VALUE OF SAMPLE SEPARATION INFORMATION IN A SEQUENTIAL PROBIT MODEL

Another Look at SSA's Disability Determination Process

C. GAO,[a] K. LAHIRI[a,b]* AND B. WIXON[b]

[a]Department of Economics, University at Albany-SUNY,
Albany, NY 12222, USA
[b] Division of Economic Research, Social Security Administration,
Washington, DC 20254, USA

* Correspondence to: Kajal Lahiri, Department of Economics, University at Albany-SUNY, 1400 Washington Avenue, Albany, NY 12222, USA.
Email:KL758@cnsvax.Albany.edu

SUMMARY

We have estimated a 4-step sequential probit model with and without sample separation information to characterize SSA's disability determination process. Under the program provisions, different criteria dictate the outcomes at different steps of the process. We used data on health, activity limitations, demographic traits, and work from 1990 SIPP exact matched to SSA administrative records on disability determinations. Using GHK Monte Carlo simulation technique, our estimation results suggest that the correlations in errors across equations that may arise due to unobserved individual heterogeneity are not statistically significant. In addition, we examined the value of administrative data on the basis for allow/deny determinations at each stage of the process. Following the marginal likelihood approach adopted by Benitez-Silva, Buchinsky, Chan, Rust, and Sheidvasser (1999), we also estimated the above sequential probit model without the sample separation information for the purpose of direct comparison. We found that without the detailed administrative information on outcomes at each stage of the screening process, we could not properly evaluate the importance of a large number of program-relevant survey-based explanatory variables. In terms of both in-sample and jackknife-type out-of-sample predictive analysis, the value of modeling the sequential structure of the determination process in generating correct eligibility probabilities is confirmed.

1. INTRODUCTION

Social Security Administration's (SSA's) two disability programs – Disability Insurance (DI) and Supplemental Security Income (SSI) – provide cash and medical benefits to people with long-term disability. At the end of 1997, these programs paid benefits exceeding $70 billion a year to over 10.3 million beneficiaries. SSA administers these programs with the help of the State Disability Determination Services (DDSs), which make the initial determination on whether the claimants meet the programs' statutory definition of disability. Lahiri, Vaughan and Wixon (1995, hereafter LVW) first modeled the SSA's disability determination process using a multistage sequential logit model, mimicking exactly the steps of the determination process used by the DDSs. By matching SSA administrative data on disability determinations to household survey information from the Survey of Income and Program Participation (SIPP), they found that the detailed administrative information on outcomes at each step of the determination process is crucial in identifying the role of many important survey-based variables. They also found that the predictions of overall allowance rates from the sequential model performed considerably better than those based on a naive allow/deny logit regression. In a recent interesting paper, Benitez-Silva, Buchinsky, Chan, Rust, and Sheidvasser (1999, hereafter BSBCRS) have modeled the entire disability program by incorporating not only the initial DDS determinations, but also the application and the appeal processes using data from the first three waves of the Health and Retirement Survey (HRS). In the absence of administrative data on the reasons for disability allowance or rejection at the DDS level, they identified the probability of disqualification at each stage of the

sequential process by using what they called the “marginal likelihood” method. They found that the predicted probabilities from their marginal likelihood approach are quite comparable to those obtained by LVW (1995), implying that one does not need to have administrative data on the exact stage at which an application was allowed or denied to accurately estimate the overall acceptance probability. The main purpose of the present paper is to reexamine this issue in the context of a more general econometric model than used by LVW (1995), but utilizing the same data.

As serious social security reform looms in the near future, increasing the mandatory retirement age becomes an inevitable alternative. However, the retirement and disability behaviors of persons approaching mandatory retirement are highly interrelated. Thus, for a comprehensive social security reform, one needs to understand the dynamics of disability enrolment growth. It is now well understood that for the purposes of disability modeling and forecasting, it is crucial to estimate accurately the eligibility probabilities for prospective disability applicants, see Aarts and De Jong (1992). From a methodological standpoint, econometric analysis of multi-step sequential models with incomplete sample separation information has not been very common. Interestingly, the models formulated by LVW (1995) and BSBCRS (1999) are multivariate logit models with varying degree of observability in the decision variables. Previous literature on the value of sample separation information in the study of switching regression models, disequilibrium models, and double-hurdle models found that accurate sample separation information improves the efficiency of estimation.[1] In particular, Poirier (1980) and Meng and Schmidt (1985) analyzed the cost of partial observability in bivariate probit models,

and concluded that it is worthwhile to obtain the additional information. However, all previous research in this area considered only relatively simple bivariate models. The BSBCRS (1999) study is pioneering in that it considers a 4-level sequential logit model with incomplete sample separation information. The intricacies in the estimation of this class of models is not well understood, particularly when one likes to allow for cross-equation correlation in errors.

Due to the strict sequential nature of the disability determination process, it is logical to think that the cross-equation correlations in errors can be ignored. However, non-zero correlations may result if there is significant unobserved individual heterogeneity in the equations due to misspecification. In order to study this issue, we will first extend the work of LVW (1995) by allowing for possible cross-equation correlations in a multivariate sequential probit framework. Since we model the last four steps of the disability determination process with five possible outcomes, the sequential probit model generates decision probabilities that are multivariate integrals of order up to four. In general, the evaluation of such integrals has been computationally difficult except in very special cases. However, recent developments in Monte Carlo simulation techniques have facilitated the estimation of multivariate probit models. According to the simulation experiments by Hajivassiliou, McFadden and Ruud (1996), the GHK simulator [cf., Geweke (1989), Hajivassiliou (1993) and Keane (1994)] appears to be the most reliable method for simulating orthant probabilities. Using GHK, simulated maximum likelihood (SML) estimation shows that none of the off-diagonal elements in the covariance matrix to be significantly different from zero.

---

[1] See Goldfeld and Quandt (1975), Kiefer (1978), Blundell, Ham and Meghir (1978), Schmidt (1981), and

Second, we examine the value of the detailed administrative information in predicting disability determinations. Our approach is first to model the process using a sequential probit model with administrative sample separation information. We then compare its estimation and prediction results with those without the administrative information. The advantage of this experimental set-up is that we could control for everything except for the sample separation information.

The paper is organized as follows. In section 2 we briefly introduce the disability determination process and the data set used. Section 3 describes the sequential probit model with or without the sample separation information for the intermediate step decisions of the sequential process. We also allow for possible correlation among the structural disturbances. Section 4 reports the empirical results. We also examine the predictive capacity of these models through extensive in-sample and jackknife-type out-of-sample predictive analysis. Section 5 summarizes the conclusions.

2. DISABILITY DETERMINATION PROCESS AND THE DATA

The DDS makes accept/reject decisions according to a sequential five-step screening process as depicted in Figure 1. LVW (1995) provide a detailed description and logic of the SSA determination procedure. The first step is an earnings screen and is administered at one of nearly 1300 SSA field offices across the country. Applicants earning more than the substantial gainful activity (SGA) are denied immediately. Since this step of the

---

Jones (1989), among others.

process is not part of the medical-vocational determination made by DDS, it was not modeled in LVW(1995). We will also not model the first step in this paper.

Insert Figure 1 here.

The second step is to determine the severity of impairments. An applicant is denied if the impairments do not significantly limit the physical or mental ability to accomplish basic work-related activities. Applicants are also denied at step 2 if their impairments do not meet the duration test of 12 months. Under step 3, the medical evidence obtained on an applicant's impairment is assessed. If the applicant's impairment meets or equals the criteria of at least one of over 100 "listed" impairment classifications, the applicant is granted a *medical allowance* without further evaluation. Applicants neither denied at step 2 nor allowed at step 3 are evaluated in terms of their residual capacity and vocational factors at steps 4 and 5. An applicant is denied benefits at step 4 if he or she is judged able to perform past work. The remaining applicants, including those with no recent work experience, are evaluated at the fifth and final step for their residual functional capacity in conjunction with vocational factors. If an applicant is determined incapable of working in any job in the national economy, a *vocational allowance* is awarded; otherwise the applicant is given a vocational denial. For convenience, the five final outcomes are labeled as d_2, a_3, d_4, a_5, and d_5. Overall, an applicant is granted benefits if the outcome is either a_3 or a_5.

In this work, we use the study data derived in LVW (1995) from the 1990 panel (June 1990 - May 1992) of the Survey of Income and Program Participation (SIPP) exact matched to SSA disability determination records of SIPP sample members whose

applications were acted upon in the calendar years 1986-1993.[2] The final study sample consists of 1230 survey cases who also applied for DI or SSI adult disability benefits. Among them, 223 cases (18.1%) were denied benefits at step 2. At step 3, 359 cases were granted medical allowance and 648 cases were passed on to step 4. At step 4, 188 cases were denied for being able to do past work. At the final step, 213 cases were given vocational allowance and 247 cases received denial based on capacity for some work. The sets of sample observations are denoted as D_2, A_3, D_4, A_5, and D_5, corresponding to the five outcomes denoted by d_2, a_3, d_4, a_5, and d_5. LVW (1995) showed that the percentage distribution of the study sample by step in the sequential process and outcome closely approximates the pattern experienced by the full universe of applicants.[3]

3. THE 4- STEP SEQUENTIAL MODELS

Throughout the paper, we only model the last four steps of the disability determination process as shown in Figure 1. For any applicant denoted by j, let S_i be the latent criterion function involved at step i,

$$S_i = b\text{¢}_i X_i + u_i, \tag{1}$$

for i = 2,3,4,5. Here X_i is the set of individual characteristics that are used to evaluate the criterion function at step i, and b_i is the corresponding parameter vector. To simplify

[2] In order to match the timing of the SIPP responses more closely with the administrative records, we also conducted our analysis after deleting all applicants whose adjudication took place during 1986-1988, resulting in a sample size of 927. The results were very similar to what we report later in the paper.

notation, subscript j is dropped when it is not ambiguous. Given that the application is at step i, a favorable outcome from the standpoint of the applicant is assigned if $S_i > 0$, an unfavorable outcome is assigned otherwise. To accommodate possible correlation among decisions at multiple steps, we assume that (u_2, u_3, u_4, u_5) are *iid* N(0, W), and $W = (\omega_{km})$ is *pds*, for $k, m = 1,2,3,4$.

Now we consider the probability of an applicant attaining one of the five outcomes. At step 2, an applicant may be denied benefits with probability $P\{S_2 \leq 0\}$ and passed on to step 3 with probability $P\{S_2 > 0\}$. Conditional on the applicant not being denied at step 2, the applicant may be granted medical allowance at step 3 with probability $P\{S_3 > 0 \mid S_2 > 0\}$. So the probability of an applicant attaining outcome a_3 is given by

$$\begin{aligned} & P\{S_3 > 0 \mid S_2 > 0\}P\{S_2 > 0\} \\ = & P\{S_2 > 0, S_3 > 0\}. \end{aligned}$$

The applicant may be passed on to step 4 with probability $P\{S_2 > 0, S_3 \leq 0\}$. Then conditional on the applicant not being denied at step 2 and not being awarded medical allowance at step 3, he or she may be denied benefits at step 4 with probability $P\{S_4 \leq 0 \mid S_2 > 0, S_3 \leq 0\}$. So the probability of an applicant attaining outcome d_4 is given by

$$\begin{aligned} & P\{S_4 \leq 0 \mid S_2 > 0, S_3 \leq 0\}P\{S_2 > 0, S_3 \leq 0\} \\ = & P\{S_2 > 0, S_3 \leq 0, S_4 \leq 0\}. \end{aligned}$$

[3] For the derivation of the study sample and other details, see Lahiri, Vaughan, and Wixon (1995).

Similarly, the probability of an applicant attaining outcome a_5 is

$$P\{S_2 > 0, S_3 \leq 0, S_4 > 0, S_5 > 0\},$$

and the probability of an applicant attaining outcome d_5 is

$$P\{S_2 > 0, S_3 \leq 0, S_4 > 0, S_5 \leq 0\}.$$

In the following we will use these probabilities to write the likelihood functions for our sample under sample separation and without sample separation information, with general or restricted covariance matrix W of errors.

3.1 *Sample Separation Information Available*

Under this scenario, the step at which an applicant was accepted or denied benefits is known. The joint log-likelihood function incorporating this information is given by

$$\begin{aligned} L_{NP} = {} & \textstyle\sum_j \hat{\imath}_{D2} \ln P\{S_{2j} < 0\} + \sum_j \hat{\imath}_{A3} \ln P\{S_{2j} > 0, S_{3j} > 0\} \\ & + \textstyle\sum_j \hat{\imath}_{D4} \ln P\{S_{2j} > 0, S_{3j} \leq 0, S_{4j} \leq 0\} \\ & + \textstyle\sum_j \hat{\imath}_{A5} \ln P\{S_{2j} > 0, S_{3j} \leq 0, S_{4j} > 0, S_{5j} > 0\} \\ & + \textstyle\sum_j \hat{\imath}_{D5} \ln P\{S_{2j} > 0, S_{3j} \leq 0, S_{4j} > 0, S_{5j} \leq 0\}. \end{aligned} \tag{2}$$

Like many sequential response models, however, we have unbalanced observations across steps, i.e., the applicants who are adjudicated as allowed or denied at an earlier step are not evaluated any further. This gives rise to a multivariate probit model with partial observability. As a result, the summation of the log-probability for each outcome is taken only over the sample cases that have attained that final outcome, as denoted by the sets D_2, A_3, D_4, A_5, and D_5. The evaluation of the log-likelihood function L_{NP} in general involves higher-dimension multiple integrals of normal rectangle probabilities unless W assumes a simplified structure. We use GHK simulator to simulate the probabilities and evaluate the log-likelihood function.[4] Note that in expression (2), each of the integration of the probabilities for different outcomes involves only a top-left subset of W, depending on the dimension of the integral.

Obviously, when W is an identity matrix, the log-likelihood function in (2) is equivalent to that of a 4-stage step-wise sequential probit model:

$$
\begin{aligned}
L_{SSP} = & \left(\sum_j \hat{I}_{D2} \ln P\{S_{2j} \le 0\} + \sum_j \hat{I}_{A3}\hat{E}_{D4}\hat{E}_{A5}\hat{E}_{D5} \ln P\{S_{2j} > 0\}\right) \\
& + \left(\sum_j \hat{I}_{A3} \ln P\{S_{3j} > 0\} + \sum_j \hat{I}_{D4}\hat{E}_{A5}\hat{E}_{D5)} \ln P\{S_{3j} \le 0\}\right) \\
& + \left(\sum_j \hat{I}_{D4} \ln P\{S_{4j} \le 0\} + \sum_j \hat{I}_{A5}\hat{E}_{D5} \ln P\{S_{4j} > 0\}\right) \\
& + \left(\sum_j \hat{I}_{A5} \ln P\{S_{5j} > 0\} + \sum_j \hat{I}_{D5} \ln P\{S_{5j} \le 0\}\right). \qquad (3)
\end{aligned}
$$

Note that the expressions inside each pair of parentheses are independent of each other and may be evaluated separately, cf. LVW (1995).

[4] We used the GAUSS code, available on the Internet, provided by Hajivassiliou, McFadden and Ruud (1996).

3.2 *Sample Separation Information Unavailable*

Now suppose that we ignore the administrative information about the step of the disability determination process at which an applicant was allowed or denied benefits. This structure is similar to that of hurdle models. From the standpoint of an applicant, each step in the disability determination process is a hurdle.[5] The decision by a DDS to pass the applicant an intermediate hurdle is not observed by the researcher. Note that a conventional hurdle model involves a continuous observed dependent variable with truncation, see Blundell et al. (1987).

In the absence of sample separation information in the intermediate steps, we only observe whether an applicant was awarded or denied benefits. A disability applicant would be granted benefits if he or she attains outcomes a_3 or a_5; otherwise the applicant would be denied. Using the probabilities of an applicant attaining each of the five outcomes derived above, the probability of an applicant being granted allowance is given by

$$Y = P\{S_2 > 0, S_3 > 0\} + P\{S_2 > 0, S_3 \leq 0, S_4 > 0, S_5 > 0\}, \tag{4}$$

and the probability of being denied benefits is $(1 - Y)$. The joint log-likelihood function then becomes:

[5] Early advocates of double-hurdle models include Fisher (1962) and Cragg (1971). Jones (1989) considered a trivariate model when identification of non-starters and ex-smokers is given in a study of cigarette consumption.

$$L_{MH} = \sum_j \hat{I}_A \ln Y_j + \sum_j \hat{I}_D \ln (1 - Y_j), \tag{5}$$

where $A = A_3 \grave{E} A_5$ is the set of applicants granted benefits, $D = D_2 \grave{E} D_4 \grave{E} D_5$ is the set of applicants denied benefits.

The log-likelihood function (5) places no restriction on the error covariance matrix W. The involved multiple integrals may be evaluated using the GHK simulator, similar to that for (2). Identification of the parameters in (5) is, however, complicated. We conjecture that the property of the log-likelihood function in (5) resembles that of disequilibrium models with unknown sample separation. Even if the diagonal elements in W are normalized to be 1's, under certain circumstances, some of the off-diagonal elements may converge to certain values (not necessary to be +1 or –1 in the multivariate case) such that W becomes singular.[6]

If W is assumed to be an identity matrix, (5) gets simplified. In this case, $P\{S_2 \leq 0\}$ is the conditional probability of being denied at step 2, $P\{S_3 \leq 0\}$ is the conditional probability of being passed on to step 4, $P\{S_4 \leq 0\}$ is the conditional probability of being denied at step 4, and $P\{S_5 \leq 0\}$ is the conditional probability of being denied at step 5. Then under the assumption of no correlation and normality of the structural disturbances, the probability of an applicant being awarded benefits is given by

$$\begin{aligned} Y &= P\{S_2 > 0, S_3 > 0\} + P\{S_2 > 0, S_3 \leq 0, S_4 > 0, S_5 > 0\} \\ &= P\{S_2 > 0\}P\{S_3 > 0\} + P\{S_2 > 0\}P\{S_3 \leq 0\}P\{S_4 > 0\}P\{S_5 > 0\} \end{aligned}$$

[6] See Maddala (1983, pp. 299-302) for a review of the relevant literature.

$$= F(b_2 \cdot X_2)F(b_3 \cdot X_3) + F(b_2 \cdot X_2)F(-b_3 \cdot X_3)F(b_4 \cdot X_4)F(b_5 \cdot X_5), \quad (6)$$

where $F(.)$ denotes the cumulative standard normal density. The joint log-likelihood function correspondingly gets simplified, and is akin to the marginal likelihood function analyzed by BSBCRS (1999).

4. EMPIRICAL RESULTS

4.1 *With Sample Separation Information*

For this case, the log-likelihood function in (2) with general W (*Model 1*) is evaluated using SML using the GHK simulator. In the estimation, the explanatory variables used for each step are the same as in LVW (1995). Identification of the model requires us to impose restrictions on the parameter vector b_i's or the covariance matrix W.[7] To make results of this model comparable to the results reported in LVW (1995) using a multi-stage sequential logit model, we normalize the diagonal elements of W to be unity (i.e., $\omega_{kk} = 1$ for $k = 1,2,3,4$). This normalization, however, prevents us from directly using the convenient derivatives of the multivariate normal rectangle probabilities in the GHK simulator as derived by Hajivassiliou, McFadden and Ruud (1996) to estimate standard errors.[8] Instead, we rely on the maximum likelihood algorithm to generate standard errors

[7] Keane (1992) considered identification in multinomial probit models. Since we have distinct variables for the criterion function at each step, a simple normalization in each equation would be sufficient for identification. However, it is not sufficient to let just one of the diagonal elements to be 1, which is sufficient for the multinomial probit model. The reason is that here we have multiple levels.

[8] If the constant terms are alternatively normalized to be +1 or –1, analytical derivatives of the multivariate normal rectangle probabilities may be readily derived and implemented in the GHK simulator.

for all parameters. We first used a sampling size of 50 for the GHK simulator to obtain initial estimates and then fine-tuned the estimation finally with a sampling size of 500.

The empirical results for *Model 1* are reported in Table 1. Assuming general W, the estimated covariance matrix was

$$\begin{bmatrix} 1.000 & & & \\ 0.028\ (0.261) & 1.000 & & \\ -0.239\ (0.444) & 0.189\ (0.291) & 1.000 & \\ 0.458\ (0.413) & -0.350\ (0.262) & 0.257\ (0.394) & 1.000 \end{bmatrix}$$

with standard errors in parentheses. We find that none of the off-diagonal elements is statistically significant, with p-values ranging from 0.09 to 0.46. Meanwhile all the structural parameters have estimates and standard errors comparable to those reported in LVW (1995) using a multi-stage sequential logit model. This finding that W = I has an important policy implication: The decisions at multiple steps of the DDS disability determination process are not correlated and therefore there is no loss of efficiency if the process is modeled separately. It also suggests that, given the specified explanatory variables, the unobserved heterogeneity in the four equations is not very important.

In Table 1, we also report results from estimation of a multi-stage step-wise sequential probit model with identity covariance matrix, i.e., $W = I$ (*Model 2*), by evaluating the log-likelihood function in (3). We find that all the parameter estimates are statistically significant, most of which are significant at the 5% level of significance. Between Models

1 and 2, although some of the parameter estimates differ to some extent, the difference in the values of log-likelihood function is not statistically significant. A likelihood ratio test statistic for the null hypothesis of no correlation has a value of 2.85, which is statistically negligible for a sample of 1230 observations with 6 restrictions on the off-diagonal elements of the covariance matrix.

4.2 *Without Sample Separation Information*

In the previous subsection, we found that there is no significant correlation between decisions at multiple steps. Therefore, we will assume an identity covariance matrix in the following and the simplified log-likelihood function (6) will be evaluated. We found that one has to be very careful while maximizing this likelihood function. Due to the lack of data on dependent variables as anchors at each step, difficulty arises in the selection of explanatory variables. Noticeably, many of the variables that were statistically significant in LVW (1995) as well as in subsection 4.1 lost their significance. With a number of relatively less significant variables present in the specifications, the maximum likelihood algorithm may generate unreasonably large but insignificant parameter estimates for some variables or fail to compute the variance-covariance matrix for the estimated parameters even using BHHH algorithm. Having too many overlapping explanatory variables also tends to create problems. However, it is not essential that the explanatory variables in the four equations be completely non-overlapping, as in BSBCRS (1999). In order to accommodate the maximum number of statistically significant explanatory variables in the four specifications, we took first a “general to specific”, then a “specific to general” approach in identifying regressors for each step. First we excluded all variables that had

relatively large *p*-values in Table 1 until we found a set of variables such that the maximum likelihood algorithm converged normally and produced reasonable parameter estimates. Then we retained variables with statistically significant parameter estimates, which consisted of only 14 variables in the total. Next, we examined the incremental contribution of each of the remaining variables to this core specification. We added the variables that had significant parameter estimates in the second step. After some additional experiments, we found a set of 27 variables for which most of the parameter estimates are statistically significant in the combined model. A casual inspection of this set of 27 variables (6 in step 2, 10 in step 3, 5 in step 4, and 6 in step 5) reveals that they still very much capture the basic logic of the screen at each step. We should point out that in BSBCRS (1999, Table 3) a total of only 8 variables (including two year dummies) were statistically significant at the 10% level of significance. Of these eight variables, "Back problems" had a perverse sign at step 3 and "Divorced" of step 4 has no program specific meaning. Note that BSBCRS consolidated the last two steps of the determination process into one (their step 4), but had an additional step (SGA) at the beginning. Unfortunately, the SGA step in their analysis had no statistically significant explanatory variable at the 10% level. These results are consistent with those in Meng and Schmidt (1985), Blundell, Ham, and Meghir (1987), and Jones (1989) who found substantial loss of efficiency in the absence of sample separation information. Also, it should be obvious that without very strong prior information on the specification of the regressions at each step, it would virtually be impossible to estimate multi-stage sequential probit models when sample separation information is not available. Fortunately, in this empirical case

the program provisions clearly guide us in specifying different set of covariates at different steps of the determination process.

The maximum likelihood estimates without the administrative information on sample separation and with an identity covariance matrix (*Model 3*) are reported in Table 2. Recall that the total number of explanatory variables in Table 1 (i.e., with sample separation) is 56 excluding the constant terms, and each of them is significant at the 10% level of significance. There are only 27 explanatory variables in Table 2. Among them, parameter estimates for 22 variables (excluding the constant terms) have *p*-value less than 0.05. Using this restricted set of 27 variables, we also evaluated the log-likelihood function (3) with sample separation information, i.e., a sequential probit model using the administrative information (*Model 4*). We find in Table 2 that the estimated sampling variances of parameters with sample separation were on the average about one-sixth of those without sample separation information. To examine the cost of partial observability further, Figures 2 and 3 plot the estimated probabilities of attaining a favorable outcome at each step and overall for each observation, sorted from low to high for each plotted probability. Compared to the sequential probit models using administrative information (*Models 2* and *4*), we note that *Model 3* (i.e., without sample separation information) predicts higher probabilities of favorable outcome for about 80% of the observations at step 2 and for all observations at step 5, lower probabilities of favorable outcome for about 90% of the observations at step 3 and for all observations at step 4, but overall probabilities of attaining favorable outcome are very close across models. Also, it seems that, with sample separation information, use of the full set of variables as in Table 1 or the restricted set of variables as in Table 2 makes only a slight difference. Table 3 reports

the averages of these probabilities for all models estimated, including *Model 1* in subsection 4.1. Clearly, without using the administrative information, we cannot identify the probability of allowance at each step with satisfactory accuracy. According to *Model 3*, the average probability of an applicant being granted medical allowance at step 3 is only a half of the actual probability, while the average probability of an applicant being granted vocational allowance at step 5 is 40% higher than the actual probability. The similarity of the overall average probability of being awarded benefits across all models is not surprising; it is due to a feature of probit models we are estimating.

A careful look at Figure 3 reveals that the estimated probabilities of favorable outcome at each step with sample separation and using the full set of variables are much smoother across observations, which means that additional variables help to pick up the variations in characteristics in the sample.

We also tried to evaluate the likelihood function (5) with a general or partially restricted W. We worked on the set of variables from Table 2 but excluded those with p-values greater than 0.025, giving a total of 19 variables excluding constant terms. Using the simulated maximum likelihood based on GHK simulator with a sampling size of 500, we obtained an estimate for W as:

$$\begin{bmatrix} 1.000 & & & \\ 0 & 1.000 & & \\ 0 & 0.320\ (0.564) & 1.000 & \\ -0.860\ (0.451) & 0 & 0 & 1.000 \end{bmatrix}$$

with standard errors in the parentheses. Here elements with 0's were constrained and any further relaxation resulted in singular W. We find that only the (4,1) and (1,4) elements, i.e., the covariance between the step 2 and step 5 errors is marginally significant, and has a negative sign. We, however, can not explain the estimate intuitively. The estimates of other structural parameters have been reported in Table 4. Compared to maximum likelihood estimation of (5) using the same set of variables but with $W = I$, the estimated sampling variances of parameters in this estimation with $W \neq I$ were on the average about 5% greater. The $|t|$-values for the included variables at step 2 were significantly lower in this estimation; the rest of the parameters were close to those with $W = I$, see the last two columns of Table 4. Therefore there is slight efficiency loss if W is not assumed to be diagonal in our context. This result is consistent with the finding in Meng and Schmidt (1985) and Keane (1992) that there is a trade-off between efficiency in the parameter estimates and the covariance matrix restrictions, making identification of a model like (5) to be 'fragile' and data dependent, see Poirier (1980). Note that based on the complete set of explanatory variables in Table 1, we found W to be diagonal. Thus, we attribute any non-zero values of the correlations to mis-specification in the list of explanatory variables in Table 4. As a result, we did not use the estimates with $W \neq I$ in our subsequent analysis.

4.3 *Out-of-Sample Predictive Capacity: A Jackknife Approach*

In order to investigate further the value of the structural approach and the sample separation information, we studied the out-of-sample predictive capacity of various model specifications through jackknifing. A model is repeatedly estimated by omitting one

observation at a time. The estimated model is then used to generate a prediction for the probability of attaining an observed outcome for the omitted observation at each step of the process and for the overall probability. This process is replicated for all observations one by one for *Models 2* and *3*. It should be noted that the computation of jackknife predictions for *Model 3* is time consuming and takes more than twenty days on a Pentium II 450MHz PC. We computed Brier's (1950) quadratic probability score (QPS) for attaining a favorable outcome at each step and for the composite allowance probability. The sign test, Wilcoxon's signed-rank test, and the Morgan-Granger-Newbold (MGN) test statistics are also computed for the loss differentials to test the null hypothesis that the competing predictions have no difference in their capacity to predict the binary outcomes. See Hettmansperger (1984) and Diebold and Mariano (1995) for expositions on these statistics. We used the quadratic loss function in computing these test statistics, even though the use of a linear loss function generated very similar conclusions. The reported test statistics have been studentized and are distributed asymptotically standard normal under the null.

When we compared the in-sample predictions with those from the jackknife approach, we found that the QPS scores and the average probabilities of attaining favorable outcomes at each step and also for the composite allowance were remarkably similar across all models. However, the sign test, Wilcoxon's signed-rank test, and MGN test statistics were highly significant and indicated that the jackknife predictions were considerably inferior in predicting the binary outcomes at each step and also the final outcome. This is an expected result, and shows that the jackknife approach gives a better evaluation of the true predictive capacity of an estimated model than those based on

simple in-sample predictions. The important point to note is that the QPS statistics and the average allowance probabilities are not able to discern these differences.

In Table 5 we compare the jackknife predictions from Model 2 (i.e., sequential model with administrative data on sample separation) and those from Model 3 (i.e., sequential model without sample separation information). The QPS statistics for the overall outcome are almost the same for the two models, even though they are substantially better at the individual steps when sample separation information is available. We find a similar result when we look at the sign test, Wilcoxon's signed-rank test, and MGN test. At the individual steps they are generally statistically significant at the 5% level, but for the overall probability the computed statistics are insignificant. These results underscore the usefulness of the marginal likelihood approach adopted by BSBCRS (1999) in cases where the distribution of only the final eligibility probabilities is needed. However, we should emphasize that this encouraging predictive result for *Model 3* may not necessarily carry over to other empirical contexts.

We also generated jackknife predictions based on an estimated reduced-form model, which only used the binary final outcomes and ignored the sequential structure of the disability determination process. Table 6 reports the parameter estimates from a probit regression of a reduced-form model with explanatory variables that were statistically significant at the 10% level. A comparison of the jackknife predictions from the reduced-form model with those from *Model* 2 for the final overall outcome revealed that that reduced-form model is inferior to *Model 2*. The sign test, Wilcoxon's signed-rank test, and MGN test statistics are calculated to be –4.73, –3.22, and -2.41 respectively, which are highly significant. When we compared the reduced form predictions with those from

Model 3, these three test statistics are found to be -5.36, -3.60, and -1.74, which are statistically significant at the 10% level. Thus, we find little evidence to suggest that in the absence of sample separation information the conventional reduced form approach over the sequential one should be adopted.

5. CONCLUSIONS

We have estimated a 4-step sequential probit model to characterize SSA's disability determination process with and without the sample separation information on the outcomes in the intermediate steps. Under the program provisions, different criteria dictate the outcomes at different steps of the determination process. We used data on health, activity limitations, demographic traits, and work from 1990 SIPP exact matched to SSA administrative records on disability determinations. Using GHK Monte Carlo simulation technique, our estimation results suggest that the correlations in errors across equations that may arise due to unobserved individual heterogeneity is not statistically significant. In addition, we examined the value of the administrative data on the basis for allow/deny determinations at each stage of the process. Following the approach taken by BSBCRS (1999), we also estimated the above sequential probit model without the sample separation information for the purpose of a direct comparison. We found that without this detailed administrative information on outcomes at each stage of the screen, we could not properly evaluate the importance of a large number of program-relevant survey-based explanatory variables. A considerable loss in estimation efficiency was also observed

when the sample separation information was not used. In terms of both in-sample and jackknife-type out-of-sample predictive analysis, the value of structural modeling over the conventional allow/deny reduced form regression is clearly established.

ACKNOWLEDGEMENTS

We are grateful to our discussant Maarten Lindeboom at the Rotterdam Conference, and to Benjamin Bridges, Jianting Hu, John Rust, and Denny Vaughan for many helpful comments and discussions. The views expressed in this paper are our own, and do not necessarily represent those of the Social Security Administration.

REFERENCES

Aarts, L.J.M., and P.R. De Jong (1992), *Economic Aspects of Disability Behavior*. Amsterdam: North Holland.

Benitez-Silva, H., M. Buchinsky, H.-M. Chan, J. Rust and S. Sheidvasser (1999), 'An empirical analysis of the social security disability application, appeal, and award process', *Labour Economics*, 6, 147-178.

Blundell, R.W., J. Ham and C. Meghir (1987), 'Unemployment and female labor supply', *Conference Papers*, Supplement to *Economic Journal*, 97, 44-64.

Brier, G.W. (1950), 'Verification of forecasts expressed in terms of probability', *Monthly Weather Review*, 75, 1-3.

Cragg, J.G. (1971), 'Some statistical models for limited dependent variables with applications to the demand for durable goods', *Econometrica*, 39, 829-44.

Diebold, F.X. and R.S. Mariano (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics*, 13, 253-263.

Fisher, J.A. (1962) 'An analysis of consumer durable goods expenditures in 1957', *Review of Economics and Statistics*, 27, 431-47.

Geweke, J. (1989), 'Efficient simulation from the multivariate normal distribution subject to linear inequality constraints and the evaluation of constraint probabilities', Mimeo. (Duke University, Durham, NC)

Goldfeld, S.M., and R.E. Quandt (1975), 'Estimation in a disequilibrium model and the value of information', *Journal of Econometrics*, 3, 325-348.

Hajivassiliou, V. (1993), 'Simulation estimation methods for limited dependent variable models'. in: G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., *Handbook of Statistics*, Vol. 11 (Econometrics). (North-Holland, Amsterdam)

Hajivassiliou, V., D. McFadden, and P. Ruud (1996), 'Simulation of multivariate normal rectangle probabilities and their derivative: Theoretical and computational results', *Journal of Econometrics*, 72, 85-134.

Hettmansperger, T.P. (1984), *Statistical Inference Based on Ranks*, Wiley, New York.

Jones, A.M. (1989), 'A double-hurdle model of cigarette consumption', *Journal of Applied Econometrics*, 4, 23-39.

Keane, M. (1992), 'A note on identification in the multinomial probit model', *Journal of Business and Economic Statistics*, 10, 193-200.

Keane, M. (1994), ‘A computationally efficient practical simulation estimator for panel data, with applications to estimating temporal dependence in employment and wages’, *Econometrica*, 62, 95-116.

Kiefer, N.M. (1978), Discrete parameter variation: Efficient estimation of a switching regression model’, *Econometrica*, 46, 427-434.

Lahiri, K., D.R. Vaughan and B. Wixon (1995), ‘Modeling SSA’s sequential disability determination process using matched SIPP data’, *Social Security Bulletin* 58, No. 4, 3-42.

Meng, C.L. and P. Schmidt (1985), ‘On the cost of partial observability in the bivariate probit model’, *International Economic Review*, 26, 71-85.

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

Poirier, D. (1980) ‘Partial observability in bivariate probit models’, *Journal of Econometrics*, 12, 209-217.

Schmidt, P. (1981), ‘Further results on the value of sample separation information’, *Econometrica*, 49, 1339-1343.

Table 1. Sequential Probit Model with Sample Separation Information

Variable Description	Mnemonic	With general *W* Estimate (*t*-ratio)	When *W* = *I* Estimate (*t*-ratio)
Step 2: Severe impairment			
Constant		0.681 (5.294)	0.710 (5.535)
With recent work experience and disability determination in 1990	WORK90C	-0.412 (-2.479)	-0.376 (-2.276)
Three or more severe ADLs, wave 6	SAL36	0.685 (1.935)	0.776 (2.260)
One or more severe IADLs, wave 3	SIL13	0.400 (2.627)	0.388 (2.532)
Prevented from working (wave 2) and never able to work at a job	T8338W2D	0.941 (1.871)	0.872 (1.695)
Gender (male)	SEXD	0.251 (2.566)	0.255 (2.588)
General health status good (wave 6)	T8800W6B	-0.229 (-1.797)	-0.282 (-2.146)
General health status poor (wave 6)	T8800W6E	0.190 (1.559)	0.199 (1.622)
White-south (Black/other and north in the base)	RACESTDA	-0.281 (-1.905)	-0.320 (-2.159)
White-north (Black/other and south in the base)	RACESTDB	0.298 (2.195)	0.283 (2.089)
Black-south (White/other and north in the base)	RACESTDC	-0.536 (-3.322)	-0.567 (-3.506)
Work limited because of mental condition	MEDFRP31	0.334 (2.163)	0.337 (2.153)
Reports inability to work in at least 2 waves	TDIREP12	0.255 (2.333)	0.268 (2.434)
Work limiting condition caused by accident	T8326W2D	-0.178 (-1.417)	-0.230 (-1.799)
Age 18-34 (35 plus in the base)	AGE12	-0.203 (-1.826)	-0.203 (-1.807)
Work limited less than 12 months	WPRVDUD1	-0.307 (-1.984)	-0.297 (-1.915)
Needs help in getting around the house	T8840W3D	-0.669 (-2.093)	-0.709 (-2.172)
Work limited, but to do prior work (both in wave 2)	WORKV2D2	0.669 (3.321)	0.620 (2.996)
Work limited, but able to work occasionally or irregularly	WORKV1D3	-0.509 (-2.019)	-0.499 (-1.984)
Step 3: Listing impairment			
Constant		-0.483 (-4.886)	-0.470 (-5.808)
At least 1 overnight hospital stay in last 12 months	T9100W3D	0.190 (1.809)	0.175 (1.656)
Reports at least two mental conditions (wave 3)	TDI12W3D	0.803 (2.473)	0.817 (2.459)
Has two or more severe ADLs (wave 6)	TAS12W6D	0.389 (1.886)	0.367 (1.814)
Has at least two IADLs (wave 3)	TIL12W3D	0.285 (1.885)	0.288 (1.935)
With recent work experience and disability determination occurred in 1991	WORK91C	0.275 (2.017)	0.259 (1.891)
With recent work experience and disability determination occurred in 1992	WORK92C	0.263 (2.224)	0.274 (2.306)
Never married	MSF	0.264 (2.360)	0.247 (2.213)
Work limiting condition caused by accident	T8326W2D	-0.422 (-3.094)	-0.369 (-2.682)
Aged 55 or older (18-54 in the base)	AGE56	-0.157 (-1.589)	-0.165 (-1.672)
Work limited because of musculoskeletal condition	MEDGRP32	-0.207 (-1.699)	-0.246 (-2.009)
Work limited because of sensory/neurological condition	MEDGRP33	0.450 (2.662)	0.449 (2.603)
Unable to walk 3 city blocks	T8832W3D	-0.234 (-1.828)	-0.265 (-2.074)
Needs help in doing light house work	T8859W6D	0.371 (2.357)	0.360 (2.277)
Has difficulty lifting 10 lbs., and reports presence of work limitation (both in wave 6)	LFTCNW6D	-0.315 (-2.560)	-0.329 (-2.637)
Has difficulty walking up stairs and reports presence of work limitation (both in wave 6)	WUPCNW6D	0.245 (2.034)	0.279 (2.275)
Needs help in getting out of bed or chair	T8848W3D	-0.510 (0.306)	-0.589 (-1.959)
Needs help in getting around inside the home	T8840W3D	0.501 (0.335)	0.566 (1.697)

Table 1. (Cont.) Sequential Probit Model with Sample Separation Information

Variable Description	Mnemonic	With general *W* Estimate (*t*-ratio)	When *W* = *I* Estimate (*t*-ratio)
Step 4: Capacity for past work			
Constant		0.655 (2.869)	0.509 (3.874)
Any one of the five mental disability reported or a mental condition reported as causing work or activity limitation	MENTDISD	0.349 (2.432)	0.339 (2.403)
Unable to lift 10 lbs., and prior work was very physically demanding using strength, stoop, climb criteria	NSTRLIFT	0.402 (1.695)	0.440 (1.847)
No recent work experience	NOWORKD	0.472 (2.912)	0.461(2.928)
Work limited, but able to perform prior work (both in wave 2)	WORKV2D2	-0.483 (-2.628)	-0.492 (-2.701)
Principal occupation of prior work was in sales or service	OCCSIPP3	-0.442 (-2.930)	-0.456 (-3.067)
Never married	MSF	0.443 (2.448)	0.432 (2.473)
Prior work physically demanding according to broad strength, stoop, climb criteria	SIPPOCC4	0.416 (3.244)	0.413 (3.209)
Has work limitation and has difficulty lifting and carrying 10 lbs.	LFTCNW6D	-0.313 (-2.801)	-0.300 (-2.594)
White-north (Black/other and south in the base)	RACESTDB	-0.282 (-2.110)	-0.227 (-2.001)
Step 5: Capacity for other work			
Constant		-1.134 (-4.719)	-0.739 (-3.917)
Aged 55 or older (18-54 in the base)	AGE56	1.330 (5.993)	1.454 (7.830)
Gender (male)	SEXD	-0.244 (-1.576)	-0.350 (-2.413)
Disability determination occurred in 1988	SSAY88D	0.463 (1.534)	0.561 (1.729)
Disability determination occurred in 1991	SSAY91D	0.316 (1.560)	0.388 (1.937)
Disability determination occurred in 1992	SSAY92D	0.371 (1.872)	0.453 (2.131)
Mental condition is cause of work or activity limitation	MENTDISD	0.569 (3.165)	0.572 (3.172)
Prior work physically demanding according to broad strength, stoop, climb criteria	SIPPOCC4	0.332 (2.319)	0.344 (2.262)
Able to work only occasionally or irregularly (wave 2)	WORKV1D3	-1.095 (-1.832)	-1.127 (-1.709)
Unable to work (wave 2)	WORKV1D4	0.305 (2.177)	0.261 (1.721)
Under age 35 and has a mental condition	YONDMENT	0.831 (2.874)	0.868 (2.844)
Has one or more severe functional or ADL imitations, has more than 12 years of education, under age 55	FLADLEDY	-0.295 (-2.150)	-0.330 (-2.197)
No severe functional limitation, no ADL, under age 55	NFLADLYD	-0.333 (-1.976)	-0.351 (-1.944)
Average log-likelihood		-1.383	-1.384

Table 2. Sequential Probit Model without/with Sample Separation Information, $W = I$

Variable Description	Mnemonic	Without Sample Separation Estimate (*t*-ratio)	With Sample Separation Estimate (*t*-ratio)
Step 2: Severe impairment			
Constant		1.838 (3.266)	1.052 (13.83)
With recent work experience and disability determination in 1990	WORK90C	-1.074 (-2.938)	-0.413 (-2.762)
General health status good (wave 6)	T8800W6B	-0.681 (-1.934)	-0.309 (-2.466)
White-south (Black/other and north in the base)	RACESTDA	-1.350 (-2.668)	-0.414 (-3.956)
Black-south (White/other and north in the base)	RACESTDC	-1.828 (-3.391)	-0.728 (-5.975)
Reports inability to work in at least 2 waves	TDIREP12	1.021 (3.218)	0.371 (4.120)
Work limiting condition caused by accident	T8326W2D	-0.545 (-1.416)	-0.190 (-1.666)
Step 3: Listing impairment			
Constant		-1.807 (-4.806)	-0.616 (-8.824)
At least 1 overnight hospital stay in last 12 months	T9100W3D	0.511 (2.209)	0.191 (1.906)
Reports at least two mental conditions (wave 3)	TDI12W3D	1.394 (1.722)	0.875 (2.689)
Has two or more severe ADLs (wave 6)	TAS12W6D	0.857 (2.161)	0.308 (1.650)
With recent work experience and disability determination occurred in 1991	WORK91C	0.677 (2.256)	0.195 (1.453)
With recent work experience and disability determination occurred in 1992	WORK92C	-0.899 (-1.155)	0.259 (2.239)
Never married	MSF	1.016 (3.405)	0.357 (3.336)
Work limiting condition caused by accident	T8326W2D	-1.878 (-2.328)	-0.513 (-4.301)
Work limited because of sensory/neurological condition	MEDGRP33	1.045 (3.035)	0.533 (3.306)
Needs help in doing light house work	T8859W6D	0.702 (2.323)	0.408 (2.750)
Has difficulty walking up stairs and reports presence of work limitation (both in wave 6)	WUPCNW6D	0.704 (2.532)	0.046 (0.479)
Step 4: Capacity for past work			
Constant		0.795 (2.464)	0.528 (4.262)
No recent work experience	NOWORKD	-0.215 (-1.048)	0.549 (3.617)
Work limited, but able to perform prior work (both in wave 2)	WORKV2D2	-0.509 (-1.663)	-0.397 (-2.239)
Prior work physically demanding according to broad strength, stoop, climb criteria	SIPPOCC4	0.244 (1.360)	0.373 (3.137)
Has work limitation and has difficulty lifting and carrying 10 lbs.	LFTCNW6D	-0.454 (-2.365)	-0.279 (-2.554)
White-north (Black/other and south in the base)	RACESTDB	-0.408 (-2.059)	-0.171 (-1.558)
Step 5: Capacity for other work			
Constant		0.193 (0.749)	-0.564 (-3.684)
Aged 55 or older (18-54 in the base)	AGE56	0.829 (2.637)	1.424 (8.053)
Disability determination occurred in 1992	SSAY92D	0.616 (2.074)	0.256 (1.297)
Mental condition is cause of work or activity limitation	MENTDISD	1.256 (2.919)	0.862 (5.745)
Able to work only occasionally or irregularly (wave 2)	WORKV1D3	-1.370 (-2.638)	-1.135 (-2.152)
Has one or more severe functional or ADL limitations, has more than 12 years of education, under age 55	FLADLEDY	-0.542 (-2.499)	-0.300 (-2.170)
No severe functional limitation, no ADL, under age 55	NFLADLYD	-0.269 (-1.317)	-0.225 (-1.382)

Average log-likelihood	-0.607	-1.445

Table 3. In-Sample Prediction Analysis: Average Probabilities for Favorable Outcomes

Model	Step 2	Step 3	Step 4	Step 5	Overall
Model 1. With sample separation information, general *W*, and full set of variables [a]	0.819	0.352	0.715	0.475	0.471
Model 2. With sample separation information, *W* = *I*, and full set of variables [a]	0.819	0.351	0.719	0.464	0.472
Model 3. Without sample separation information, *W* = *I*, and restricted set of variables [b]	0.870	0.177	0.661	0.648	0.464
Model 4. With sample separation information, *W* = *I*, and restricted set of variables [b]	0.818	0.351	0.715	0.456	0.463
Actual value	0.819	0.357	0.710	0.463	0.465

Notes: [a] The set of variables in Table 1 is used.
[b] The set of variables in Table 2 is used.

Table 4. Sequential Probit Model without Sample Separation Information: $W \neq I$ *vs.* $W = I$

Variable Description	Mnemonic	$W \neq I$ Estimate (*t*-ratio)	$W = I$ Estimate (*t*-ratio)
Step 2: Severe impairment			
Constant		1.589 (3.336)	1.457 (3.204)
With recent work experience and disability determination in 1990	WORK90C	-1.106 (-2.999)	-1.222 (-3.217)
White-south (Black/other and north in the base)	RACESTDA	-1.023 (-2.306)	-1.046 (-2.362)
Black-south (White/other and north in the base)	RACESTDC	-1.418 (-2.861)	-1.628 (-3.587)
Reports inability to work in at least 2 waves	TDIREP12	0.929 (2.664)	1.307 (3.354)
Step 3: Listing impairment			
Constant		-1.469 (-4.124)	-1.537 (-4.418)
At least 1 overnight hospital stay in last 12 months	T9100W3D	0.560 (2.522)	0.507 (2.326)
Has two or more severe ADLs (wave 6)	TAS12W6D	0.805 (2.108)	0.783 (2.114)
With recent work experience and disability determination occurred in 1991	WORK91C	0.651 (2.333)	0.673 (2.415)
Never married	MSF	0.749 (2.758)	0.789 (2.925)
Work limiting condition caused by accident	T8326W2D	-2.022 (-2.127)	-1.873 (-2.196)
Work limited because of sensory/neurological condition	MEDGRP33	0.861 (2.543)	0.873 (2.634)
Needs help in doing light house work	T8859W6D	0.644 (2.194)	0.610 (2.114)
Has difficulty walking up stairs and reports presence of work limitation (both in wave 6)	WUPCNW6D	0.534 (2.053)	0.501 (1.942)
Step 4: Capacity for past work			
Constant		0.788 (2.166)	0.720 (2.057)
Has work limitation and has difficulty lifting and carrying 10 lbs.	LFTCNW6D	-0.430 (-2.153)	-0.442 (-2.200)
White-north (Black/other and south in the base)	RACESTDB	-0.388 (-1.821)	-0.318 (-1.604)
Step 5: Capacity for other work			
Constant		0.105 (0.413)	-0.004 (-0.015)
Aged 55 or older (18-54 in the base)	AGE56	0.902 (2.858)	0.904 (2.845)
Disability determination occurred in 1992	SSAY92D	0.451 (1.743)	0.503 (1.767)
Mental condition is cause of work or activity limitation	MENTDISD	1.170 (2.652)	1.242 (2.629)
Able to work only occasionally or irregularly (wave 2)	WORKV1D3	-1.264 (-2.867)	-1.359 (-2.812)
Has one or more severe functional or ADL limitations, has more than 12 years of education, under age 55	FLADLEDY	-0.455 (-2.153)	-0.459 (-2.051)
Average log-likelihood		-0.614	-0.615

Table 5. Jackknife Prediction Analysis: Comparison between *Models 2* and *3*

	Step 2	Step 3	Step 4	Step 5	Overall
Number of observations	1230	1007	648	460	1230
QPS-*Model 2*	0.364	0.282	0.369	0.175	0.221
QPS-*Model 3*	0.413	0.403	0.332	0.234	0.222
Sign test	0.228	-0.725	2.278*	-4.663*	1.483
Wilcoxon's signed-rank test	-8.318*	-10.576*	3.336*	-5.790*	0.580
MGN test	-10.941*	-16.053*	5.383*	-5.896*	-0.565

Note: QPS is defined as mean squared prediction errors. It ranges from 0 to 1, with a score of 0 corresponding to perfect accuracy. The sign test, Wilcoxon's signed-rank test, and MGN test statistics are studentized. * denotes significant at a 5% level.

Table 6. Probit Reduced-form Regression

Variable Description	Mnemonic	Parameter Estimate (*t*-ratio)
Constant		-0.225 (-3.432)
With recent work experience and disability determination in 1990	WORK90C	-0.375 (-2.519)
White-south (Black/other and north in the base)	RACESTDA	-0.141 (-1.492)
Black-south (White/other and north in the base)	RACESTDC	-0.363 (-3.030)
Work limiting condition caused by accident	T8326W2D	-0.374 (-3.736)
At least 1 overnight hospital stay in last 12 months	T9100W3D	0.322 (3.547)
Never married	MSF	0.156 (1.502)
Age 55 or older (18-54 in the base)	AGE56	0.436 (5.276)
Work limited because of sensory/neurological condition	MEDGRP33	0.488 (3.270)
Able to work only occasionally or irregularly (wave 2)	WORKV1D3	-0.829 (-2.992)
Under age 35 and has a mental condition	YONDMENT	0.657 (4.638)
Average log-likelihood		-0.640

Figure 1. SSI disability determination process