor force statuses can be drawn by treating reported EMP, UN, and OLF separately. However, our results also indicate that the censored probit specification of these three states is insufficiently general to acceptably capture the classification ambiguities inherent in our (and all similar) data sets. By constraining reported UN and OLF statuses to be true indications of a woman's labor force situation, the censored probit specification ignores the very real likelihood that women with characteristics indistinguishable from OLF (UN) individuals are misclassified as UN (OLF).[28]

[28] For completeness, we report maximum likelihood estimates of the censored and Poirier probit models in Appendix B (Appendix Table 1). As discussed above, the Poirier probit model may not identify LF and EMP equations. We have ascribed the estimates to be LF or EMP due to their similarity to the generalized censored probit estimates. For the most part, the inferences that may be drawn from these results are comparable to those arising from the generalized censored probit model and will not be discussed in detail. But a noticeable difference is that race (black) is not an important explanatory variable in the labor-force equation estimated by the censored probit, while it is in the same equation estimated by the Poirier probit. Overall, the Poirier probit results are similar to those obtained from the generalized censored probit.

TABLE 3
TESTS FOR RESTRICTED MODELS

Tests for Restricted Models	The Poirier Model ($P_1 = P_2$) ($df = 17$)	The Censored Model ($P_1 = 1$ and $P_2 = 0$) ($df = 34$)
LR	47.1 ($p = 0.00$)[1]	137.3 ($p = 0.00$)
Wald	24.9[2] ($p = 0.10$)	738.3[3] ($p = 0.00$)
LM	68.2[2] ($p = 0.00$)	182.4[3] ($p = 0.00$)

Notes: 1) p-values are in the parentheses.
2) Tests for the restriction $\gamma_1 = \gamma_2$.
3) The censored model is equivalent to the general model with two sets of restrictions on γ_1 and γ_2. The first set of restriction is that the constant term in γ_1 is arbitrarily large while all other coefficients in γ_1 equal zeros. Another set of restrictions is that the constant term in γ_2 is a negative number whose absolute value is large while all other coefficients in γ_2 equal zeros. We chose three for the constant term in γ_1 and negative three for the constant term in γ_2. The computed Wald and LM statistics are for testing these restrictions.

Having shown that appropriate modeling of labor force statuses requires the distinction among the three reported states, EMP, UN, and OLF, and that classification ambiguities are likely to be present, it is important to determine the degree to which the estimated misclassification of UN (OLF) women as OLF (UN) results in over or under estimates of unemployment rates. As discussed above, if the sole classification ambiguity arises due to misclassified UN women ($P_1 < 1$), estimates of the unemployment rate would be overstated. Conversely, if the sole classification ambiguity arises among OLF women ($P_2 > 0$), estimated unemployment rates would be too low. Finally, if both classification ambiguities are present, estimates of the unemployment rate could be either too high or too low depending on the relative degree of misclassification.

Predicted probabilities generated from our generalized censored probit estimates are presented in Table 4. Panel A contains conditional probability estimates along with standard errors and 95% confidence intervals. We estimate that P_1, the probability that a woman with attributes consistent with true UN status is correctly reported as UN, equals 47%, with confidence bounds of 23 to 71%. The comparable estimate of P_2, the likelihood that a woman with OLF characteristics is misclassified as UN, equals approximately

TABLE 4

PROBABILITIES PREDICTED BASED ON THE GENERALIZED CENSORED PROBIT ESTIMATES

A. Conditional Probabilities

	From Generalized Censored Probit Estimates
Conditional Prob. of reported UN given UN attributes (P_1)	0.470[1] [0.124][2] {0.227, 0.713}[3]
Conditional Prob. of reported UN given OLF attributes (P_2)	0.121[4] [0.038] {0.047, 0.195}
Conditional Prob. of UN attributes given reported UN	0.492[5] [0.074] {0.347, 0.637}
Conditional Prob. of UN attributes given reported OLF	0.224[6] [0.083] {0.061, 0.387}

B. Unconditional Probabilities

	From Sample	From Generalized Censored Probit Estimates
Unconditional Prob. of Employment	0.695	0.695[7] [0.008] {0.677, 0.711}
Unconditional Prob. of Unemployment	0.044	0.076[8] [0.016] {0.045, 0.107}
Unconditional Prob. of OLF	0.262	0.229[9] [0.017] {0.196, 0.262}

Notes: 1) Computed by the sample mean of P_1 for all nonworkers in the sample.
2) Asymptotic standard errors are in the square brackets [ ].
3) 95% confidence intervals are in the braces { }.
4) Computed by the sample mean of P_2 for all nonworkers in the sample.
5) Computed by the sample mean of

$$\frac{P_1\{\Phi(X\beta_{lf})-F(X\beta_{lf}, X\beta_{emp}, \rho)\}}{[P_1\{\Phi(X\beta_{lf})-F(X\beta_{lf}, X\beta_{emp}, \rho)\}+P_2\{1-F(X\beta_{lf}, X\beta_{emp}, \rho)\}]}$$

for all nonworkers in the sample.
6) Computed by the sample mean of

$$\frac{(1-P_1)\{\Phi(X\beta_{lf})-F(X\beta_{lf}, X\beta_{emp}, \rho)\}}{[(1-P_1)\{\Phi(X\beta_{lf})-F(X\beta_{lf}, X\beta_{emp}, \rho)\}+(1-P_2)\{1-F(X\beta_{lf}, X\beta_{emp}, \rho)\}]}$$

for all nonworkers in the sample.
7) Computed by the sample mean of $F(X\beta_{lf}, X\beta_{emp}, \rho)$ for all sample observations.
8) Computed by the sample mean of $[\Phi(X\beta_{lf})-F(X\beta_{lf}, X\beta_{emp}, \rho)]$ for all sample observations.
9) Computed by the sample mean of $[1-F(X\beta_{lf}, X\beta_{emp}, \rho)]$ for all sample observations.

12%, with 95% confidence bounds of 5 to 20%. Thus our estimates imply classification ambiguities for both groups, as P_1 is significantly less than one while P_2 significantly exceeds zero.[29]

The impact of these estimated classification ambiguities on projected unconditional probabilities of EMP, UN, and OLF is presented in Panel B of Table 4. Since our generalized censored probit model did not consider potential misclassification of EMP, the sample and estimated probability of EMP both equal 69.5%. While the sample proportion of UN equals 4.4%, our model estimates that, after adjusting for potential classification ambiguities, the actual unemployment rate can increase to 7.6% (with 95% confidence bounds of 4.5 to 10.7%). This result indicates that the failure to control for misclassification of UN and OLF women in net may result in an underestimate of the proportion who are UN of nearly 73% of the sample proportion of UN. This underestimate of UN is, by definition, offset by an overestimate of the probability of OLF. In contrast to the sample proportion of 26.2%, our estimate of this likelihood is 22.9% (with confidence bounds from 19.6 to 26.2%). Our result demonstrates that failure to adjust estimates of UN and OLF for potential misclassification of women's statuses may lead to incorrect inferences as to the degree of UN (OLF) in the labor market.

V. Conclusion

This paper has developed a generalized censored probit likelihood function that nests both the Poirier probit and censored probit as special cases. The Poirier probit approach permits the labor market to be categorized by two distinct states — either employed or not employed — while the censored probit approach permits three distinct states — employed, unemployed, or out-of-labor force. Thus the generalized censored probit likelihood function that we develop permits us to determine if the labor market may be more appropriately categorized by two distinct states or by three. In

[29] It is also possible to compute the conditional probabilities that an individual with characteristics consistent with true UN status is reported as UN or OLF. The last two rows of Panel A of Table 4 find these estimates to be 49% and 22%, respectively, and also provide confidence bounds for these estimates.

addition, our generalized censored probit model permits parameterization and estimation of the degree of empirical ambiguity between individuals reported as UN or OLF relative to individuals with characteristics distinguishing them as truly UN or OLF.

Our empirical results demonstrate that distinguishing between reported UN and OLF is useful to identify and estimate individuals' labor force and employment decisions. However, we also find that the censored probit specification of these two decisions would not be adequate as it fails to consider possible sample classification ambiguities. Our generalized censored probit estimates show that the impacts of most covariates are consistent with our expectations. Further, the model specification utilized satisfies both the Hausman and Hausman Score tests as an appropriate formulation of the model. Estimates of UN and OLF probabilities based on our model show that failure to incorporate classification ambiguities results in UN rates that are understated and OLF rates that are overstated.

Our generalized censored probit model provides a new tool with which to determine the number of independent states that underlie an economic process and the degree of classification ambiguity present in sample data. When either issue is of interest, or when more accurate measures of event likelihoods in the presence of possible misclassifications is desired, estimation of our generalized censored probit model would be appropriate.

Our model could be further generalized at least in two directions. First, we have assumed that individual workers' classified unemployment or out of labor force statuses could be misclassified, while their classified employment statuses are the same as their true employment statuses. This assumption could be relaxed by allowing potential classification errors in employment status. Second, our model has assumed that the latent variables determining the conditional probabilities of misclassification errors are uncorrelated with the error terms in the labor-force and employment index equations. This assumption could be relaxed by allowing correlations among those unobservable variables. Each of these generalizations would make the model very complicated and difficult to estimate by the conventional maximum-likelihood approach that we use, although. Markov Chain Monte Carlo (MCMC) methods might be helpful for the estimation of these generalized models. The application of the MCMC method to these generalized models would be an intriguing future research topic.

(Received 3 November 2006; Revised 16 December 2006)

Appendix A

In this appendix, we explain how the two specification tests introduced in Section II-*C* can be conducted in empirical studies. The two statistics HT_g and HST_g are straightforward extensions of Newey (1987) and Peters and Smith (1991).

For notational convenience, we use $\lambda_o = (\theta'_o, \gamma_o)'$ to denote the true value of the parameter vector λ for the generalized censored probit model. In addition, we use subscripts "p" and "g" to refer to the Poirier and generalized censored probit models, respectively. Thus $\hat{\theta}_p$ and $\hat{\lambda}_g = (\hat{\theta}'_g, \hat{\gamma}'_g)'$ indicate the maximum-likelihood estimators of θ and λ for the Poirier and generalized censored probit models, respectively, while $s_p(\theta) = \partial l_p(\theta)/\partial\theta$ and $s_g(\lambda) = \partial l_g(\lambda)/\partial\lambda$ denote score vectors. For future use, we denote score vectors for individual i by $s_{pi}(\theta)$ and $s_{gi}(\lambda)$. We define the Hessian matrices for the models by $H_p(\theta) = \partial s_p(\theta)/\partial\theta'$ and $H_g(\lambda) = \partial s_g(\lambda)/\partial\lambda'$. We also define corresponding information matrices by $J_p(\theta) = [-H_p(\theta)]^{-1}$ and $J_g(\lambda) = [-H_g(\lambda)]^{-1}$. Then the variance-covariance matrices $V(\hat{\theta}_p)$ and $V(\hat{\lambda}_g)$, can be consistently estimated by $J_p(\hat{\theta}_p)$ and $J_g(\hat{\lambda}_g)$, respectively. Finally, letting $M = (I_q, \ 0_{qxr})$ where q and r are the number of parameters in θ and γ, respectively, $V(\hat{\theta}_g) = MJ_g(\hat{\lambda}_g)M'$.

The foundation of the Hausman statistic HT_g is the fact that $\hat{\theta}_p$ is a consistent estimator of θ_o if the generalized censored probit model (9) is correctly specified. Thus, under the null hypothesis (say, H_o^g) that the general model is correctly specified, the difference between $\hat{\theta}_p$ and $\hat{\theta}_g$ should be small. This observation motivates use of HT_g. As mentioned in Section II-*B*, which parts of $\hat{\theta}_p = (\hat{\beta}'_{lf,p}, \hat{\beta}'_{emp,p}, \hat{\rho}_p)'$ should be treated as the estimates of β_{lf} and β_{emp} cannot be determined in the Poirier model. A possible treatment for this problem is to compare $\hat{\theta}_p$ and $\hat{\theta}_g$ and choose the part of $\hat{\theta}_p$ close to $\hat{\beta}_{lf,g}(\hat{\beta}_{emp,g})$ as the Poirier estimate of $\beta_{lf}(\beta_{emp})$. As one might correctly point out, this method could be biased toward acceptance of the general model. (Equivalently, the test statistic computed choosing a different part of $\hat{\theta}_p$ as the Poirier estimate of β_{lf} would be biased toward rejection of the model.) Thus, the Hausman test result should be interpreted with some caution. It is worth noting that this problem does not apply to the Hausman score test introduced below.

In practice, the variance-covariance matrix $V(\hat{\theta}_p - \hat{\theta}_g)$ must be estimated. Following Hausman (1978), it can be shown that $V(\hat{\theta}_p - \hat{\theta}_g) = V(\hat{\theta}_p) - V(\hat{\theta}_g)$. Thus, $V(\hat{\theta}_p - \hat{\theta}_g)$ can be easily estimated by the difference between $J_p(\hat{\theta}_p)$ and $MJ_g(\hat{\lambda}_g)M'$. Unfortunately, however, this estimate is not necessarily positive definite, and the Hausman statistic computed with this estimate could be negative. In order to avoid this problem, we estimate $V(\hat{\theta}_p - \hat{\theta}_g)$ following Newey (1987, p. 130). Define:

$$B_g(\hat{\theta}_p, \hat{\lambda}_g) = [J_p(\hat{\theta}_p), \ -MJ_g(\hat{\lambda}_g)]; \ D_g(\hat{\theta}_p, \hat{\lambda}_g) = \sum_i d_{g,i}(\hat{\theta}_p, \hat{\lambda}_g)' d_{g,i}(\hat{\theta}_p, \hat{\lambda}_g),$$

where $d_{g,i}(\hat{\theta}_p, \hat{\lambda}_g) = [s_{p,i}(\hat{\theta}_p)', s_{g,i}(\hat{\lambda}_g)']$. Then, it can be shown that:

$$\hat{V}(\hat{\theta}_p - \hat{\theta}_g) = B_g(\hat{\theta}_p, \hat{\lambda}_g) D_g(\hat{\theta}_p, \hat{\lambda}_g) B_g(\hat{\theta}_p, \hat{\lambda}_q)' \qquad \text{(A.1)}$$

is a consistent estimator of $V(\hat{\theta}_p - \hat{\theta}_g)$.

The HST_g statistic is motivated by the fact that under H_o^g, $E[s_{p,i}(\theta_o)] = 0$. Thus if H_o^g is correct, $N^{-1}s_p(\hat{\theta}_g)$ should be close to zero, since $\hat{\theta}_g$ is a consistent estimator of θ_o. Accordingly, any significant deviation of $N^{-1}s_p(\hat{\theta}_g)$ from zero can be regarded as a sign of misspecification. Empirical use of the HST_g statistic requires estimation of $V[s_p(\hat{\theta}_g)]$. However, following Peters and Smith (1991, pp. 181-2), the variance-covariance matrix $V[s_p(\hat{\theta}_g)]$ can be consistently estimated by:

$$\hat{V}[s_p(\hat{\theta}_p)] = [I_q, \ H_p(\hat{\theta}_p)MJ_g(\hat{\lambda}_g)] D_g(\hat{\theta}_g, \ \hat{\lambda}_g) [I_q, \ H_p(\hat{\theta}_p)MJ_g(\hat{\lambda}_g)]'. \qquad \text{(A.2)}$$

Note that HST_g is computed by using $\hat{\lambda}_g$ only. In contrast to HT_g, it does not require computation of $\hat{\theta}_p$. Nonetheless, following Peters and Smith, it can be shown that the two statistics are asymptotically identical under H_o^g.

Appendix B

APPENDIX TABLE 1

CENSORED AND POIRIER PROBIT ESTIMATES

Regressors	Censored Probit Model		Poirier Probit Model		
	Labor Force	Employment	Labor Force[1)]	Employment[2)]	Prob. of UN Given not EMP with restriction $P_1 = P_2$
Constant	2.747* (14.58)[3)]	0.541 (1.264)	1.476* (3.365)	3.219* (7.166)	2.084* (5.136)
HSGRAD	0.172** (2.314)	0.357* (2.597)	0.303** (2.263)	0.045 (0.300)	-0.092 (0.630)
COGRAD	0.622* (6.206)	0.404** (1.729)	0.399** (2.125)	0.547* (2.651)	0.345 (1.524)
KIDS5	-0.512* (12.21)	-0.034 (0.205)	-0.424* (4.835)	-0.346** (2.542)	-0.445* (5.144)
KIDS17	-0.053** (1.851)	-0.005 (0.115)	-0.061 (0.958)	-0.079 (1.230)	-0.004 (0.063)
BLACK	0.081 (1.060)	-0.438* (3.759)	0.426** (2.008)	-0.474* (2.719)	0.627* (3.995)
AGE	-0.059* (15.53)	0.016 (0.781)	-0.033* (1.873)	-0.053* (6.361)	-0.071* (7.073)
EXP	0.081* (15.52)	0.010 (0.364)	0.157 (4.989)	0.030** (2.537)	0.062* (5.171)
SMSA	-0.118** (1.884)	0.204** (1.921)	-0.209* (1.854)	0.123 (1.030)	-0.358* (2.611)
REGNC	0.041 (0.441)	0.185 (1.262)	0.061 (0.386)	0.097 (0.505)	-0.239 (1.267)
REGS	-0.002 (0.028)	0.378* (2.741)	0.183 (1.201)	0.046 (0.271)	-0.502* (2.766)
REGW	-0.106 (1.098)	0.466* (2.578)	0.314 (1.563)	-0.283 (1.531)	-0.767* (3.328)
WNLINC	-0.011 (0.108)	0.021 (0.084)	0.405 (0.922)	-0.070 (0.670)	-0.199 (0.666)
HLINC	-0.057* (4.690)	0.028 (0.820)	-0.107 (4.845)	0.005 (0.156)	-0.092** (2.391)
HNLINC	-0.017 (0.426)	0.297** (2.032)	0.231 (1.486)	-0.143 (2.186)	-0.435** (1.824)
WPHLIM	-0.381* (4.104)	-0.537* (2.632)	-0.081 (0.252)	-0.649** (4.276)	0.155 (0.895)
UNEMPR	-5.940* (4.480)	-2.065 (0.718)	-6.924* (2.701)	-2.632 (1.047)	-3.649 (1.350)
ρ	-0.133 (0.176)		-0.047 (0.053)		
Log-likelihood	-1650.1		-1336.2		-268.7
# of observation	2,706		2,706		826

Notes: 1), 2) Chosen compared with the generalized probit results.
3) Absolute value of t-statistic in parentheses.
4) *Significant at $\alpha = .01$ (two-tail test).
**Significant at $\alpha = .10$ (two-tail test).

References

Ahn, S. C. Three Essays on Share Contracts, Labor Supply, and the Estimation of Models for Dynamic Panel Data. Unpublished Ph.D. Dissertation, Michigan State University, 1990.

Blundell, R., Ham, J., and Meghir, C. "Unemployment and Female Labor Supply." *Economic Journal* 97 (Supplment 1987): 44-64.

__________. "Unemployment, Discouraged Workers, and Female Labor Supply." *Research in Economics* 52 (1998): 103-31.

Blundell, R., and Meghir, C. "Bivariate Alternatives to the Tobit Model." *Journal of Econometrics* 34 (Nos. 1-2 1987): 179-200.

Bowlus, A. J. A Search Interpretation of Male-Female Wage Differentials. Unpublished Manuscript, University of Western Ontario, 1995.

Burdett, K., and Mortensen, D. "Search, Layoffs, and Labor Market Equilibrium." *Journal of Political Economy* 88 (No. 4 1980): 652-72.

Burgess, P. "Compliance with Unemployment Insurance Job Search Requirements." *Journal of Law and Economics* 35 (No. 2 1992): 371-96.

Caudill, S. B., and Mixon, F. G. Jr. "Analysing Misleading Discrete Responses: A Logit Model Based on Misclassified Data." *Oxford Bulletin of Economics and Statistics* 67 (No. 1 2005): 105-13.

Clark, K. B., and Summers, L. H. *The Dynamics of Youth Unemployment, in the Youth Labor Market Problems: Its Nature, Causes and Consequences*. in R. Freeman and D. Wise (ed.), The University of Chicago Press, pp. 199-234, 1982.

Devine, T. J., and Kiefer, N. M. *Empirical Labor Economics: The Search Approach*. Oxford University Press, 1991.

Ehrenberg, R. G., and Smith, R. S. *Modern Labor Economics: Theory and Public Policy*. 3*rd* Edition, Glenview, Illinois: Scott, Foresman and Company, 1987.

Farber, H. S. "Worker Preference for Union Representation." *Research in Labor Economics* 2 (Supplment 1983): 171-205.

Filer, R. K., Hamermesh, D. S., and Rees, A. E. *The Economics of Work and Pay*. 6*th* Edition, New York: Harper Collins College

Publishers, 1996.

Flaim, P. O., and Hogue, C. R. "Measuring Labor Force Flows: A Special Conference." *Monthly Labor Review* 108 (July 1985): 7-17.

Flinn, C. J., and Heckman, J. "Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?" *Journal of Labor Economics* 1 (No. 1 1982): 169-91.

Gönül, F. "New Evidence on Whether Unemployment and Out of Labor Force Are Distinct States." *The Journal of Human Resources* 27 (No. 2 1992): 329-61.

Hausman, J. A. "Specification Tests in Econometrics." *Econometrica* 46 (No. 6 1978): 1251-71.

__________. "The Effect of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation." *Journal of Public Economics* 14 (No. 2 1980): 161-94.

Hausman, J., Abrevaya, J., and Scoot-Morton, F. M. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics* 87 (No. 2 1998): 239-269.

Heckman, J. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42 (No. 4 1974): 670-94.

Lewbel, A. "Identification of the Binary Choice Model with Misclassification." *Econometric Theory* 16 (No. 4 2000): 603-9.

Meng, C., and Schmidt, P. "On the Cost of Partial Observability in the Bivariate Probit Model." *International Economic Review* 26 (No. 1 1985): 71-85.

Newey, W. K. "Generalized Method of Moments Specification Testing." *Journal of Econometrics* 29 (No. 3 1985): 229-56.

__________. "Specification Tests for Distributional Assumptions in the Tobit Model." *Journal of Econometrics* 34 (Nos. 1-2 1987): 125-45.

Peters, S., and Smith, R. J. "Distributional Specification Tests against Semiparametric Alternatives." *Journal of Econometrics* 47 (No. 1 1991): 175-94.

Poirier, D. J. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics* 12 (No. 2 1980): 209-17.

Sundt, L. The Role of Employment Opportunities in Determining Employment Status. Unpublished Manuscript, University of Arizona, 1990.

Comments and Discussion

*Comments by Dae Il Kim**

The authors attempt in this paper to identify the distinction among various labor market statuses using a simple parametric model allowing for coding (reporting) errors. In particular, they focus on the classification of the "unemployed" and "non-participating," and use the possible non-randomness of errors in reported labor market statuses to identify their model. The primary result is that "being unemployed" is distinguishable from "being out of labor force," regardless of whether these statuses are reported or identified from the model after correcting for coding errors. However, the reported unemployment tends to understate the "actual" unemployment due to the coding errors.

The authors motivate an important and interesting discussion in the paper and propose quite a simple and clear-cut econometric framework to handle the issue. Their exposition is so clear that even a reader without deep knowledge in econometrics like myself can understand the issue and their logics with no difficulty. I must applaud them for their excellent job of producing the results and making them understood easily.

That being said, I wish to suggest a few points hoping that they could help improve the paper. First, the authors have done a nice job of identifying the possible coding errors in reported labor market statuses, but they need to go one step further to show that their "corrected" statuses make more senses than the "reported" statues. For example, they may wish to compare the next period's job access rates of those who report themselves as out of labor force but are identified as highly probable to be an unemployed worker from the model to the rates of those who report themselves and are identified as less so. The authors should expect to see a significantly higher job access rate among the former than among the latter if their correction of coding errors is a valid one. Alternatively, the authors

* Professor, School of Economics, Seoul National University, Seoul 151-746, Korea, (Tel) +82-2-880-6364, (E-mail) dikim@snu.ac.kr.

may wish the job access rates of those who report themselves as out of labor force but are identified as highly probable to be an unemployed worker from the model to the rates of those who report themselves and are identified as unemployed. The authors should expect to see no significant differences between these groups. Given that the authors are using a panel data, this is readily doable.

Second, I would like to ask the authors to explain a little bit more on how the coefficients in the reporting error probability regressions (γ_1 and γ_2 in P_1 and P_2 regressions) are "connected" to the coefficients in participation and employment equations. This is a meaningful question because the authors use the same regressors in all equations in the model. In particular, when Poirier probit (Appendix B) is compared to the authors' model (Table 2), the major differences are in the estimates of *KIDS17* and *BLACK*. The race variable (*BLACK*) in participation equation has a smaller coefficient in the authors' model and has a large positive coefficient in P_2 regression. This can be interpreted as indicating that black women are in principle not so much participating but are likely to be misclassified as unemployed. However, no such clear-cut interpretation is readily available for *KIDS17* variable. Insignificant coefficients on *KIDS17* in P_1 and P_2 regressions appear to imply that there is no serious mis-coding connected to this variable, but it suddenly has a significant coefficient in the authors' participation equation. Why?

Third, the authors use an old single year sample, and I think they need to expand their analysis to a longer and more recent period because there is no reason to believe that the structure of coding errors is stable over time, or through business cycles. Further the sample is limited to married women, whose reporting errors, if any, may not have a similar structure as men or single women who are more strongly attached to the market. Indeed, the authors' result that "true" unemployment is much higher than "reported" may not be readily extend to other population groups.

Lastly, I do not quite understand why various non-labor income variables widely differ in their effects on women's labor market behavior. The results suggest that the income sources, not just amounts, matter. I do not deny the possibility, but wish to see an intuitive explanation for that.

Again, I very much enjoyed reading the paper, and wish to thank the authors for their very interesting paper.

*Comments by Keunkwan Ryu**

It is often argued that there is no clear boundary between unemployment and out of labor force, which is partly evidenced by workers who are observed to enter employment directly from out of labor force. Considering potential misclassification between unemployment and out of labor force, they model the observed labor market statuses and present several tests for their model.

Let me first review their model and testing ideas. There are two latent variables, say y_1^* and y_2^*. True (latent) labor market statuses are determined by the signs of these two latent variables. The sign of y_1^* determines true labor force participation: In labor force (+) or out of labor force (−). Given $y_1^*>0$, the sign of y_2^* determines true employment status: Employment (+) or unemployment (−). They model that the error terms in these two equations can be correlated.

Allowing for potential misclassification between unemployment and out of labor force, they introduce two conditional probabilities as follows: P_1 is the probability of correctly classifying a true unemployed worker as unemployed, and P_2 is the probability of incorrectly misclassifying a person out of labor force as unemployed. These probabilities are set as functions of their observed characteristics.

To simplify the notation, let OLF=out of labor force, UN=unemployment, and EMP=employment. By attaching *, let us denote the corresponding true state (ones without * denote observed states.)

Depending on whether to allow for misclassification possibility and whether to partition unemployment and out of labor force, three different ways of assigning probabilities to the observed labor market states are possible. They are as follows.

L_1 (no allowance for misclassification, separate treatment of OLF and UN):
The likelihood function is based on $P(OLF)=P(OLF^*)=P(y_1^*<0)$, $P(UN)=P(UN^*)=P(y_1^*>0,\ y_2^*<0)$, and $P(EMP)=P(EMP^*)=P(y_1^*>0,\ y_2^*>0)$.

L_2 (misclassification or not, combined treatment of OLF and UN):

* Professor, School of Economics, Seoul National University, Seoul 151-746, Korea, (Tel) +82-2-880-6397, (E-mail) ryu@snu.ac.kr.

The likelihood function is based on $P(OLF$ or $UN)=P(OLF^*$ or $UN^*)$ and $P(EMP)=P(EMP^*)$.

L_3 (allowance for misclassification, separate treatment of OLF and UN): The likelihood function is based on $P(OLF)=P(OLF^*)(1-P_2)+P(UN^*)(1-P_1)$, $P(UN)=P(OLF^*)P_2+P(UN^*)P_1$ and $P(EMP)=P(EMP^*)$. This approach is robust to misclassification and efficient as well.

Let us study the relationship among the above three approaches. First, if there is no misclassification, L_3 degenerates to L_1. Second, L_2 is robust to misclassification, but less efficient than either L_1 or L_3 due to non-separation of OLF and UN.

One can design several test statistics by utilizing the relationships among the three likelihood approaches. First, by comparing L_2 and L_3, one can design a model specification test using the Hausman's test idea testing whether the assumed misclassification is the correct scheme.

In fact, this test has been suggested in their paper. One can think of the following two other tests as well. Second, by comparing L_1 and L_2, one can test the null hypothesis whether there is misclassification. Finally, by comparing L_1 and L_3, one can test whether there is misclassification and/or whether the assumed misclassification scheme is correct. The final two tests are not mentioned in their paper, but would be useful addition to the first test.

As a numerical matter, they may consider Markov Chain Monte Carlo using "data augmentation plus Gibbs sampling" to replace the current numerical maximization. For this purpose, they may augment the likelihood function to include y_1^*, y_2^*, and true classification in addition to the model parameters, and then apply Gibbs sampling scheme.

The empirical results of their paper are interesting and convincing.