

# A STRUCTURAL MODEL OF SOCIAL SECURITY'S DISABILITY DETERMINATION PROCESS

Jianting Hu, Kajal Lahiri, Denton R. Vaughan, and Bernard Wixon\*

*Abstract*—We estimate a multistage sequential logit model reflecting the structure of the disability determination process of the Social Security Administration (SSA). The model is estimated using household survey information exact-matched to SSA records on disability adjudications from 1989 to 1993. Under program provisions, different criteria dictate outcomes at different steps of the determination process. We find that, without the multistaged structural approach, effects of many important health, disability, and vocational factors are not readily discernible. As a result, split-sample predictions of overall allowance rates from the sequential model perform considerably better than do those for the conventional allowed/denied logit regression.

## I. Introduction

THE disability determination process used by the Social Security Administration (SSA) acts as the gatekeeper for several public programs: Disability Insurance (DI), Supplemental Security Income (SSI) and—for those judged disabled—Medicare and, in most states, Medicaid. Between 1985 and 1997, annual benefits under the two cash programs, DI and SSI, grew from \$26 billion to \$68 billion—increasing more than 150% in current dollars. Factoring in Medicare and Medicaid, total benefits to the disabled have grown to well over \$100 billion per year. Entitlement spending on this scale argues not only for more analysis of applications decisions, but for an expanded focus on the complex procedure for determining eligibility. The eligibility criteria implicit in the screening process reflect basic normative decisions that define the population the program is intended to serve. Hence, it is important to understand the structure of the screening process and how it affects the size and characteristics of the targeted population.

Unfortunately, analysts studying disability program growth face a daunting information gap. Administrative data, tied to day-to-day operations, have no information on those who have not applied and little socioeconomic information that is useful for understanding application behavior. In contrast, household surveys of the general population provide information on nonapplicants and on a range of socioeconomic details; however, they contain little information on an individual's actual interactions with the administering agency—information that is needed to relate survey information to the agency's eligibility decisions.

SSA addressed this issue by implementing three major surveys: the 1966 Social Security Survey of the Disabled,

the 1972/1974 Survey of Disabled and Non-Disabled Adults, and the 1978 Survey of Disability and Work. These surveys asked questions on health impairments and socioeconomic status that were designed for analysis of the disability programs. The resulting data have been utilized in a number of important studies of disability enrollment.<sup>1</sup> However, little effort has been devoted to modeling the structure of the disability determination process.

In this paper, we model the outcome of SSA disability determinations using survey responses on health, functional limitations, demographic traits, and work behavior for adult applicants. To do this, we exact-match data from the 1990 Survey of Income and Program Participation (SIPP) to SSA records on disability determination. We adopt a structural approach, using a multistage logit model that mimics the sequential process by explicitly modeling decisions made under distinct criteria. For example, SSA's decisions at some steps are based solely on medical evidence and health-related activity limitations, whereas decisions at other steps also take into account age, education, and past work. The matched data, combined with the structural representation of the process, enable more-appropriate use of survey data in explaining these differential eligibility outcomes.

The paper is organized as follows. Section II outlines the sequential disability determination process, by step. Section III introduces the four-stage sequential logit model. Section IV describes the data used in the study. Section V presents empirical findings, and section VI reports split-sample analysis results. Finally, section VII summarizes our conclusions.

## II. Disability Determination

The two disability programs administered by SSA use the same disability determination process, but have different, albeit complementary, objectives. The DI program is part of the comprehensive social insurance program—Old Age, Survivors, and Disability Insurance (OASDI). It is funded through payroll tax contributions. The SSI program, by contrast, is means-tested and guarantees a minimal level of income to the poorest of the aged, blind, or disabled. SSI is funded from general revenues. Compared to eligibility decisions for other programs, disability determination is complex and not widely understood. Yet, in recent years, total applications for the two programs have reached as high as 2.5 million per year, so the budgetary and welfare impact of the process are undeniable. The process modeled in this

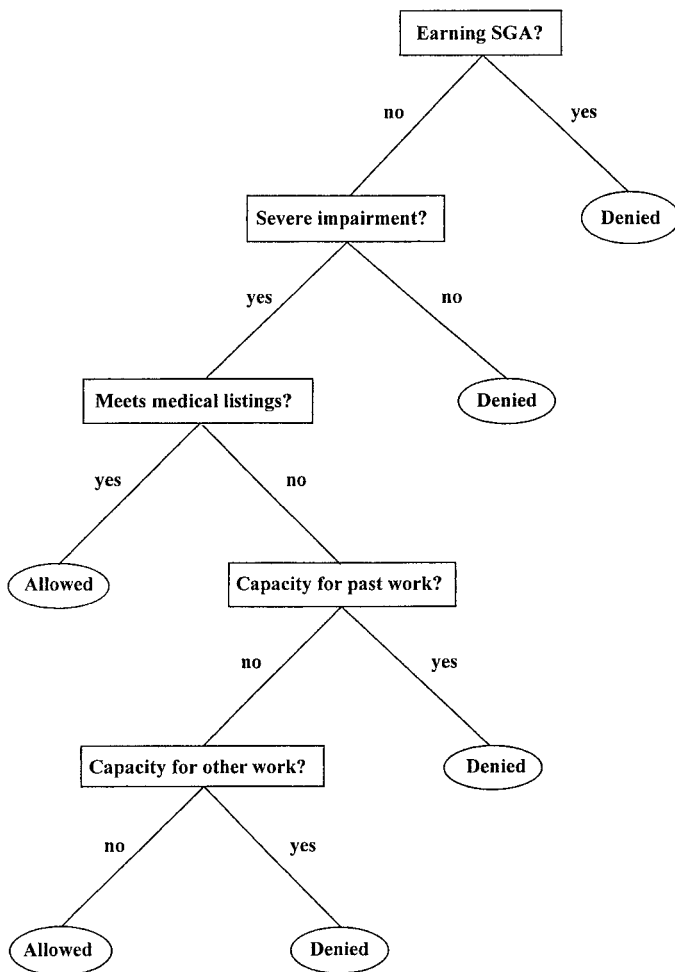
<sup>1</sup> See, for example, Halpern and Hausman (1986) and Leonard (1986) who used the Survey of Disabled and Non-Disabled Adults, and Haveman, De Jong, and Wolfe (1991), Stern (1989), and Kreider (1998) who used the 1978 Survey of Disability and Work.

Received for publication October 1, 1998. Revision accepted for publication May 4, 2000.

\* Social Security Administration, State University of New York at Albany and Social Security Administration, Social Security Administration, and Social Security Administration, respectively.

The authors wish to thank Howard Bradley, Benjamin Bridges, Richard Burkhauser, Susan Grad, Brent Kreider, Michael Leonasio, Robin Sickles, Peter Wheeler, and two anonymous referees for their helpful comments. This paper is an abridged version of our own working paper (Hu et al., 1997).

FIGURE 1.—SSA DISABILITY DETERMINATION PROCESS



study is that presently used by the state Disability Determination Service (DDS) agencies. It is used in initial determinations or DDS reconsiderations and it accounted for more than 80% of all allowances in fiscal year 1990. Other decisions made under the multilayered appeals process—by administrative law judges (ALJs), for example—are not modeled in this analysis.<sup>2</sup> As shown in figure 1, disability determination involves a sequence of five steps.

A. Step 1: An Earnings Screen

SSA field offices screen out applicants who work. Applicants are denied if they engage in activities that are both substantial and gainful. Specifically, applicants earning more than the substantial gainful activity (SGA) amount are denied immediately; otherwise, the application is referred to the DDS.<sup>3</sup> The maximum SGA amount was \$500 per month

<sup>2</sup> There are five levels of decision making: initial, reconsideration, administrative law judge, Appeals Council, and Federal District Court.

<sup>3</sup> The field office also verifies that the applicant has contributed for the requisite number of quarters or, in the case of SSI, does a preliminary check of eligibility based on income and assets. An applicant must be financially eligible for one or both programs to be referred to the DDS.

in 1993. Because this step is not based on medical criteria, it is not modeled here.

B. Step 2: A Medical Screen to Deny the Least Severely Impaired

An applicant is denied at step 2 if his or her impairment (or combination of impairments) is judged not severe. This judgment is based on a conceptual threshold: an impairment is considered nonsevere if it does not significantly limit the physical or mental abilities to perform basic work-related activities. Basic work activities include physical functions (such as walking, standing, or lifting), sensory functions (such as seeing, hearing, or speaking), and routine mental functions (such as understanding simple instructions, responding to supervision, or adapting to changes in the work environment). Applicants are also denied if their impairments fail the duration test, that is, if the impairment is not expected to result in death, and if the impairment has neither lasted twelve months nor is expected to last twelve months.

C. Step 3: A Medical Screen to Allow the Most Severely Impaired

Medical evidence on an applicant's impairment is assessed under step 3 using codified clinical criteria relating to both the nature and severity of the impairment. These codified criteria, including more than 100 impairments, are called the *Listings of Impairments*. Applicants with impairments that "meet" the listings are allowed immediately, solely on the basis of medical criteria. Moreover, if an applicant has an impairment that is not included in the listings but is considered medically equivalent to a listed impairment, the impairment is said to "equal the listings" and the applicant is allowed. Applicants neither denied at step 2 nor allowed at step 3 have impairments that, although severe, are not severe enough to consider them disabled purely on medical grounds. Such applicants are evaluated further in steps 4 and 5. During 1990–1993, more than two-thirds of all initial allowances for adult applicants were based on the listings.

D. Step 4: Can Severely Impaired Applicants Work in Past Jobs?

In this step, an applicant's residual functional capacity (RFC) to meet the requirements of major past jobs is considered. The evaluation of residual functional capacity determines to what extent the applicant can perform basic work-related activities despite a severe impairment. This analysis takes into account, for example, whether the applicant can walk, lift objects, follow instructions, and tolerate environmental conditions found in the workplace. The work-related functions considered are the same ones considered in step 2, but they are evaluated differently. In step 2, the presence of such limitations is taken as evidence of a severe limitation, but in step 4 the extent of the limitations

is considered, by comparing the applicant's residual capacity with the demands of past jobs. Hence, step 4 also involves determining the requirements of jobs held in the fifteen years prior to application, if they were held long enough for the applicant to learn the requisite skills. For example, if two applicants have similar leg impairments, the applicant who held a desk job is more likely to be judged able to do sedentary work and hence denied, whereas the applicant who held a physically demanding job would probably be evaluated further in the final step. In general, applicants judged able to perform past work are denied; the remaining applicants, including those with no recent work experience, are passed on to step 5.

#### E. Step 5: Can Severely Impaired Applicants Do Any Work in the Economy?

Under step 5, the applicant's residual functional capacity is considered, along with vocational factors, to determine whether the applicant can work in jobs other than those previously held. The analysis of residual capacity from step 4 indicates whether the applicant can do sedentary, light, medium, heavy, or very heavy work.<sup>4</sup> Under step 5, the applicant's age, education, and work experience are used to determine whether the applicant can work in employment consistent with his or her residual capacity. This determination is based on a table called the vocational grid. The grid, like the medical listings, lends objectivity to the determination process, facilitating uniform administration. To illustrate the use of the grid, suppose an applicant's impairment permits only sedentary work. If the applicant is of advanced age, low education (six years or less), and has held only arduous, unskilled jobs, then the applicant is judged disabled using the vocational grid.

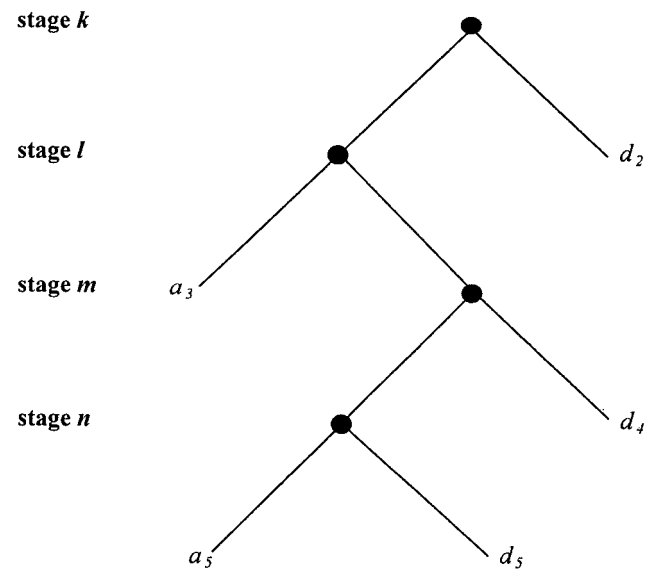
### III. The Structural Model

The intent of the sequential process is to provide a cost-effective procedure for determining disability that can be administered uniformly throughout the nation. Intuitively, decisions made at early steps are more straightforward to adjudicate, because they require only development of medical information, whereas medical-vocational decisions—made at steps 4 and 5—require both medical and vocational information. Hence, the process is made more efficient not only by breaking the determination into separate steps, each with distinct decision criteria, but by ordering the steps in terms of their informational requirements.<sup>5</sup>

<sup>4</sup> A separate but analogous procedure is used for nonexertional impairments such as mental impairments or environmental impairments (which relate to workplace conditions).

<sup>5</sup> The sequential process can be justified rigorously in terms of McFadden's (1976) optimizing framework which Mashaw (1983) characterizes as "bureaucratic rationality." According to Mashaw, this objective has been of crucial importance in the historical development of the screening process.

FIGURE 2.—THE SEQUENTIAL MODEL



We model the last four steps of the determination process, which include four decision nodes and five final outcomes,  $d_2$ ,  $a_3$ ,  $d_4$ ,  $a_5$ , and  $d_5$ . The decision nodes correspond to the four steps modeled and, for convenience, will be denoted  $k$ ,  $l$ ,  $m$ , and  $n$ , respectively. As shown in figure 2,  $d_2$  = denial at the second step based on severity of impairments;  $a_3$  = allowance at the third step based on listed impairments;  $d_4$  = denial at the fourth step based on residual capacity for past work;  $a_5$  and  $d_5$  are allowance and denial, respectively, at the fifth step based on capacity for any work in the economy.

Because the decisions are sequential, the probability of an outcome at a decision node can be taken as independent of choices at prior nodes. This gives rise to the sequential logit model that separates all choices into a series of conditional binary choices (Amemiya, 1985). At each node— $k$ ,  $l$ ,  $m$ , or  $n$ —the value "1" indicates the favorable outcome from the standpoint of the applicant, that is, either an allowance (at nodes  $l$  and  $n$ ) or a pass on to the next step (at nodes  $k$  and  $m$ ). Suppose  $W_k$ ,  $X_l$ ,  $Y_m$ , and  $Z_n$  are sets of explanatory variables and  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  are corresponding parameter vectors used to evaluate "severity" at each step. Based on a logit regression, the probability of denial at step 2 can be written as

$$\Pr(d_2) = P_{k=0} = 1 - F(\alpha'W_k), \quad (1)$$

where  $P_{k=1} = F(\alpha'W_k) = \exp(\alpha'W_k)/[1 + \exp(\alpha'W_k)]$ . Similarly,

$$\begin{aligned} \Pr(a_3) &= P_{l=1} \\ &= P_{l=1|k=1} \cdot P_{k=1} = F(\beta'X_l)F(\alpha'W_k), \end{aligned} \quad (2)$$

where  $P_{l=1|k=1} = F(\beta'X_l) = \exp(\beta'X_l)/[1 + \exp(\beta'X_l)]$  is the probability of allowance at step 3, conditional on not being denied at step 2. Likewise,

$$\begin{aligned}
 \Pr(a_4) &= P_{m=0} = P_{m=0|l=0,k=1} \cdot P_{l=0|k=1} \cdot P_{k=1} \\
 &= [1 - F(\gamma' Y_m)] \cdot [1 - F(\beta' X_l)] \\
 &\quad \times F(\alpha' W_k),
 \end{aligned} \tag{3}$$

where  $P_{m=0|l=0,k=1} = 1 - F(\gamma' Y_m) = 1 - \exp(\gamma' Y_m) / [1 + \exp(\gamma' Y_m)]$  is the probability of denial at step 4, conditional on not being denied at step 2 and not being allowed at step 3. And, finally,

$$\begin{aligned}
 \Pr(a_5) &= P_{n=1} \\
 &= P_{n=1|m=1,l=0,k=1} \cdot P_{m=1|l=0,k=1} \\
 &\quad \times P_{l=0|k=1} \cdot P_{k=1} \\
 &= F(\delta' Z_n) \cdot F(\gamma' Y_m) \\
 &\quad \times [1 - F(\beta' X_l)] \cdot F(\alpha' W_k),
 \end{aligned} \tag{4}$$

where  $P_{n=1|m=1,l=0,k=1} = F(\delta' Z_n) = \exp(\delta' Z_n) / [1 + \exp(\delta' Z_n)]$  is the probability of allowance in the last step, conditional on being not denied at step 2, not allowed at step 3, and not denied at step 4. The corresponding probability of denial at step 5 can be expressed as

$$\begin{aligned}
 \Pr(d_5) &= P_{n=0} \\
 &= P_{n=0|m=1,l=0,k=1} \cdot P_{m=1|l=1,k=1} \\
 &\quad \times P_{l=0|k=1} \cdot P_{k=1} \\
 &= [1 - F(\delta' Z_n)] \cdot F(\gamma' Y_m) \\
 &\quad \times [1 - F(\beta' X_l)] \cdot F(\alpha' W_k)
 \end{aligned} \tag{5}$$

We estimate parameter vectors  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  sequentially over surviving subsamples. The successive sample splits representing the outcome at each step of the screening process are based on the SSA disability determination records. Note that the probability that an applicant with given characteristics is eligible for disability benefits is  $\Pr(a_3) + \Pr(a_5)$ .

Although the determination process yields five allow/deny choices, they are not ordered in terms of a single underlying index of disability severity. Only  $d_2$  and  $a_3$ —representing the least severely impaired and the most severely impaired—can unequivocally be ordered in terms of a single latent-severity distribution. Hence, to estimate the eligibility probability for an applicant, one needs to generate the outcome probabilities at each step of a discontinuous, nonlinear process. This point has not been appreciated much in recent research on the economic effects of disability. Disability determination has typically been modeled simply by using one underlying index of severity, expressed as a function of health and other factors.

#### IV. The Matched Data

The study relies on two principal data sources: the 1990 panel of the SIPP and SSA's disability determination records

(the so-called 831 data). These data sets are exact-matched for SIPP sample members who applied for disability and whose applications were adjudicated during 1989–1993. In addition, we add information on occupational characteristics from the *Dictionary of Occupational Titles* (DOT) (DOL, 1993), workloads from DDS Staffing and Workload Analysis Reports, and staffing, workload, processing time, and demographics at the district-office level from the SSA's Profiling System Database.

The 1990 SIPP panel includes information on work history, work disability, general health status, functional and activity limitations, health conditions, healthcare utilization, and related topics. Two types of health measures have proven central to our analysis: activities of daily living (ADLs) (such as dressing, eating, and using the toilet) and instrumental activities of daily living (IADLs) (such as preparing meals and getting around inside the home). The SIPP information was collected in four topical modules over a twenty-month period from June–September 1990 to February–May 1992. Programmatic information on disability applicants is from the SSA Form 831, used by DDS agencies to document disability determinations for applicants. These data include administrative, diagnostic, and statistical items, although there is considerable nonreporting of statistical items. The most critical item from the 831 data, the regulation basis code, characterizes the determination outcome in terms of the steps of the sequential process.

We take information on physical demands, general educational development (GED), specific vocational preparation (SVP), and environmental conditions of occupations from the DOT. The Master Crosswalk Database program from the National Occupational Information Coordinating Committee is used to aggregate the nine-digit occupational information in the DOT to the level of 1980 three-digit census occupations. (See Hundley (1993).)

## V. Empirical Results

### A. Step 2

In this step, which follows the SGA assessment, an applicant is denied if the impairment(s) is not severe, judged in terms of work-related physical, mental, and sensory limitations. Because this is the first step modeled, we use the whole sample of 927 applicants.<sup>6</sup> The results are given in table 1. Individuals with three or more severe ADLs (*SEVEREADL*), one or more severe IADLs (*SEVEREIADL*), poor health (*POORHEALTH*), functional limitations due to mental conditions (*MENTAL*), or who never worked because of a disability (*NEVERWORKED*) are more likely to be passed on (that is, they are not denied).

<sup>6</sup> See Lahiri, Vaughan, and Wixon (1995) for details on the development of the study sample. All empirical results reported here were generated using SAS/STAT version 6.11. Throughout the analysis, all explanatory variables with  $p$ -values greater than 0.10 were dropped and the equation reestimated. Variable definitions are given in appendix A.

TABLE 1.—SEQUENTIAL LOGIT: STEP 2 OF THE SSA DISABILITY DETERMINATION MODEL

Variable	Coefficient	<i>p</i> -Value	Odds Ratio	Marginal Effect	Mean
<i>SEVEREADL</i>	1.569	0.10	4.80	0.190	0.03
<i>SEVEREIADL</i>	0.732	0.06	2.07	0.091	0.09
<i>POORHEALTH</i>	0.722	0.00	2.05	0.087	0.28
<i>GOODHEALTH</i>	-0.468	0.03	0.62	-0.057	0.18
<i>MENTAL</i>	1.144	0.00	3.13	0.139	0.13
<i>DURATION</i>	-0.884	0.00	0.41	-0.107	0.07
<i>OCCASWORK</i>	-1.046	0.01	0.35	-0.127	0.03
<i>NEVERWORKED</i>	1.252	0.08	3.49	0.152	0.03
<i>YOUNG</i>	-0.320	0.10	0.72	-0.039	0.24
<i>ENVIRON</i>	-0.718	0.03	0.48	-0.087	0.20
<i>TITLEII</i>	0.405	0.04	1.49	0.049	0.71
<i>WORK/90</i>	-0.754	0.00	0.47	-0.091	0.11
<i>GENDER</i>	0.577	0.00	1.78	0.070	0.48
<i>RACE</i>	0.327	0.08	1.38	0.040	0.71
<i>YOUNGBLACK</i>	-0.878	0.06	0.41	-0.106	0.03
<i>WORKLOAD</i>	-0.008	0.04	0.99	-0.001	0.06
<i>LOWINCOMEAREA</i>	-0.022	0.01	0.97	-0.003	21.90
<i>BOSTON</i>	-0.814	0.06	0.44	-0.099	0.05
<i>CHICAGO</i>	-0.808	0.00	0.44	-0.098	0.18
<i>ATLANTA</i>	-1.035	0.00	0.35	-0.125	0.23
<i>DALLAS</i>	-0.995	0.00	0.37	-0.121	0.13
pseudo- $R^2 = 0.30$			Goodman-Kruskal Gamma = 0.53		

Dependent variable is 1 if passed on (751 obs.), 0 if denied (176 obs.). See appendix A for definition of variables.

In terms of odds ratios and marginal effects, the impairment-related variables are the most important factors leading to pass-ons. Concurrent applicants—defined as those applying for both DI and SSI—and DI applicants (*TITLEII*) are also more likely to be passed on. Such applicants, as compared to the SSI applicants, are expected to apply with more severe impairments on average, because their opportunity cost of applying is higher. As expected, individuals reporting good health, *GOODHEALTH*, a work limitation but able to work occasionally or irregularly (*OCCASWORK*), or a work limitation of less than twelve months duration (*DURATION*) have a higher probability of denial. The last variable reflects a statutory criterion, the duration test. In terms of health measures, our results show that self-reports of both objective and subjective measures are useful predictors of severity (Sickles & Taubman, 1997). We also find that applicants younger than 35 years of age (*YOUNG*) or whose prior occupations involved hazardous conditions (*ENVIRON*) tend to be denied. The negative coefficient for the latter variable suggests that marginally impaired individuals who worked under hazardous conditions have a higher propensity to apply than those with similar impairments but who are not subject to hazardous conditions. This finding, consistent with Burtless (1987), indicates that hazardous work conditions tend to affect a person's ability to continue working at full capacity, rather than the probability of becoming severely disabled in the sense of having a higher probability of death.

We find that applicants with cases adjudicated in 1990 and with recent work experience (*WORK/90*) tend to be denied. These are likely to be mostly DI and concurrent applicants. This finding is consistent with the longstanding observation that recessionary times induce disability appli-

cations by some with impairments of dubious severity (Leonard, 1986; Stapleton et al., 1995). Conceivably, at the height of the last recession (which officially lasted from July 1990 to March 1991), those who were out of work, with unemployment and health insurance benefits running out, may have found enrolling for disability an attractive alternative, even if only marginally impaired. In addition, SGA—the amount that a beneficiary may earn and still be considered disabled—was raised from \$300 to \$500 in 1990, perhaps inducing some marginally impaired workers to apply. Interestingly, no time dummy variable representing the recession years played a role for people without recent work experience (mostly SSI). This is consistent with evidence reported by Rupp and Stapleton (1995) that the last recession affected DI and concurrent applicants more than it did SSI applicants.

To control for regional differences in allowance rates, we introduce dummies to represent the ten federal regions. There has long been concern over regional differences in allowance rates, specifically, whether applicants from one region are treated differently from those in another (Gallicchio & Bye, 1980; Marvel, 1982; Parsons, 1991). Four region dummies—Boston, Chicago, Dallas, and Atlanta—are significant and all have negative coefficients. Thus, ceteris paribus, applicants from these regions, compared to those from the rest of the country, have a higher likelihood of denial at step 2 for the period studied. This may be explained in part by the reaction of state agencies and ALJs to court decisions in their regions, class-action suits, or so-called acquiescence rulings (whereby SSA disagrees with a court decision but agrees to abide by the decision within the court's jurisdiction). Differences in administrative factors such as experience, training, management, in-

dividual subjectivity, and the local adjudicative climate may play a role in explaining the strong negative coefficient for the Dallas and Atlanta regions.<sup>7</sup> The negative coefficient for the Boston region is unexpected because its overall allowance rate has historically been much higher than the national average. One explanation for the negative coefficient is that this area was hard hit by the last recession, possibly inducing some marginally impaired persons to apply, many of whom were (appropriately) denied at this step.

Differences in allowance rates by region may reflect variation in the economic circumstances driving applications, yielding differing mixes of more and less severely disabled among regional applicant pools. To control for these forces, we also introduce variables representing the economic conditions of the District Office (DO) catchment area in which the applicant resides.<sup>8</sup> These variables are percentage of the civilian labor force unemployed, median home value, median household income, percentage Hispanic, percentage black, percentage female, median years of school completed, percentage age 65 or older, and percentage of households with income under \$15,000. Only the last variable (*LOWINCOMEAREA*) is significant, and it has a negative coefficient. This is consistent with the hypothesis that low-income areas generate more applicants with non-severe impairments who are screened out at this step.

The regression results appear to show that gender and race, which are not intended to play a role in the determination, affect the likelihood of being found severely disabled. Specifically, males and whites (compared to blacks and others) have a higher likelihood of being passed on to the next step. This result is consistent with findings of Bound, Schoenbaum, and Waidmann (1995) that blacks in poor health are more likely to leave the workforce than are whites in poor health. They argue that the difference in responsiveness, at least in part, reflects the different social and economic conditions faced by the two populations. Two U.S. GAO reports (1992, 1994) also found that, for the most part, racial and gender differences in allowance rates could be explained by differences in age, types of impairment, occupation, and other demographic characteristics. However, unlike the previous studies, here we are observing race and gender effects by the steps of the determination process, after controlling—to the extent the data permit—for the severity of impairments, occupational characteristics, and a host of other policy-relevant characteristics of the applicant. In addition to the gender and race effects noted above, our results indicate that younger black applicants (age less than 35 years) have an additional likelihood of being denied (*YOUNGBLACK*). The significance of the *YOUNGBLACK*

variable in our regression underscores a concern expressed in GAO (1992).

The findings from our sequential model support an interpretation that women and blacks may, on average, apply with less severe impairments, possibly due to more-limited occupational opportunities. Although race and gender effects have been noted in overall allow/deny models estimated by other researchers, under our sequential approach these effects are observed only for step 2, which may suggest that these groups are being screened at this initial step for the same reason. This interpretation is consistent with an SSA Quality Assurance (QA) finding that a higher proportion of women apply with marginal impairments than do men.<sup>9</sup>

There has been speculation about possible effects of a surge in workload pressure and backlogs on the accuracy of DDS decisions. Although DDS staffing in 1993 was almost the same as for 1986, during 1989–1993 claims increased from 1.6 million to 2.6 million—an increase of more than 60% in just four years. Not surprisingly, both the backlog of pending claims and the number of dispositions per staff year increased dramatically. In a report to Congress, SSA (1996) considered whether workload pressures could have contributed to the increase in award rates. Several arguments have been made for this hypothesis. First, when workload pressures increase, it may be easier for decision-makers to allow than to deny, because denials require lengthy written rationales (Koitz, Kollmann, & Neisner, 1992). Second, there has been concern within SSA that procedures adopted at the beginning of 1992 to expedite cases more likely to be allowed may have affected the accuracy of some decisions. Specifically, some disability examiners, facing heavy workload pressures, may have made more favorable decisions to exploit the streamlined procedure for likely allowances (DHHS, 1992). Third, to adjust for backlogs, SSA reduced medical reviews of DDS decisions, and questioned fewer DDS decisions under its QA process. In addition, rising caseloads are considered a catalyst for changes in the adjudicative climate that favor allowances (Koitz et al., 1992).

There has been little evidence on the effects of workload pressure and backlogs. To test for such effects, we construct variables at the district office and DDS levels for each year between 1989 and 1993. The variables constructed are (1) difference in dispositions per staff-year from the national average (*WORKLOAD*) for DDS agencies, (2) average processing time for DI and SSI applicants, separately, at the DO level, and (3) actual dispositions as a percentage of bud-

<sup>7</sup> The 709 report (DHHS, 1992) defined adjudicative climate as “the perceptions of individual disability adjudicators, based on the prevailing national attitudes regarding disability, that may affect how they apply existing formal policy in instances where some judgement is required within the specified evaluation procedures” (p. 30).

<sup>8</sup> Muller (1982) studied effects of local labor markets on disability allowances using microdata.

<sup>9</sup> Historically, there has been considerable controversy over the application of the severity criterion, and it has been challenged in many courts. In 1987, the Supreme Court (*Bowen vs. Yuckert* (96 L. 119, 1987)) upheld the severity requirement as both efficient and reliable, because it allows SSA to identify “at an early stage those claimants whose medical impairments are so slight that it is unlikely they would be found to be disabled even if their age, education, and experience were taken into account” (p. 7). See Bloch (1992).

TABLE 2.—SEQUENTIAL LOGIT: STEP 3 OF THE SSA DISABILITY DETERMINATION MODEL

Variable	Coefficient	<i>p</i> -Value	Odds Ratio	Marginal Effect	Mean
<i>HOSPITAL</i>	0.489	0.02	1.63	0.109	0.24
<i>SEVEREADL</i>	1.114	0.06	3.04	0.248	0.03
<i>2IADL</i>	0.721	0.00	2.05	0.161	0.22
<i>MENTAL/NOWORK</i>	0.948	0.01	2.58	0.211	0.06
<i>UNABLETOWORK</i>	-0.352	0.09	0.70	-0.078	0.61
<i>ACCIDENT</i>	-0.468	0.09	0.62	-0.104	0.18
<i>MUSCULO</i>	-0.668	0.01	0.51	-0.149	0.22
<i>NEURO</i>	0.872	0.02	2.39	0.195	0.06
<i>WAITTIME</i>	-0.002	0.03	0.93	-0.001	140.60
<i>NEVERMARRIED</i>	0.358	0.10	1.43	0.080	0.21
<i>WORK/91</i>	0.581	0.02	1.78	0.130	0.15
<i>WORK/92</i>	0.651	0.00	1.97	0.145	0.19
<i>SEATTLE</i>	0.826	0.09	2.28	0.184	0.03
pseudo- $R^2 = 0.16$			Goodman-Kruskal Gamma = 0.41		

Dependent variable is 1 if allowed (264 obs.), 0 if passed on (487 obs.). See appendix A for definition of variables.

geted (that is, projected) dispositions at the DDS level. Only the first variable, *WORKLOAD*, is significant at step 2, but it has an unexpected, negative sign. Thus, we find no evidence for the hypothesis that higher workloads are associated with more-lenient decision making by adjudicators at this step. This is consistent with evidence from SSA's ongoing regional QA reviews of DDS decisions showing that the error rate in DDS allowance decisions is only about 2.5% (DHHS, 1992) and has not changed over the last ten years. Our regression results, on the other hand, indicate that adjudicators under workload pressure may tend to deny more often at step 2. We should point out that dispositions per staff-year varied considerably in our sample across states and over time. In 1989, it ranged from 189 for Tennessee to 251 for Mississippi. In 1993, workload varied from 211 for Hawaii to 333 for Mississippi. The negative effect of the workload variable can be rationalized in a number of ways. First, QA results over the last couple of years have consistently shown that DDSs make more erroneous denials than allowances. For instance, the SSA regional branch QA review found that, in 1991, 7.6% of sampled denials and only 2.8% of sampled allowances were incorrect. Second, as part of the "preeffectuation review" (PER), at least half of all DI allowances (both initial and reconsideration) are reviewed before they go into effect. However, denials are not subject to preeffectuation reviews. Thus, the state-agency examiners face a PER process in which an award is more likely to be questioned than a denial. One might expect that this asymmetric review process would tilt the screening toward denials, particularly when adjudicators face heavy workloads. Nonetheless, the odds ratio and marginal effect for this variable are only 0.992 and -0.001, respectively.

The pseudo- $R^2$  value is 0.30, implying that approximately 30% of the variation of the underlying latent dependent variable is explained.<sup>10</sup> In addition, the Goodman-

Kruskal Gamma measure of rank correlation between the observed outcomes and the predicted probabilities is 0.53.

### B. Step 3

Step 3 of the sequential process involves assessing the nature and severity of impairment(s) based on SSA's detailed list of medical standards, the Listings. Out of a total sample of 927, 176 applications (19%) are denied in step 2, leaving a sample of 751 at step 3. Of these, 264 (35.2%) are allowed at this step. It is worth noting that the survey data on health and disability used here were not designed to estimate eligibility under the Listings, so how well these data explain step 3 allowances is an empirical question.

In table 2, we report logit regression results for step 3, with thirteen statistically significant explanatory variables. Several variables on health and activity limitations increase the probability of allowance, including having three or more severe ADLs (*SEVEREADL*), having two or more IADLs (*2IADL*), having a mental disability and no work history (*MENTAL/NOWORK*). Throughout our analysis, the presence of a reported mental impairment proves the most consistent health-related variable explaining allowances. For step 3, the odds ratio associated with the *MENTAL/NOWORK* variable is 2.58 and the marginal effect is 0.21. Applicants reporting work limitations due to musculoskeletal conditions (*MUSCULO*) tend to be passed on to the next step (odds ratio 0.51), whereas those reporting work limitations due to sensory/neurological conditions (*NEURO*) tend to be allowed more often at this step (odds ratio 2.39). Our finding that those with musculoskeletal impairments are less likely to be allowed at step 3 is consistent with other studies conducted over the past three decades (Nagi, 1969). (See Lahiri, Vaughan, and Wixon (1995) for more details.) Applicants reporting an overnight hospital stay during the last twelve months (*HOSPITAL*) or

<sup>10</sup> We report the pseudo- $R^2$  measure proposed by McKelvey and Zavoina (1975). Many recent studies have favored the McKelvey-Zavoina measure over a number of competing pseudo- $R^2$  measures for logit models. This

measure mimics the ordinary least-squares  $R^2$  in the underlying linear latent model the best, and it also is least vulnerable to the varying proportion of any particular outcome in the sample.

TABLE 3.—SEQUENTIAL LOGIT: STEP 4 OF THE SSA DISABILITY DETERMINATION MODEL

Variable	Coefficient	<i>p</i> -Value	Odds Ratio	Marginal Effect	Mean
<i>FL/HEAVYOCCU</i>	1.552	0.04	4.72	0.287	0.06
<i>SAMEWORK</i>	-0.739	0.03	0.47	-0.137	0.10
<i>NOMENTAL/GED</i>	-0.725	0.10	0.48	-0.134	0.23
<i>WHITECOLLAR/EDU</i>	-0.657	0.01	0.51	-0.121	0.17
<i>NEVERWORKED</i>	1.653	0.00	5.23	0.305	0.17
<i>ENVIRON</i>	0.644	0.01	1.90	0.119	0.36
<i>DIPROCESSTIME</i>	0.012	0.01	1.01	0.002	88.00
<i>BOSTON</i>	1.007	0.09	2.73	0.186	0.06
pseudo- <i>R</i> <sup>2</sup> = 0.18			Goodman-Kruskal Gamma = 0.43		

Dependent variable is 1 if passed on (354 obs.), 0 if denied (133 obs.). See appendix A for definition of variables.

who never married (*NEVERMARRIED*) also have a higher probability of being allowed. We speculate that these variables reflect variation in impairment severity that is not captured by the survey health measures. The most severely impaired may be less likely to marry, for example. Perhaps the major finding for this step is that activity limitations, medical events, and medical conditions are key explanatory variables with relatively high odds ratios and marginal effects. This is expected, because at this step the DDS judges whether an impairment meets or equals the listings.

*WORK/91* and *WORK/92* (that is, 1991 and 1992 year dummies crossed against applicants with recent work experience) are positive, increasing the odds of allowance. Interestingly, *WORK/90* has a positive effect on denials in step 2. We argue above that many marginally impaired workers, after being laid off, apply for benefits and are denied at step 2. However, *WORK/91* and *WORK/92* suggest that such variables have a dual effect: that there are also a number of individuals who have worked despite severe physical or mental impairments. Because employers may have to make accommodations for the health problems of such employees, they may be more vulnerable to layoffs during periods of high unemployment. Moreover, some of these workers, if laid off, can qualify for benefits under the listings. Brehm and Rush (1988) estimated that one in eight men had an impairment that met or equaled the medical listings, yet continued to work even four years after the onset of the disabling condition. *WORK/91* and *WORK/92* may reflect such applicants.<sup>11</sup> Note that the structural approach permits us to trace at what steps business-cycle effects are felt; in this case, the evidence suggests dual and offsetting effects.

We find that applicants reporting a work disability due to an accident (*ACCIDENT*) and those unable to work in at least one of the four interviews (*UNABLETOWORK*) tend to be passed on to the next step. The accident variable may have a negative coefficient, because accident victims are often counseled by hospitals to apply for disability, although many do not meet the listings. Halpern and Hausman (1986) found a similar negative effect, but it was not significant at the 10% level. We also consider effects of average waiting

time (*WAITTIME*), finding that longer times between application and decision at the district-office level reduce the probability of allowance. This negative effect may indicate that, as the waiting time increases, disability examiners tend to pass marginal applicants on to the next step rather than allow at step 3. Like the effect of the *WORKLOAD* variable in step 2, if anything, the estimate shows that adjudicators were slightly less lenient under work pressure. However, as we observed for the *WORKLOAD* variable, the odds ratio (0.93) is close to 1, and the marginal effect is -0.001. Finally, of all the federal regions, only Seattle is significant, and it has a positive sign. As pointed out in the so-called 709 report (DHHS, 1992), following the *Morrison, Doe*, and *Decker* court decision in 1990 (involving the emphasis placed on the opinion of a claimant's treating physician), the initial allowance rate in the Seattle region increased from 38% in 1988-1989 to 50% by mid-1990.

As a measure of the overall explanatory power of this equation, the pseudo-*R*<sup>2</sup> is 0.16 and the Gamma coefficient is 0.41. Our estimated model could be improved if better information on the severity of medical impairments were available in the survey.

### C. Step 4

Step 4 involves evaluating the applicant's residual functional capacity to perform past work, taking into account physical, mental, and other limitations. Out of 487 applicants passed on to this step, 133 (27.3%) are denied. The remaining applicants are passed on to the final step which considers their ability to do any work. The estimates, mainly involving explanatory variables with program-specific underpinnings, are reported in table 3.

We find that applicants reporting difficulty walking a quarter of a mile, walking upstairs, and lifting or carrying 10 lbs., and whose previous occupations required heavy or very heavy work (*FL/HEAVYOCCU*) tend to be denied less often. The odds ratio and the marginal effect of *FL/HEAVYOCCU* are 4.72 and 0.287, respectively. Applicants with no recent work experience or no work history (*NEVERWORKED*) are also much more likely to be passed on to the next step, with an odds ratio of 5.23 and a marginal effect of 0.305. This finding is expected, given that step 4 assesses applicants' ability to perform past work and that

<sup>11</sup> Note that, although the recession of the early 1990s officially ended in March 1991, the percentage of individuals with declining family incomes remained elevated through 1991 and 1992.



TABLE 4.—SEQUENTIAL LOGIT: STEP 5 OF THE SSA DISABILITY DETERMINATION MODEL

Variable	Coefficient	p-Value	Odds Ratio	Marginal Effect	Mean
<i>OLD/LOWEDU</i>	1.136	0.06	3.11	0.279	0.11
<i>YOUNG/NOMENTAL</i>	-1.453	0.00	0.23	-0.356	0.49
<i>OLD</i>	1.007	0.02	2.73	0.247	0.30
<i>UNSKILLED</i>	0.807	0.10	2.24	0.198	0.17
<i>HEAVYOCCU</i>	-0.797	0.03	0.45	-0.195	0.18
<i>OCCASWORK</i>	-2.477	0.02	0.08	-0.608	0.04
<i>YOUNG/SKILLED</i>	-0.765	0.08	0.46	-0.188	0.20
<i>DUMMY91</i>	0.877	0.02	2.40	0.215	0.16
<i>DUMMY92</i>	0.992	0.00	2.69	0.243	0.21
<i>DALLAS</i>	-1.112	0.01	0.32	-0.273	0.12
<i>LOWINCOMEAREA</i>	-0.027	0.05	0.97	-0.001	21.50
pseudo- $R^2 = 0.46$			Goodman-Kruskal Gamma = 0.69		

Dependent variable is 1 if allowed (159 obs.), 0 if denied (195 obs.). See appendix A for definition of variables.

persons lacking recent work experience are properly evaluated in step 5. Of applicants similarly impaired who are considered at step 4, those working in environmentally hazardous conditions (*ENVIRON*) are less likely to be able to work in the same job and are passed on to the next step (Bresnitz et al., 1994). Applicants with no mental conditions and whose previous occupations required higher general educational development, in terms of reasoning, language, and mathematical skills (*NOMENTAL/GED*), tend to be judged able to continue their previous work and are often denied, probably because their work is not physically demanding. We also find, as expected, that applicants who report being able to do the same kind of work despite their disability (*SAMEWORK*) and those working in white-collar occupations and having more than twelve years of education (*WHITECOLLAR/EDU*) have higher probabilities of being denied.

Of all the variables we construct to represent the high workload of recent years, only one—the mean processing time for DI claims at the DO level (*DIPROCESSTIME*)—is significant, but its marginal effect is small. The mean processing time for SSI claims is not significant; this is expected, because SSI applicants, who have less work experience than DI applicants, are readily passed on to the next step. The Boston region (consisting of Maine, New Hampshire, Vermont, Massachusetts, Connecticut, and Rhode Island) is significant with a positive coefficient, implying that applicants from this region are less likely to be denied at step 4. This finding may reflect two court cases (*Aldrich* in Vermont and *Schisler* in Vermont, New York, and Connecticut) which dealt with the importance of the opinions of the treating physicians in evaluating a claimant's subjective complaints of pain and in other cases. In early 1990, adjudicators in the Vermont DDS received extensive training in evaluating pain cases. Following this training, the overall allowance rate in Vermont soared to more than 60% and continued in the 50%–55% range for two years. Connecticut also experienced an increase in the allowance rate over the 1990–1991 period, which may be attributed to these court cases. The pseudo- $R^2$  for the step 4

regression is 0.18, and the gamma measure for model fit is 0.43.

#### D. Step 5

At step 5, the applicant's age, education, and work experience are used, in conjunction with the analysis of residual functional capacity, to determine whether he or she can do any work in the economy. The regression results are presented in table 4. Out of 354 applicants evaluated at this step, 159 (44.9%) are allowed and 195 (55.1%) are denied. The variables we find to have notable positive effects on the probability of allowance are aged 55 or older (*OLD*), old with low education (*OLD/LOWEDU*), and *UNSKILLED*. These three variables are dominant in terms of their odds ratios and marginal effects; furthermore, they have clear programmatic underpinnings, given the roles of age and education in the vocational grid. On the other hand, younger applicants (55 or younger) with no mental conditions (*YOUNG/NOMENTAL*), applicants who do occasional work (*OCCASWORK*), and applicants who are young and skilled (*YOUNG/SKILLED*) tend to be denied more often on grounds that they could do some work. Of these, the *OCCASWORK* variable is most dominant, having an odds ratio of 0.08 and a marginal effect of -0.608. These variables reflect the structure of the vocational grid; hence, the survey variables play roles consistent with the statutory criteria underlying step 5.

In addition, year dummies for 1991 and 1992 are strongly positive, indicating that applications reaching step 5 were more likely to be allowed in those years. There are several possible explanations. Following a series of adverse circuit-court decisions, the Omnibus Budget Reconciliation Act of 1990 (OBRA 90) changed the way that SSA adjudicates disability claims of surviving divorced spouses and widowed beneficiaries (DWBs). Prior to OBRA 90, disability was determined for such claimants by using medical criteria to establish the inability to do any "gainful activity." OBRA 90 made the criteria for DWBs consistent with that for other Title II disability claims, requiring adjudicators to also

TABLE 5.—PSEUDO-STRUCTURAL MODEL OF DISABILITY DETERMINATION

Variable	Coefficient	p-Value	Odds Ratio	Marginal Effect	Mean
<i>WORK/90</i>	-0.629	0.02	0.53	-0.156	0.11
<i>OCCASWORK</i>	-1.387	0.03	0.25	-0.343	0.03
<i>LOWINCOMEAREA</i>	-0.014	0.08	0.98	-0.003	21.9
<i>TITLEII</i>	0.345	0.05	1.41	0.085	0.71
<i>YOUNGBLACK</i>	-1.150	0.03	0.31	-0.284	0.03
<i>HOSPITAL</i>	0.373	0.02	1.45	0.092	0.38
<i>SEVEREADL</i>	0.584	0.00	1.79	0.144	0.20
<i>NEVERMARRIED</i>	0.470	0.02	1.60	0.116	0.21
<i>ACCIDENT</i>	-0.619	0.00	0.53	-0.153	0.18
<i>NEURO</i>	0.589	0.10	1.80	0.146	0.06
<i>DIPROCESSTIME</i>	0.012	0.00	1.01	0.003	86.60
<i>YOUNG/NOMENTAL</i>	-1.003	0.00	0.36	-0.248	0.53
pseudo-R <sup>2</sup> = 0.18			Goodman-Kruskal Gamma = 0.44		

Dependent variable is 1 if allowed (423 obs.), 0 if denied (504 obs.). See appendix A for definition of variables.

assess the ability of DWBs to do prior work or any work in the economy. Because many such claimants were over age 50, with limited education and work experience, they became excellent candidates for favorable vocational grid decisions at this step. During 1991–1992, the DWB allowance rate was approximately twice the rate prior to the statutory change. Since 1993, this rate has stabilized. In addition, responding to an escalating backlog of claims, SSA made major administrative changes in 1992 to expedite case processing, such as funding overtime and other short-term changes. As a result, allowance rates in 1991 and 1992 might have been artificially high; possibly the rates may have risen as expedited cases were processed and fallen as tougher cases were taken up in 1993. Our data are consistent with this hypothesis, showing that step 5 allowance rates in 1991 and 1992 were 54% and 51%, respectively, compared to only 34% in 1993. These dummies may also partly reflect SSA's increased effort on SSI outreach. Changes in state and local programs may have increased incentives to apply for DI and SSI as well. In addition, as argued above, the effects of the economic recession of the early 1990s and longer-term structural changes in the job market adversely affecting low-skilled workers were played out in this context. As Burkhauser, Haveman, and Wolfe (1993) have shown, more vulnerable workers, like the disabled and minorities, take the longest to recoup the ground lost during a recession. (See also footnote 11.) To control for regional variation, we introduce the regional dummies and DO-specific variables representing local socioeconomic conditions. Only the Dallas region (New Mexico, Texas, Oklahoma, Arkansas, and Louisiana) and percentage of households with income less than \$15,000 (*LOWINCOMEAREA*) are significant and both have negative signs. *DALLAS* and *LOWINCOMEAREA* are also significant with negative coefficients for step 2. However, given its odds ratio (0.97) and marginal effect (-0.001), *LOWINCOMEAREA* does not seem important.

The model fit for step 5 is the best of the four regressions. The pseudo-R<sup>2</sup> is 0.46 and the gamma coefficient is 0.69. The good fit at this step may be attributable to the objective nature of the criterion used at this step (the vocational grid

and to the exploitation of survey variables on demographic and labor force characteristics, rather than self-rated health variables.

#### E. The Pseudo-Structural Approach

We mentioned above that it is difficult to identify effects of program-specific variables on overall eligibility outcomes unless they are introduced at the appropriate steps of the determination. The typical approach has been to calculate eligibility probabilities by regressing overall allowed/denied decisions on a host of health, socioeconomic, and demographic variables. To illustrate this approach, we run a similar logit regression with all explanatory variables for the four individual regressions. The results are reported in table 5, which lists only variables that are statistically significant at the 10% level.

The estimated pseudo-structural equation with overall allowed/denied decisions as the dependent variable has a pseudo-R<sup>2</sup> value of 0.18 and a gamma value of 0.44, but only twelve of the 49 distinct independent variables have p-values less than 0.10. Only five of the twelve are impairment-related: *SEVEREADL*, *ACCIDENT*, *HOSPITAL*, *NEURO*, and *YOUNG/NOMENTAL*. The seven nonimpairment variables are *WORK/90*, *OCCASWORK*, *LOWINCOMEAREA*, *TITLEII*, *YOUNGBLACK*, *NEVERMARRIED*, and *DIPROCESSTIME*. By comparing estimates of the pseudo-structural equation with those from the four-step sequential model, we find that some objective measures of disability based on ADLs, functional limitations, and mental conditions (which were major factors in steps 2 and 3) are no longer important. In addition, almost all variables representing residual functional capacity, age, education, skill, and occupation that are significant in the last two steps are no longer relevant.

#### VI. Predictive Performance

One potential advantage of structural modeling is that covert changes in the nature and characteristics of appli-

cants can be reflected in model predictions in an appropriate manner. The ultimate success of the structural model depends on how well it predicts final allowances and denials at the level of individuals. Predicting an aggregate percentage such as the percentage of the population eligible for a public benefit—or, as in our case, the percentage of applicants allowed—is also an important problem often faced by policymakers. In our structural model, there are two ways a disability applicant can be allowed under the determination process: (1) at step 3, one's impairment can meet or equal the listings, conditioned on not being denied at step 2; and (2) at step 5, one can be allowed, conditioned on not being denied at step 2, not being allowed in step 3, and, finally, not being denied in step 4. Thus, as explained in section III, the final probability of being eligible can be expressed as

$$\begin{aligned} \Pr(a_3) + \Pr(a_5) &= F(\hat{\beta}'X_i)F(\hat{\alpha}'W_k) \\ &+ F(\hat{\delta}'Z_n) \cdot F(\hat{\nu}'Y_m) \\ &\times [1 - F(\hat{\beta}'X_i)] \cdot F(\hat{\alpha}'W_k). \end{aligned} \quad (6)$$

See also equation (2) and (4). Note that each applicant will have a nonzero probability of a particular outcome at each step of the sequential process. Given the observed health indicators and other characteristics of an individual, we first generate  $F(\hat{\alpha}'W_k)$ ,  $F(\hat{\beta}'X_i)$ ,  $F(\hat{\gamma}'Y_m)$ , and  $F(\hat{\delta}'Z_n)$  for each individual from the four estimated logit regressions. These individual probabilities are appropriately processed as shown in equation (6) to calculate the total probability of allowance for that particular individual. Note that  $[\Pr(a_3) + \Pr(a_5)] + [\Pr(d_2) + \Pr(d_4) + \Pr(d_5)] = 1$  (that is, the final probabilities of allowance and denial must sum to unity for each individual). To obtain a consistent estimate of the prediction for the percentage of allowances, the probabilities are then averaged over the sample (Amemiya, 1985).

We studied the out-of-sample predictive capacity of the sequential model and compared it with that of the pseudo-structural model through extensive split-sample analysis. First, we estimated the four regressions corresponding to the steps of the sequential determination process, using data from 1989 to 1993. We repeated the estimation five times, each time “jackknifing” all observations for one of the five years at a time. We then calculated the out-of-sample prediction for the final eligibility probability using equation (6) for all individuals included in the sample year set aside.<sup>12</sup> For instance, we first estimated the structural equations for steps 2 through 5 using data for 1989–1992. (The sample sizes for the regressions for steps 2 through 5 are 641, 517, 325, and 233, respectively.) These estimated regressions are used to predict  $a_3$  and  $a_5$  probabilities for each of the 286 applicants in 1993 using equation (6). The out-of-sample

<sup>12</sup> In order not to benefit from retrospection, we did not include the year dummies and variables defined using these in generating the “jackknife” predictions. Thus, we did not include *WORK/90* for step 2 and for the pseudo-structural model, *WORK/91* and *WORK/92* for step 3, nor *DUMMY91* and *DUMMY92* for step 5.

predicted allowance rate was found to be 38.9%. The actual allowance rate for the 1993 applicants was 39.5%, indicating a 1.6% underestimation of allowances. We generated similar out-of-sample predictions using the single pseudo-structural equation of table 5 as well. When calculated over all 927 sample applicants, the predicted average allowance rate was 44.8% using the sequential approach and 48.2% using the pseudo-structural approach. The actual allowance rate in our sample was 45.6%. Thus, the structural model underestimated by 1.8%, whereas the pseudo-structural approach overestimated by 2.6%.

Diebold and Mariano (1995) recently catalogued a number of procedures to test for the null hypothesis of equal accuracy for two competing probability forecasts based on individual-level forecast differentials. Let us denote  $d_i =$  (forecasts from the sequential model—the forecasts from the pseudo-structural model, for individual  $i$ ). We calculated four different test statistics: a simple asymptotic  $N(0, 1)$  test based on the mean and variance of  $d_i$ , studentized versions of the sign test and Wilcoxon's Signed-Rank test, and the Morgan-Granger-Newbold Student's  $t$  test for equal forecast accuracy. The statistics were calculated as  $-7.86$ ,  $-10.64$ ,  $-11.56$ , and  $-11.38$ , respectively—all resoundingly rejecting the equal-accuracy hypothesis at any reasonable level of significance. These significance tests on the out-of-sample predictive efficiency complement the evidence we have presented in Lahiri et al. (1995) and Hu et al. (1997) on the superiority of the sequential model for predictive purposes.

As a first attempt to model the structure of the SSA disability determination process using self-reported SIPP survey data and data from other sources, we find that the sequential model yields quite accurate predictions. This holds despite evidence that the standard (*viz.*, the DDS determination decision) against which we are evaluating the survey predictors is subject to some uncertainty. As we note above, quality assurance results consistently show that nearly 5% of the initial decisions can be erroneous.<sup>13</sup>

## VII. Conclusions

In this paper, we model the structure of the SSA disability determination process using household survey data on health, demographic traits, work, and activity limitations. We use data from the 1990 panel of the Survey of Income and Program Participation that have been exact-matched to administrative records on disability determinations for SIPP sample members who applied for disability benefits between 1989 and 1993. The study also exploits data on

<sup>13</sup> Based on an experiment, Nagi (1969) reported that, in the opinion of independently appointed clinicians, SSA adjudicators misjudged almost 27% of all initial claims. Similar findings were reported in Gallicchio and Bye (1980) and other studies. Hence, to facilitate equity and uniformity in decisions, Congress enacted the Disability Insurance Amendments of 1980 (PL 96-265), to tighten the performance standards of state DDSs. Since then, error rates on initial DI and SSI decisions have fallen steadily, stabilizing at less than 5% for the last ten years.

current and past occupations of applicants from the *Dictionary of Occupational Titles*, workload pressures at the DDSs, and local-area economic conditions from unpublished SSA sources.

We model the structure of the disability evaluation process using a four-step sequential logit model. The key is that, under the program provisions, different decision criteria dictate the outcomes at distinct steps of the determination process. For instance, steps 2 and 3 are medical in nature, whereas the last two steps are based, in part, on applicants' residual functional capacity, age, education, and past occupations. Detailed administrative information on outcomes at each step of the determination make the multistaged approach possible.

The typical approach in previous studies has been to run a single logit regression of overall allowed/denied decisions, on a host of health, socioeconomic, and demographic variables. We compare such a pseudo-structural equation approach with our model of the actual sequential structure. Specifically, we demonstrate that, without the multistaged structural approach, it is difficult to estimate the impact of more objective survey health variables (such as functional limitations, ADLs, and IADLs) on the determination outcome. Also, some variables representing the vocational grid are not significant in the pseudo-structural equation. Hence, many variables with program-specific underpinnings that we find important in our staged sequential regressions do not have discernible effects in the pseudo-structural logit regression. As a result, we find that the split-sample predictions of overall allowance rates from the sequential logit model perform considerably better than do those generated by the pseudo-structural model.

Throughout our analysis, the presence of a reported mental impairment proves to be an important health-related variable explaining allowances and pass-ons. This is of interest, given the changes in the handling of mental impairments introduced in the mid- and late 1980s. Our regression results also demonstrate that survey data on functional, ADL, IADL, and work limitations, including those related to mental impairments, can, when used in the context of a structural model, be of considerable value in predicting SSA allowances and denials.

We find evidence that particular federal regions or judicial districts affect allowance probabilities. Crude allowance rates vary significantly across regions, raising concerns that applicants from different regions of the country may be treated differently. Considering this issue requires evaluating class-action law suits and SSA's "acquiescence" rulings. Our estimates suggest that applicants from the Seattle region (in step 3) and those from the Boston region (in step 4) had elevated probabilities of favorable treatment (allowance or pass on) attributable to three major court decisions: *Aldrich* in Vermont; *Schisler* in Vermont, Connecticut, and New York; and *Morrison, Doe, and Decker* in Seattle. Four federal regions (Dallas, Atlanta, Boston, and Chicago) have

negative and significant effects in step 2, possibly due to differences in the applicant pools for these regions. Further research is needed.

Finally, we consider several workload variables at each step to test whether DDS adjudicators became more lenient due to increased workload pressures in the years studied. At step 2, one variable tested (*WORKLOAD*) is significant, but with an unexpected sign. Similarly, at step 3, the variable *WAITTIME* is significant but with an unexpected sign. A third variable *DIPROCESSTIME* is significant at step 4 with the expected, positive sign. However, for all three variables, the odds ratios and marginal effects suggest their quantitative effects are minuscule. Hence, we see almost no evidence that workload pressures led to more lenient decisions. We estimate the eligibility model using data only on applicants, without appropriately controlling for economic and other factors that induce prospective applicants to actually apply. Presumably, much of the unobserved heterogeneity that exists among applicants can be explained once we develop and estimate a model describing application behavior.

#### REFERENCES

- Amemiya, Takeshi, *Advanced Econometrics* (Cambridge: Harvard University Press, 1985).
- Bloch, Frank S., *Disability Determination: The Administrative Process and the Role of Medical Personnel*, Studies in Welfare Programs and Policies 13 (Westport: Greenwood Press, 1992).
- Bound, John, Michael Schoenbaum, and Timothy Waidmann, "Race and Education Differences in Disability Status and Labor Force Attachment," *Journal of Human Resources* 30 (Supplement 1995), S227–S267.
- Brehm, Henry P., and Thomas V. Rush, "Disability Analysis of Longitudinal Health Data: Policy Implications for Social Security Disability Insurance," *Journal of Aging Studies* 2:4 (1988), 379–399.
- Bresnitz, Eddy, Howard Frumkin, Lawrence Goldstein, David Neumark, Michael Hodgson, and Carolyn Needleman, "Occupational Impairment and Disability among Applicants for Social Security Disability Benefits in Pennsylvania," *American Journal of Public Health* 84 (November 1994), 1786–1790.
- Burkhauser, Richard V., Robert H. Haveman, and Barbara L. Wolfe, "How People with Disabilities Fare When Public Policies Change," *Journal of Policy Analysis and Management* 12 (Spring 1993), 251–269.
- Burtless, Gary, "Occupational Effects on the Health and Work Capacity of Older Men" (pp. 103–142), in Gary Burtless (Ed.), *Work, Health, and Income Among the Elderly* (Washington, DC: Brookings, 1987).
- Diebold, Francis X., and Roberto S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–263.
- Gallicchio, Sal, and Barry Bye, "Consistency of Initial Disability Decisions Among and Within States," Staff Paper 39, Office of Research and Statistics, Social Security Administration (1980).
- Halpern, Janice, and Jerry A. Hausman, "Choice Under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program," *Journal of Public Economics* 31 (November 1986), 131–161.
- Haveman, Robert, Philip R. de Jong, and Barbara Wolfe, "Disability Transfers and the Work Decision of Older Men," *The Quarterly Journal of Economics* 106:3 (1991), 939–949.
- Hu, Jianting, Kajal Lahiri, Denton R. Vaughan, and Bernard Wixon, "A Structural Model of Social Security's Disability Determination Process," working paper #72, Office of Research, Evaluation, and Statistics, Social Security Administration (1997).
- Hundley, Greg, "The Effect of Comparable Worth in the Public Sector on Public/Private Occupational Relative Wages," *Journal of Human Resources* 28 (Spring 1993), 318–342.

- Koitz, David, Geoffrey Kollmann, and Jennifer Neisner, *Status of the Disability Programs of the Social Security Administration*, CRS Report to Congress (Washington, D.C.: Library of Congress, September 1992).
- Kreider, Brent, "Workers' Applications to Social Insurance Programs When Earnings and Eligibility Are Uncertain," *Journal of Labor Economics* 16:4 (1998), 848–877.
- Lahiri, Kajal, Denton R. Vaughan, and Bernard Wixon, "Modeling Social Security's Sequential Disability Determination Using Matched SIPP Data," *Social Security Bulletin* 58 (April 1995), 3–42.
- Leonard, Jonathan S., "Labor Supply Incentives and Disincentives for Disabled Persons" (pp. 64–94), in M. Berkowitz and M. Anne Hill (Eds.), *Disability and the Labor Markets* (Ithaca, NY: Cornell University Press, 1986).
- Marvel, Howard P., "Social Security Disability Insurance," *Journal of Human Resources* 17:3 (1982), 393–412.
- Mashaw, J. F., *Bureaucratic Justice: Managing Social Security Disability Claims* (New Haven, CT: Yale University Press, 1983).
- McFadden, Daniel, "The Revealed Preference of a Government Bureaucracy: Empirical Evidence," *Bell Journal of Economics* 7:1 (1976), 55–72.
- McKelvey, R., and W. Zavoina, "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology* 4 (1975 Summer), 103–120.
- Muller, L. Scott, "The Impact of Local Labor Market Characteristics on the Disability Process," working paper #27, Office of Research and Statistics, Social Security Administration (1982).
- Nagi, Saad Z., *Disability and Rehabilitation: Legal, Clinical, and Self-Concepts and Measurement* (Columbus, OH: Ohio State University Press, 1969).
- Parsons, Donald O., "Self-Screening in Targeted Public Transfer Programs," *Journal of Political Economy* 99 (August 1991), 859–876.
- Rupp, Kalman, and David Stapleton, "Determinants of the Growth in the Social Security Administration's Disability Programs—An Overview," *Social Security Bulletin* 58 (April 1995), 43–70.
- Sickles, Robin C., and Paul Taubman, "Mortality and Morbidity Among Adults and the Elderly" (pp. 559–643), in Mark R. Rosenzweig and Oded Stark (Eds.), *Handbook of Population and Family Economics* (Amsterdam: North-Holland, 1997).
- Stapleton, David, Burt Barnow, Kevin Coleman, Kimberly Dietrich, Jeff Furman, and Gilbert Lo, *Labor Market Conditions, Socioeconomic Factors and the Growth of Applications And Awards for SSDI and SSI Disability Benefits: Final Report*, The Office of the Assistant Secretary for Planning and Evaluation, Lewin-VHI, Inc. and The Department of Health and Human Services (1995).
- Stern, Steven, "Measuring the Effects of Disability on Labor Force Participation," *Journal Human Resources* 24 (January 1989), 361–395.
- U.S. Department of Health and Human Services (DHHS), *The Social Security Disability Insurance Program: An Analysis*, The 709 Report (Washington, D.C.: DHHS, 1992).
- U.S. Department of Labor (DOL), *Selected Characteristics of Occupations Defined in the Dictionary of Occupational Titles, Revised Fourth Edition*, Employment and Training Administration, U.S. Employment Service (Washington, D.C.: U.S. GPO, 1993).
- U.S. General Accounting Office (GAO), *Social Security: Racial Difference in Disability Decisions Warrants Further Investigation* (Washington, D.C.: GAO, April 1992).
- , *Social Security Disability: Most Gender Differences Explained* (Washington, D.C.: GAO, May 1994).
- U.S. Social Security Administration (SSA), *Growth in Disability Insurance Program Costs*, Report to the Congress (February 1996).
- BOSTON* Region I (Maine, New Hampshire, Vermont, Massachusetts, Connecticut, and Rhode Island)
- CHICAGO* Region V (Minnesota, Michigan, Indiana, Ohio, Wisconsin, and Illinois)
- DALLAS* Region VI (New Mexico, Texas, Oklahoma, Arkansas, and Louisiana)
- DIPROCESSTIME* Mean overall processing time for all DI initial claims
- DUMMY91* Decision year was 1991
- DUMMY92* Decision year was 1992
- DURATION* Work limited but lasted less than twelve months in wave 6
- ENVIRON* Previous occupation involved four or more hazardous work conditions. For each nine-digit occupation, DOT identifies the following work conditions: exposure to weather, extreme cold, extreme heat, wet and/or humid, noise, vibration, atmospheric conditions; proximity to moving mechanical parts, electrical shock, high or exposed places, or radiant energy, and working with explosives, toxic or caustic chemicals, and other hazards. The (0, 1) dummies indicating four or more hazardous work conditions at the nine-digit level were aggregated to the three-digit level using Crosswalk
- FL/HEAVYOCCU* Having one of the three functional limitations (FL)—difficulty walking a quarter of a mile, walking upstairs, and lifting and carrying 10 lbs. (wave 6), and previous occupation required moving objects of 10–20 lbs. (heavy work) or in excess of 20 lbs. (very heavy work) constantly, as defined in DOT. First, the (0, 1) dummies for the strength factor at the nine-digit level are aggregated to the three-digit level using the Crosswalk Database Program. We then define heavy occupation as a dummy taking the value 1 if the aggregated value at the three-digit level was nonzero, and 0 otherwise.
- GENDER* Male
- GOODHEALTH* Excellent or very good health on a five-point self-evaluative scale in wave 6
- HEAVYOCCU* Previous occupation requiring heavy or very heavy work
- HOSPITAL* Hospital stay overnight or longer in past twelve months in both waves 3 and 6
- LOWINCOMEAREA* Percentage of households with income under \$15,000 at district-office level (Census data)
- MENTAL* Functional limitations caused by mental conditions in wave 6
- MENTAL/NOWORK* Mental condition and no work history in both waves 3 and 6
- MUSCULO* Functional limitation due to musculoskeletal condition in wave 6
- NEURO* Functional limitation due to sensory or neurological disorder in wave 6
- NEVERMARRIED* Never married (wave 3)
- NEVERWORKED* Never been able to work at a job reported in wave 2
- NOMENTAL/GED* No mental conditions and occupation requiring general educational development (GED). The GED scale ranging from 1 to 6 is composed of three divisions: Reasoning Development, Mathematical Development, and Language Development. Using Crosswalk, we used the average of these scores aggregated over all nine-digit occupations.
- OCCASWORK* Able to work occasionally or irregularly in wave 2

## APPENDIX A

## Variable Definitions

<i>2IADL</i>	Two or more IADLs in both waves 3 and 6
<i>ACCIDENT</i>	Health condition caused by an accident or injury in wave 2
<i>ATLANTA</i>	Region IV (Kentucky, Tennessee, North Carolina, South Carolina, Alabama, Mississippi, Georgia, and Florida)

<i>OLD</i>	Older than 55 years	<i>WAITTIME</i>	Average waiting time (in days) between filing date and decision date at DO
<i>OLD/LOWEDU</i>	Older than 55 years, education less than twelve years, and previous occupation unskilled as defined in DOT (SVP requiring short demonstration up to one month only)	<i>WHITECOLLAR/EDU</i>	Education $\geq$ 12 years and white-collar occupation (sales and services)
<i>POORHEALTH</i>	Poor on a five-category self-evaluative scale of health status, in wave 6	<i>WORK/90</i>	Recent work experience (wave 2) and decision year of 1990
<i>RACE</i>	White	<i>WORK/91</i>	Recent work experience and the decision year was 1991
<i>SAMEWORK</i>	Work limitations, but able to do the same kind of work reported in wave 2	<i>WORK/92</i>	Recent work experience and the decision year was 1992
<i>SEATTLE</i>	Region X (Washington, Oregon, Idaho, and Alaska)	<i>WORKLOAD</i>	Difference in dispositions per FTE staff year between state and national levels collected from DDS staffing and workload reports for 1989–1993
<i>SEVEREADL</i>	Having three or more severe ADLs in wave 6	<i>YOUNG</i>	Age less than 35 on filing date
<i>SEVEREIADL</i>	Having one or more severe IADLs reported in waves 3 and 6	<i>YOUNGBLACK</i>	Age less than 35 years and black
<i>TITLEII</i>	Title II or concurrent (SSA 831 data)	<i>YOUNG/NOMENTAL</i>	Younger than 54 years and no mental conditions
<i>UNABLETOWORK</i>	Unable to work at least once in one of the four waves	<i>YOUNG/SKILLED</i>	Previous occupation was skilled (SVP more than six months), age less than 54 years, and had one or more severe functional limitations or ADLs
<i>UNSKILLED</i>	Previous occupation requiring SVP between short demonstration and up to one month		