NBER WORKING PAPER SERIES

DISABILITY BENEFITS, CONSUMPTION INSURANCE, AND HOUSEHOLD LABOR SUPPLY

David Autor
Andreas Ravndal Kostol
Magne Mogstad
Bradley Setzler

Working Paper 23466
http://www.nber.org/papers/w23466

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2017

This research was supported by the U.S. Social Security Administration through grant #1 DRC12000002-03 to the National Bureau of Economic Research as part of the SSA Disability Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER. The project also received financial support from the Norwegian Research Council. We are grateful to Richard Blundell, Raj Chetty, Amy Finkelstein, Kai Liu, Nathan Hendren, Hamish Low, Luigi Pistaferri, Alessandra Voena, three anonymous referees, and the editor of this journal for valuable input and guidance, and to Knut Brofoss, Espen Vihle and Runar Narvland for their help in accessing the data and in understanding the institutional details. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by David Autor, Andreas Ravndal Kostol, Magne Mogstad, and Bradley Setzler. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Disability Benefits, Consumption Insurance, and Household Labor Supply
David Autor, Andreas Ravndal Kostol, Magne Mogstad, and Bradley Setzler
NBER Working Paper No. 23466
June 2017, Revised August 2018
JEL No. H53,H55,I38,J22

ABSTRACT

There is no evaluation of the consequences of Disability Insurance (DI) receipt that captures the effects on households' net income and consumption expenditure, family labor supply, or benefits from other programs. Combining detailed register data from Norway with an instrumental variables approach based on random assignment to appellant judges, we comprehensively assess how DI receipt affects these understudied outcomes. To consider the welfare implications of the findings from this instrumental variables approach, we estimate a dynamic model of household behavior that translates employment, reapplication and savings decisions into revealed preferences for leisure and consumption. The model-based results suggest that on average, the willingness to pay for DI receipt is positive and sizable. Because spousal labor supply strongly buffers the household income and consumption effects of DI allowances, the estimated willingness to pay for DI receipt is smaller for married than single applicants.

David Autor
Department of Economics, E52-438
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
dautor@mit.edu

Andreas Ravndal Kostol
Norges Bank Research
Bankplassen 2, Oslo
and Statistics Norway
andreas.kostol@norges-bank.no

Magne Mogstad
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
magne.mogstad@gmail.com

Bradley Setzler
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
bradley.setzler@gmail.com

1 Introduction

Over the past 50 years, disability insurance (DI) rolls have risen dramatically in many OECD countries. In the U.S., SSDI benefits receipt has risen from less than 1 percent to 4.7 percent of the non-elderly adult population between the program's inception in 1956 and the present (U.S. Social Security Administration, 2017). In many European countries, the increases are even more striking, from 1 percent to 7 percent in the U.K and from 2 percent to almost 10 percent in Norway. These increases have made DI one of the largest transfer programs in most OECD countries. In the U.S., for example, outlays for DI exceed those for food stamps, traditional cash welfare, or the Earned Income Tax Credit.[1] For families without small children, DI is often the primary cash benefit available after unemployment benefits expire, and it has become an increasingly important component of the social safety net in numerous industrialized countries (OECD, 2010).

To potentially curtail DI program growth, several countries have significantly tightened disability screening criteria, and many others are considering similar policies.[2] These enhanced gatekeeping policies can reduce the fiscal burden of disability insurance, both by lowering the DI caseload and by increasing tax revenue if rejected applicants return to work. At the same time, stricter screening may result in net welfare losses if individuals and families value public disability insurance at more than its fiscal cost.[3] Assessing this tradeoff requires a comparison of the public costs and private benefits of DI awards for applicants at the margin of allowance versus denial, since it is their outcomes that would be changed by shifts in screening stringency. To implement this comparison, we need data on two economic quantities that are rarely measured: the economic value that individuals and families place on disability insurance; and the full cost of DI allowances to taxpayers, summing over DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts. Credibly estimating these quantities is typically hindered both by a lack of comprehensive linked data measuring these many outcomes, and by the difficulty of distinguishing the causal effects of DI receipts from the many unobserved factors that simultaneously determine disability status, earnings, tax payments and transfer receipts, and consumption.

This paper addresses both the measurement and the identification challenge in the context of Norway's DI system, enabling us to offer empirical evidence on the fiscal costs, income and consumption gains, and welfare consequences of DI receipt. Our work draws on two strengths of

[1] In 2016 the U.S. paid out \$143 billion to 10.6 million disabled workers and their families, with an additional \$49 billion in federal SSI payments to blind and disabled workers, and approximately \$107 billion on Medicare expenditures for disabled workers (OASDI Trustees Report, 2017, Table III.A.5; SSI Annual Report, 2017, Table IV.C1; 2017, Table II.B1). In 2013, DI cash payments constituted 1.4 percent of GDP in the U.S. and 1.7 percent of GDP across the European OECD-countries (OECD, 2015).

[2] For example, the U.S. tightened the criteria for new disability awards in the late 1970s and introduced an aggressive program of continuing disability reviews in 1980; however, Congress responded by halting the reviews and, in 1984, liberalizing the program's screening criteria along several dimensions. Another example is the Netherlands: in 1994, the eligibility criteria were tightened and the growth in DI rolls reversed.

[3] In the U.S., all private disability insurance is provided through employer-based group policies. These policies 'wrap-around' the public SSDI system, so that most of the wage insurance risk and all of the medical cost risk is ultimately borne by the public program (Autor *et al.* , 2014). There is not a strong standalone private market in disability insurance, likely because of adverse selection. In the Norwegian setting that we study, private disability insurance is rare.

the Norwegian environment. First, Norwegian register data allow us to characterize the household impacts and fiscal costs of disability receipt by linking employment, taxation, benefits receipt, and assets at the person and household level. Our measure of fiscal costs includes virtually all forms of government cash transfers and revenues from (direct) taxes, and accounts for changes in labor supply and substitution to other transfer programs. Our measures of household impacts of DI receipt include net government transfer payments from all sources, employment and earnings of DI applicants (both allowed and denied) and their spouses, as well as households' total income and measures of their consumption expenditure. Second, we obtain plausibly exogenous variation in DI allowances by exploiting the random assignment of DI applicants to Norwegian judges who differ systematically in their leniency. As a measure of judge leniency, we use the average allowance rate in all other cases a judge has handled. This leniency measure is used as an instrumental variable for DI receipt, as it is highly predictive of judicial rulings in incumbent cases but uncorrelated with case characteristics. This instrumental variables approach recovers the causal effects of DI allowance on individuals at the margin of program entry.

Our first set of analyses, which estimate the causal effects of DI receipt on earnings, total income, consumption expenditure, and fiscal costs, yields four main findings. First, granting DI benefits to applicants on the margin of program entry induces a fall in annual earnings of approximately $5,200, which is about 45 percent of the annual DI transfer benefit awarded. Second, DI allowances raise average household income and consumption expenditure by 16 and 18 percent, implying that DI receipt provides partial consumption smoothing across states of nature for a given individual. Third, the external costs to taxpayers from providing DI benefits—stemming from transfer payments and reduced payroll tax revenues—substantially exceed the net increases in household incomes accruing to DI beneficiaries. Fourth, the consequences of DI allowances differ substantially by marital status. Among single and unmarried applicants, DI awards have large direct impacts on household income and consumption expenditure—incrementing each by about 40 percent relative to baseline. Conversely, DI allowances do not significantly increase the household incomes or the consumption of married applicants on average; indeed, we can reject positive impacts of more than nine percent of baseline income. The reason is that spousal labor supply adjustments and benefit substitution are estimated to offset the effect of DI transfers on household incomes—though we stress that this does not imply that household welfare is unaffected by these transfers.

These causal effects estimates provide key data points for a welfare analysis, but they do not by themselves tell us how much DI allowances affect household welfare, since this also depends on the preferences for leisure and consumption. To explore these welfare implications, we estimate a dynamic model of household behavior with heterogeneous, forward-looking individuals. The model translates employment, savings and reapplication decisions of applicants and their spouses into revealed preferences for leisure and consumption. Brought to the data, the model matches well the instrumental variables estimates of the impact of DI allowances, and moreover, provides plausible parameter estimates for labor supply elasticities. We use the estimated model to compute the welfare benefits of DI receipt—by which we mean the cash equivalent value of receiving a DI allowance—and to perform counterfactual analyses that allows us to infer the extent to which the welfare value of receiving a DI allowance is influenced by household labor supply responses, savings, and the possibility of reapplying for DI. The model-based results suggest that on average the welfare effect of DI benefits is positive and sizable, and particularly so for single individuals. Notably, because spousal labor supply responses provide partial insurance against the impact of DI denials on income and consumption of married households, the welfare value of DI benefits for married households is considerably smaller than for single individuals.

Our paper contributes to an active literature analyzing the economic consequences of public disability insurance systems (for a review, see Autor & Duggan, 2006; Autor, 2011; Liebman, 2015). While the core of this literature focuses on the impacts of disability benefits on the employment and earnings effects of DI allowance, little is known about either the fiscal costs or the household level effects on labor supply and consumption.[4] Meyer & Mok (2013) and Kostol & Mogstad (2015) offer to our knowledge the only prior study that documents changes in income and consumption that follow changes in health and disability. Our identification strategy, which uses judge assignments to isolate quasi-experimental variation in disability allowances, builds on three recent studies using U.S. data to estimate labor supply impacts of DI receipt.[5] Exploiting variation in DI allowances stemming from differences in disability examiner leniency, Maestas *et al.* (2013) and Autor *et al.* (2017) find that DI receipt substantially reduces earnings and employment of applicants. French & Song (2013) pursue a similar strategy—using variation in the leniency of appeal judges rather than initial examiners—and find comparable labor supply effects of DI receipt among appellants.

Our study makes two contributions to this active literature. It combines quasi-experimental variation in judicial disability determinations with extensive register data on disability applicants and household members to provide novel evidence on the income gains, consumption benefits and fiscal costs of DI receipt. Second, the subsequent structural model estimation offers a welfare assessment of these findings. Our structural model mirrors the life-cycle model used by Low & Pistaferri (2015) to analyze the insurance value and incentive costs of DI benefits. We deviate from Low and Pistaferri in two important ways. While Low and Pisteferri model individual behavior, and hence do not consider insurance from spousal labor supply, we model household behavior, which is important given our finding of a strong spousal labor supply response. Specifically, we estimate a life-cycle model with two earners making consumption and labor supply decisions. Distinct from Low and Pistaferri, we do not model the pre-application behavior of households, largely because we do not have health information for people who do not apply for DI. Our goal is therefore limited to understanding the post-application labor supply, savings, and reapplication decisions of applicants and their spouses, taking as given their characteristics and economic circumstances at the time of application. Our counterfactual estimates do not therefore take into account potential changes in

[4]This literature includes Parsons (1980), Bound (1989), Gruber (2000), Chen & van der Klaauw (2008), and Kostol & Mogstad (2014) as well as the methodologically related papers on DI discussed immediately below. See also Autor & Duggan (2003) and Borghans *et al.* (2014) for empirical evidence on the interaction between disability insurance and other transfer programs in the U.S. and Netherlands.

[5]See also Dahl *et al.* (2014) who use judge assignment to show that the receipt of a DI in one generation causes increased DI participation in the next generation.

the number and composition of applicants.

Our paper also advances understanding of how households respond to shocks to income.[6] Most work in this literature assumes exogenous labor supply, focuses on a single earner, or imposes restrictions on the nature and type of insurance available to families. A notable exception is Blundell *et al.* (2016b), who estimate a life cycle model with two earners jointly making consumption and labor supply decisions.[7] Consistent with our findings, Blundell et al. find an important role for consumption insurance through household labor supply, while self-insurance through savings and borrowing matter less. In line with these results, Persson (2015) finds that husbands increase their labor supply to offset household income losses following the elimination of survivors insurance for their wives, and Fadlon & Nielsen (2015) find that wives offset income losses following the death of a spouse through increased labor supply.

A related literature tests for the added worker effect, that is, an increase in spousal labor supply induced by negative income shocks to the other spouse (Lundberg, 1985). Cullen & Gruber (2000) review this literature and highlight the difficulty in drawing credible inferences from observational data. The key challenge is to locate a plausibly exogenous shock to the income of one spouse exclusively that does not directly affect the labor supply of the other spouse, thus overcoming the problem of simultaneity and correlated unobservables among spouses. Our research design resolves these challenges by identifying a plausibly exogenous income shock (DI allowance) that directly affects only one member of the household (the DI applicant), thereby providing a strong test of the added-worker effect.

The remainder of the paper proceeds as follows: Section 2 reviews the key features regarding the DI program in Norway, compares the system with the U.S. system, and describes the research design. Section 3 describes the data and sample restrictions. Section 4 assesses the relevance and validity of our instrument. Section 5 estimates the causal effect of DI allowance versus denial on applicant labor earnings and receipt of transfer income. Section 6 analyzes the household impact and fiscal costs of DI allowances. Section 7 documents that DI allowances affect household income and consumption differentially according to marital status, and explores how spousal responses to the allowance decision may help explain this heterogeneity. Section 8 develops and estimates a structural model of household labor supply and uses these estimates to explore the welfare value of disability receipt for marginal applicants. The final section concludes.

2 Background

We first provide an institutional and statistical description of the Norwegian DI program. We next document how the DI system generates quasi-random disability allowances for a subset of DI

[6]This literature is reviewed by Blundell *et al.* (2008), Meghir & Pistaferri (2011) and Blundell *et al.* (2016b).

[7]A complementary exception is Finkelstein *et al.* (2015), who directly estimate the insurance value of Medicaid in-kind public health plan benefits using variation from a randomized controlled trial. Distinct from our focus, their work (a) abstracts from labor supply considerations since labor supply appears unaffected by Medicaid provision in their setting (Baicker *et al.* , 2014); and (b) estimates both the transfer and ex ante insurance values of public benefits provision, whereas we estimate only the first component.

appellants (i.e., applicants who appeal their initial denial) and explain how our research design uses this variation to estimate the economic consequences of DI allowances.

The Norwegian DI program

The Norwegian DI program is designed to provide partial earnings replacements to all workers under the full retirement age who are unable to engage in substantial gainful activity because of a medically determined physical or mental impairment that has lasted for at least a year.[8] The DI program is part of the broader Social Security System and is financed through employer- and employee-paid taxes. The level of DI benefits received is determined using a formula based on an individual's earnings history. The benefits schedule is progressive, so that low-wage workers replace a relatively larger fraction of their earnings with DI benefits. DI payments consist of two components: a basic benefit amount, independent of the applicant's earnings history; and supplementary benefits that increase in pre-disability earnings levels. By law, singles have a higher basic benefit amount than married beneficiaries, and spousal income (if present) reduces the spousal benefit further.

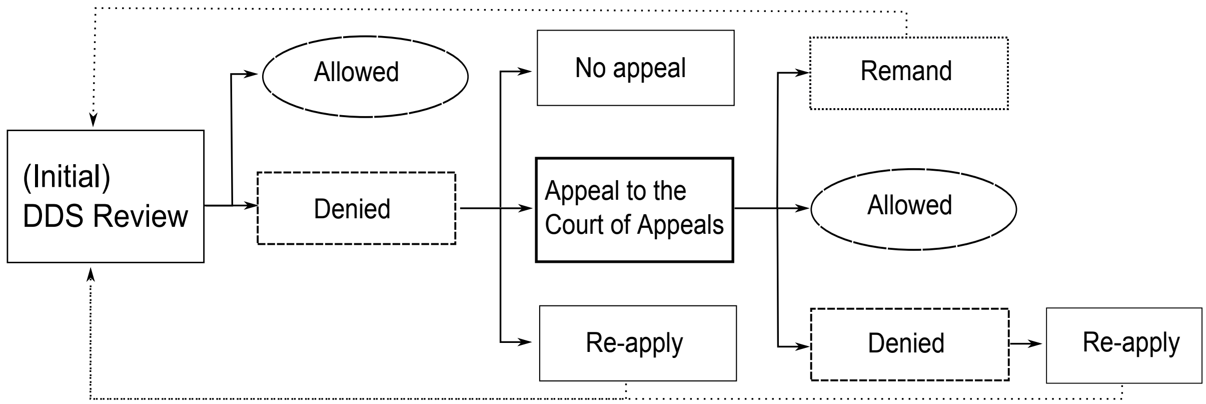
The disability determination process involves multiple steps, diagrammed in Figure 1. The first is the submission of an initial application to the Social Security Administration office for the Disability Determination Stage (DDS) review. If the applicant meets the non-medical criteria, such as age and prior employment requirements, disability examiners and medical staff assess written medical evidence regarding the applicant's ability to perform work-related activities, accounting for the applicant's health, age, education, work experience, and the transferability of her skills. If the disability examiner concludes that the applicant cannot be expected to engage in any substantial gainful activity, a disability award is made. Approximately 75 percent of applicants are awarded benefits at the DDS review. About 30 percent of beneficiaries receive partial awards. Cases that are more difficult to assess—typically claims of mental illness and lower back pain—are often denied at this step.

Those denied at the DDS review may appeal within two months to the Court of Appeals, and about 25 percent of denied applicants do so. Appellants are assigned to Administrative Law Judges (ALJs), who either allow, deny, or remand (i.e., return to the DDS for reevaluation) their cases.[9] ALJs are required to apply the same criteria used in the initial determination process, but applicants may present new supporting information in writing. Approximately 15 percent of all appealed claims are allowed at the ALJ level. If the appeal is denied, the applicant can choose to start a new DI case by reapplying to the DDS Review stage.

[8] This definition is almost identical to the one used by the U.S. SSDI program (see Social Security Act §1614).

[9] Average processing time at the DDS stage is six months, and average processing time at the appeal stage is four months. Remands are uncommon, accounting for only five percent of appeal outcomes. In our baseline analysis, we code remanded cases as rejected. As a robustness check, we recoded remanded cases as allowed or denied based on their eventual outcome after they were reconsidered by the DDS case worker. Results are comparable.

Figure 1: **DI Application and Appeals Process**



Notes: This figure summarizes the description of the application and appeal process in the Norwegian DI system.

Assignment of DI cases to judges

All Norwegian disability appeals are heard in Oslo. Prior to 1997, there was only one hearing department; subsequently, there were four equally sized departments, all housed in the same building, and with no specialization across the four departments. Within each department, the assignment of cases to Administrative Law Judge is performed by a department head who does not have knowledge of the content of cases. As stipulated in the rules set forth for the Administrative Law Court, case assignment should be done "by the drawing of lots." In practice, cases are assigned on a rotating basis depending on the date they are received and the alphabetical ordering of a judge's last name.[10]

Unusual among national disability systems, Norwegian judges are not specialized according to cases characteristic (medical condition, geographic region, or other attributes), and there is never any personal contact between judges and appellants (all evidence is presented in writing). Appellants have no option to seek a different judge from the one to whom they are assigned.

Verifying random assignment

Table 1 verifies that the hearing office assignment mechanism generates a distribution of cases across judges that is consistent with random assignment. There are 75 judges in our sample who have handled 375 cases each, on average. We measure judge leniency as the average allowance rate in all other cases a judge has handled (including the judge's past and future cases that may fall entirely outside of our estimation sample). To purge any differences over time or across departments in the characteristics of appellants or the overall leniency rate of the DI system, we always control for fully interacted year by department dummies (the level at which randomization occurs.)

[10]We verified these rules with the current Head of the Administrative Law Court, Knut Brofoss. The rules are explained in "Veileder for Saksbehandlingen i Trygderetten" (Guidelines for Processing Cases in the Court of Appeals). We have also presented our work at internal seminars with the current set of judges and department heads to confirm that we have understood how the cases are handled and assessed.

Table 1: **Testing for Random Assignment of Cases to Judges**

	(1)	(2)	(3)	(4)
Dependent Variable:	**Case Allowed**		**Judge Leniency**	
	coef.	s.e.	coef.	s.e.
A. Pre-determined characteristics				
Age	0.0044***	(0.0003)	0.0001	(0.0001)
Number of persons in household	-0.0143***	(0.0021)	-0.0003	(0.0003)
Female	0.0193***	(0.0056)	0.0008	(0.0012)
Married	0.0146**	(0.0066)	0.0005	(0.0012)
Foreign born	-0.0446***	(0.0086)	-0.0003	(0.0015)
Less than high school degree	-0.0231***	(0.0061)	-0.0005	(0.0008)
High school degree	0.0195***	(0.0061)	0.0001	(0.0007)
Any college	0.0119	(0.0116)	0.0010	(0.0014)
Children below age 18	-0.0601***	(0.0058)	-0.0009	(0.0010)
Musculoskeletal disorders	-0.0171***	(0.0059)	0.0005	(0.0017)
Mental disorders	0.0088	(0.0075)	-0.0003	(0.0024)
Circulatory system	0.0235	(0.0158)	0.0000	(0.0023)
Respiratory system	-0.0196	(0.0151)	-0.0021	(0.0021)
Neurological system	0.0459**	(0.0206)	0.0011	(0.0021)
Endocrine diseases	0.0418***	(0.0174)	-0.0029	(0.0031)
B. Pre-determined economic variables				
Average indexed earnings ($1,000)	0.0009***	(0.0002)	0.0000	(0.0000)
Total transfers ($1,000)	-0.0004	(0.0003)	0.0001	(0.0001)
Liquid assets ($1,000, per capita)	0.0004**	(0.0002)	0.0000	(0.0001)
Total gross wealth ($1,000, per capita)	0.0001***	(0.0000)	0.0000	(0.0000)
Total liabilities ($1,000, per capita)	0.0001	(0.0001)	0.0000	(0.0000)
Disposable income ($1,000, per capita)	0.0006*	(0.0004)	0.0000	(0.0002)
F-statistic for joint significance	24.36		0.78	
[p-value]	[.001]		[.72]	
Observations	14,092		14,092	

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Notes: This table reports an F-test of whether the hearing office complied with the random allocation procedure described in Section 2. The baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. Columns report OLS regressions of appellant characteristics on (column 1) a dummy variable for whether the case was allowed; and (column 3) our measure of judge leniency. F-statistics are obtained from OLS estimation on the combined set of appellant characteristics. Each regression controls for fully interacted year of appeal and department dummies. Characteristics of appellants are measured prior to appeal. Variable definitions are as follows: children is equal to 1 if appellant has children under age 18 and 0 otherwise; any college is equal to one if a person has some college or has a college degree; body system codes are based on ICD-10 diagnostic codes. Pre-determined economic variables are measured one year before appeal, and average indexed earnings is mean earnings for the ten years prior to appeal. Assets, wealth, liability and disposable income are measured at the household level and normalized by the number of household members. Nominal values are deflated to 2005 and represented in US dollars using the average exchange rate NOK/$ = 6.

The first column of Table 1 uses a linear probability model to test whether appellants' (pre-determined) characteristics and economic conditions are predictive of case outcomes. As expected, demographic, economic and health variables are highly predictive of whether an appealed case is allowed. Column 3 assesses whether these same case characteristics are predictive of the leniency of the judges to which cases are assigned. We find no such relationship. Jointly, these 21 variables explain about 0.1 percent of the variation in the judge leniency measure (joint p-value of 0.72), and none is statistically significant at the 10 percent level.

A natural question is why some judges are more lenient than others. We have few detailed characteristics of judges to help illuminate this question, but we do know the number of cases that each judge has handled. We find that experienced judges appear to be slightly less lenient, but experience accounts for only a small fraction of the total variation in allowance rates across judges (see Appendix Figure A.1). Analyzing the underlying sources of the inter-judge differences in leniency is outside the scope (and reach) of our paper. What is critical for our analysis is that appellants are randomly assigned to judges (as our data confirm), that some judges are systematically more lenient than others (as documented in Section 4.1), and that cases allowed by a strict judge would also be allowed by a lenient one (consistent with the tests in Section 4.2).

Instrument and 2SLS model

We use variation in DI allowance generated from the random assignment of appeal judges as an instrumental variable to estimate the economic consequences of disability receipt. Because some judges are systematically more lenient than others, as we document below, random assignment of appellants to judges rise to exogenous variation in the probability an individual is allowed DI in the appeals process.

Our baseline instrumental variables (IV) model can be described by the following two-equation system:

$$A_i = \gamma Z_{j(i)} + X_i'\delta + \varepsilon_i \tag{1}$$

$$Y_{it} = \beta_t A_i + X_i'\theta_t + \eta_{it}. \tag{2}$$

Here, A_i is an indicator variable equal to 1 if appellant i is allowed DI at the appeal, and $Z_{j(i)}$ is the leniency measure for judge j to which appellant i is assigned. The vector X_i contains relevant control variables, including a full set of year-of-appeal by department dummies. In the second stage equation, Y_{it} is a dependent variable of interest that is measured for appellant i at some point t after the allowance decision (e.g. earnings three years after the decision).

The target of our estimation is the average of β_t among individuals who are allowed DI at the appeal because they were assigned to a lenient judge. To estimate this Local Average Treatment Effect, our baseline specification uses 2SLS with first and second stage equations given by (1) and (2). The endogenous variable in our estimation is an indicator for whether an appellant was allowed DI *at appeal*, rather than whether or not the appellant is currently receiving DI when outcome Y_{it} is observed. This specification alleviates concerns about the exclusion restriction: 2SLS estimates of β_t capture the causal effects of the *initial* judicial disability determination, which may operate through a number of channels, including participation in DI, subsequent reapplications to the DI program following denial, or other behavioral changes resulting from the initial outcome at appeal. We can also estimate the reduced form effect of judge leniency on appellant outcomes by directly regressing Y on Z and X.

3 Data and Background

3.1 Data and Sample Restrictions

Our analysis draws on multiple administrative data sources that are linked by unique person-level identifiers. Information on DI benefits comes from social security registers that contain complete records for all individuals who entered the DI program during the period 1967-2010. These data include information on each individual's work history and medical diagnosis, the month when DI was awarded or denied, and the level of DI benefits received. These person-level records are linked to hearing office records on all DI appeals during 1989 through 2011, including dates of appeal and decision, outcomes for each appeal, and unique identifiers for both judges and appellants.

To capture complete information on DI applicants' earnings, income and assets, we merge the social security data with longitudinal administrative registers provided by Statistics Norway covering every Norwegian resident from 1967 to 2010. These register data enumerate individual demographic information (including sex, age, and education) and, since 1993, all sources of annual income, including earnings, self-employment income, capital income, and cash transfers, as well as most types of assets holdings and liabilities, such as as real estate, financial portfolio, and debt. Income data are reported in annual amounts, while the values of assets holdings and liabilities are measured as of the last day of each year.

The Norwegian data have several advantages over register data collected by many other countries. Because most components of income and wealth are third-party reported (e.g. by employers, banks and financial intermediaries), the coverage and reliability are rated as exceptional by international quality assessments (see e.g Atkinson *et al.* 1995). Because in Norway, most register data are a matter of public record, there is no attrition from the original sample due to non-response or non-consent. The income and wealth data pertain to all Norwegian residents, and are therefore not limited to those employed in jobs covered by social security, individuals who respond to wealth surveys, or households that file estate tax returns. Measures of income and wealth are recorded without any top or bottom coding.[11] Finally, unique identifiers allow us to match spouses to one another and parents to children, thereby constructing measures of per capita household income and consumption.

A key challenge in estimating the consumption effects of DI receipt is the lack of reliable longitudinal data on consumption expenditures. One approach to measuring expenditures is to use survey data, but expenditure surveys typical have small sample sizes and face significant measurement issues (see Pistaferri, 2015 for a discussion). A second option is to create measures of consumption from the accounting identity that total consumption expenditure is equal to income plus capital gains minus the change in wealth over the period. Browning & Leth-Petersen (2003) shows how

[11] Some individuals are reported with negative income components (e.g. negative cash transfers). In these cases, we truncate the income components at zero. We also top-code a handful of observations with extremely large income components. The results do not change appreciably if we retain these outliers.

one can construct such measures of consumption from longitudinal data on income and assets. Eika *et al.* (2017) perform a similar exercise combining tax data on income and wealth with detailed information on households' financial and real estate transactions. Their analysis shows that the measures of consumption derived from such data sets conform well to those reported in family expenditure surveys and to the aggregates from national accounts. We use their measures here, and refer the reader to Eika *et al.* (2017) for more details.

Our empirical analysis considers individuals who appeal an initially denied DI claim.[12] To observe individuals for at least four years after the appeal decision, our estimation sample consists of individuals whose appeal decision was made during the period 1994-2005. To reduce sampling variation in the instrumental variable, we follow Maestas *et al.* (2013) and French & Song (2013) in excluding observations for which the assigned appeal judge has handled fewer than 10 cases during the 1989 through 2011 period.[13] To circumvent the issue of older appellants substituting between DI and early retirement, we also exclude appellants who are above age 62 at the time of appeal.

In Table 2, we document characteristics of the sample of individuals who apply for DI and the subsample who appeal an initially denied DI claim (our baseline sample). Relative to the full sample of initial applicants, those who appeal are more likely to be female, are less educated, are more likely to be foreign born, and have lower prior earnings and assets. Seventy percent of DI appellants claim mental or musculoskeletal disorders, whereas this figure is 63 percent for the full set of DI applicants.

3.2 Institutional Background

There are a number of similarities and some key differences between the DI systems in the U.S. and in Norway (see Autor & Duggan, 2006; Kostol & Mogstad, 2014). In both countries, DI is one of the largest transfer programs. However, the prevalence of receipt of DI benefits is lower in the U.S. than in Norway, as shown in Figure 2, while the time trends are similar.[14] From 1961 to 2012, DI prevalence increased from 2.2 to 9.7 percent in Norway and from 0.8 to 5.0 percent in the U.S. Norway's prevalence has leveled off at about 10 percent in recent years. The U.S. SSDI prevalence rate rose steeply through 2013, after which time growth peaked and reversed (Social Security Advisory Board, 2015).[15]

[12] Some individuals have several denied DI claims over the period we consider. In such cases, we restrict our sample to the individual's first denied DI claim.

[13] Including these judges does not change the estimates appreciably, and neither does excluding judges who handle fewer than 50 cases.

[14] The cross-country difference in DI coverage is unlikely to explain the entire discrepancy in the incidence of DI: although virtually all non-elderly adults are covered in Norway, more than 80 percent of all non-elderly adults are covered in the U.S. The remaining difference could be a function of underlying differences in screening stringency, the generosity of the programs, or the frequency with which people apply for disability benefits. Milligan & Wise (2011) argue that differences in health are unlikely to explain much of the observed differences in DI rates across developed countries.

[15] The U.S. Supplemental Security Income program (SSI) also provides disability benefits to adults and children with work-limiting disabilities. DI and SSI therefore jointly provide disability benefits to a larger share of U.S. adults than does DI alone. However, the U.S. DI program is more comparable to the Norwegian DI program than is the U.S. SSI program since SSI primarily provides benefits to adults with little work history. In this sense, SSI is more akin to

Table 2: **Descriptive Statistics of Applicants and Appellants**

	DI applicants		**DI appellants**		**Test of equal means**
	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
A. Pre-determined characteristics					
Age (at the time of decision)	48.55	[9.98]	46.61	[9.30]	-25.17
Number of persons in household	2.37	[1.17]	2.79	[1.30]	39.28
Female	0.56	[0.50]	0.63	[0.48]	17.5
Married	0.57	[0.50]	0.57	[0.49]	0.73
Foreign born	0.08	[0.27]	0.18	[0.38]	32.81
Less than high school degree	0.43	[0.50]	0.50	[0.50]	16.97
High school degree	0.42	[0.49]	0.39	[0.49]	-8.17
Any college	0.13	[0.34]	0.11	[0.31]	-7.64
Children below age 18	0.3	[0.46]	0.58	[0.49]	66.48
Musculoskeletal disorders	0.37	[0.48]	0.44	[0.50]	17.67
Mental disorders	0.26	[0.44]	0.26	[0.44]	1.42
Circulatory system	0.08	[0.27]	0.04	[0.19]	-27.59
Respiratory system	0.03	[0.17]	0.03	[0.16]	-4.12
Neurological system	0.06	[0.23]	0.04	[0.19]	-12.3
Endocrine diseases	0.02	[0.14]	0.04	[0.20]	14.05
B. Pre-determined economic variables					
Average indexed earnings ($1,000)	32.76	[23.66]	25.81	[21.25]	-39.3
Total transfers ($1,000)	14.81	[14.90]	15.78	[14.06]	8.21
Liquid assets ($1,000, per capita)	23.85	[43.85]	9.63	[21.29]	-72.06
Total gross wealth ($1,000, per capita)	173.13	[212.10]	91.81	[105.93]	-83.76
Total liabilities ($1,000, per capita)	54.72	[67.25]	38.43	[49.21]	-37.97
Disposable income ($1,000, per capita)	26.54	[14.88]	24.08	[13.11]	-22.14
DI allowed	0.79	[0.41]	0.13	[0.33]	
Observations	240,900		14,092		

Standard deviations [in square brackets]

Notes: This table reports descriptive statistics for applicants and appellants. The applicant sample consists of all claims made during the period 1992-2003 by individuals who are at most 61 years of age. The appellant sample consists of the subset of applicants who filed an appeal during the period 1994-2005 (see Section 3 for further details). All characteristics are measured the year before application/appeal unless otherwise stated. The final column reports t-statistics of the test of equality between characteristics of applicants and appellants. Variable definitions are as in Table 1.

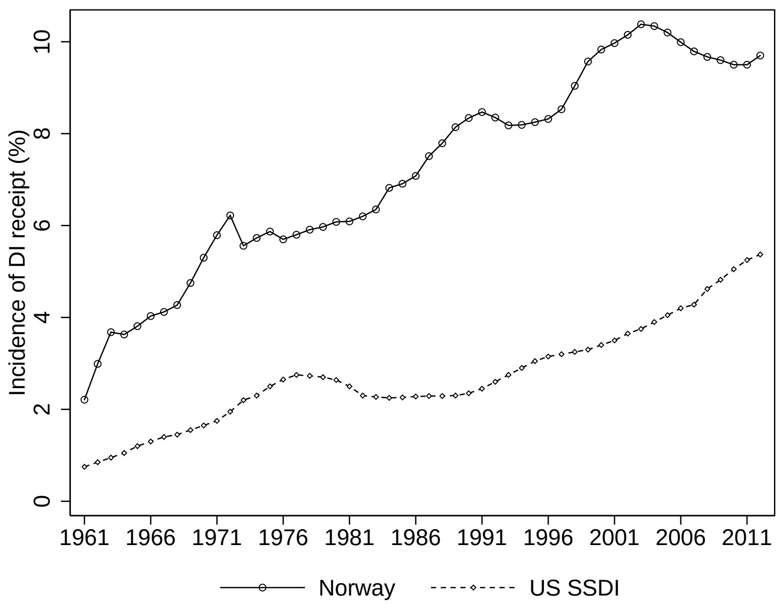
In both countries, the expansion of the DI rolls in recent decades appears to be driven in significant part by changes in disability screening criteria, which led to a steep rise in the share of DI recipients suffering from difficult-to-verify disorders such as mental illness and musculoskeletal disease.[16] Because mental illness and musculoskeletal disease have low associated mortality rates—and

the social assistance program in Norway, which is a need-based and means-tested program, with the difference that SSI applies only to individuals with disabilities.

[16]See Autor & Duggan (2006) and Liebman (2015) for discussions of this phenomenon. In the U.S., the 1984 congressional reforms shifted the focus of screening from medical to functional criteria. In Norway, the medical eligibility criteria were relaxed earlier and more gradually.

moreover, because mental illness typically has an early onset—DI recipients with such diagnoses tend to participate in the program for relatively long periods. DI exit rates in both countries have decreased in the last few decades, with progressively fewer DI recipients either reaching retirement age or dying in a given year (see Appendix Figures A.2 and A.3). The aging of the Baby Boom cohorts into their peak (near-elderly) disability age brackets has contributed substantially to the expansion of the U.S. DI rolls since the mid-1990s (Liebman, 2015).

Figure 2: **Trends in DI Receipt in Norway and the U.S.**



Notes: This figure displays trends in DI receipt in Norway and the U.S. (see Section 2).. U.S. trends are based on Autor & Duggan (2006) for 1957-2005 and SSA Office of the Chief Actuary for 2006-2012. Norwegian trends are based on SSA Statistical Supplements. Incidence of DI receipt defined as the percent of the relevant adult population receiving DI benefits (age 18-67 in Norway; age 25-64 in the US).

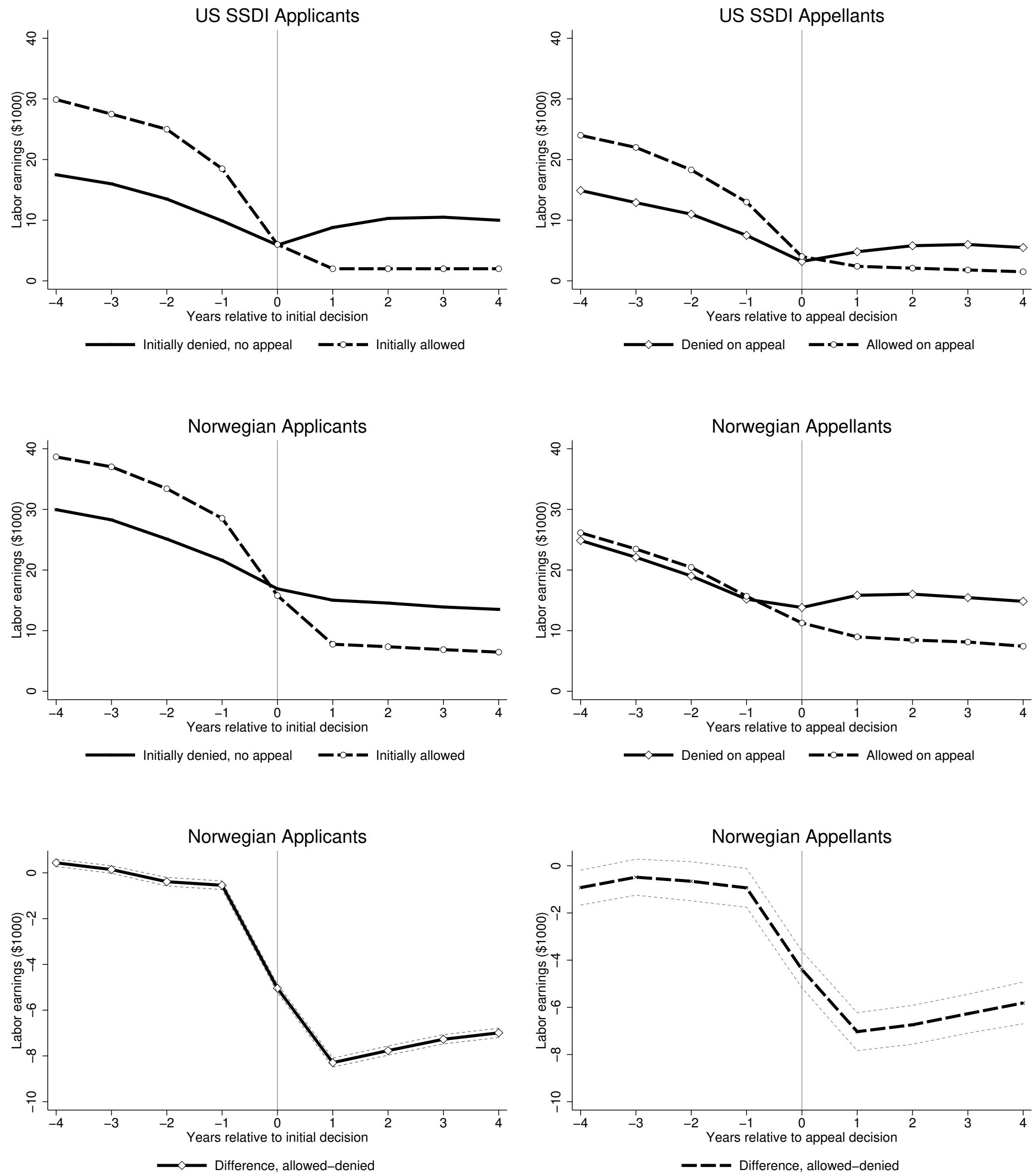
There are noteworthy differences between the U.S. and Norwegian DI programs. One difference is their income replacement rates. Kostol & Mogstad (2014) compute the replacement rate for a typical Norwegian applicant in according to the SSDI rules and the Norwegian program. For the worker they consider, the pre-tax income replacement rate would be 31 percent in the U.S. program and 58 percent in the Norwegian program. These calculations disregard income taxation, dependent benefits, and health insurance, however. Both countries' DI programs provide dependent benefits. In addition, DI recipients in the U.S. receive health insurance coverage through the federal Medicare program, which is a substantial in-kind benefit.[17] In Norway, by contrast, all citizens are eligible for health insurance through the Social Insurance System. Another difference is that the appeal process plays a more important role in the U.S. than in Norway. While 48 percent of initially rejected applicants appeal in the U.S. (French & Song, 2013), only 25 percent of those rejected in Norway appeal. Success rates at appeal are also considerably higher in the U.S. than Norway.

[17] Autor & Duggan (2006) estimate that Medicare benefits account for approximately 40 percent of the present value of an SSDI award.

Despite these differences in prevalence, benefits structure, and appeals behaviors, there are important similarities between the applicant, appellant and participant populations across the two countries. Almost 60 percent of DI recipients in both countries suffer from difficult-to-verify mental and musculoskeletal disorders (see Appendix Table A.1). And in both countries, appellants are younger, have lower prior earnings, and are more likely to suffer from difficult-to-verify disorders than are average DI applicants (see Appendix Table A.2). As a further comparison among the two programs, Figure 3 uses Norwegian and U.S. data (the latter from Maestas *et al.* 2013) to plot earnings trajectories of DI applicants and appellants in Norway and the U.S., before and after their DI determinations. We focus on years $t-4$ through years $t+4$ surrounding the initial DI determination (lefthand panel) and the year of the initial appeal decision (righthand panel).

The patterns are quite similar across the two countries. DI applicants who are allowed at the initial determination have on average substantially higher prior earnings than those who are denied. This likely reflects the fact that workers with high prior earnings who seek DI benefits often face severe impairments that necessitate a sudden cessation of employment; conversely, applicants with low prior earnings may in part be compelled to seek DI due to a lack of employment opportunities rather than by severe health shocks per se. Similarly, earnings diverge immediately between allowed and denied appellants following the appeal decision in both countries, and this gap is not closed over the subsequent four post-decision years. The figures in the bottom panel plot the difference between denied and allowed applicants (left) and appellant labor earnings (right) over time, controlling flexibly for observable characteristics and lagged dependent variables (up to the year of the initial decision, after which they are fixed as the mean over the years prior to decision).

Figure 3: **Earnings Trajectories of Allowed and Denied DI Applicants and Appellants**



Notes: The top four figures display changes in the levels of earnings for allowed (dashed line) and denied (solid line) DI applicants (left) and for DI appellants (right) in the nine years surrounding the initial DI determination and the initial outcome at appeal in the U.S. (top panel, sourced from Maestas *et al.* 2013), and for Norway (middle panel). For the Norwegian data, the applicant sample consists of all claims made during the period 1998-2003 by individuals who are at most 61 years of age. The appellant sample filed an appeal during the period 1998-2005 (see Section 3 for further details). The figures in the bottom panel plot the difference between denied and allowed applicants (left) and appellant labor earnings (right) over the same period, controlling flexibly for observable characteristics and lagged dependent variables (up to the year of the initial decision, after which they are fixed as the mean over the years prior to decision). The dashed lines in the bottom panel represent 90% confidence intervals, where each yearly difference is estimated separately with flexible controls for individual characteristics comprising application year dummies, dummy variables for county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, educational attainment, and number of medical diagnoses, as well as polynomials of lagged averages of earnings and disposable income (not including observations after the decision). Nominal values are deflated to 2005 and represented in US dollars using the average exchange rate NOK/$ = 6.

4 Assessing the Instrument

We begin our presentation of results by providing evidence on the relevance and validity of the instrument.

4.1 Instrument Relevance

Figure 4 provides a graphical representation of the first stage of our IV model. In the background of this figure is a histogram for the density of judge leniency (controlling for fully interacted year and department dummies). The measure of judge leniency is the average judge allowance rate in all other cases a judge has handled, including the judge's past and future cases that may fall outside of our estimation sample. The mean of the leniency variable is 0.15 with a standard deviation of 0.05. The histogram reveals a wide spread in judge leniency, with a judge at the 90th percentile allowing approximately 18 percent of cases as compared to approximately 8 percent for a judge at the 10th percentile.

The solid line plotted in the figure's foreground depicts the relationship between judge leniency and the appellant's allowance rate (controlling for fully interacted year and department dummies). The graph is a flexible analog to the first stage equation (1), where we plot a local linear regression of individual allowance outcomes against judge leniency. The individual allowance rate is monotonically increasing in our leniency measure, and is close to linear. A 10 percentage point increase in the judge's allowance rate in other cases is associated with an approximately 8 percentage point increase in the probability that an individual appellant's case is allowed.

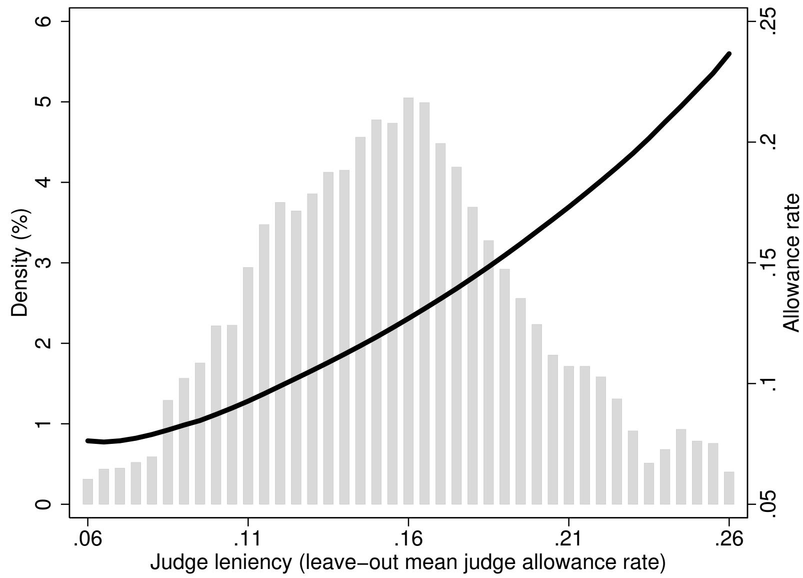
Table 3 presents estimates of our first equation for the relationship between judge leniency and DI allowance rates at appeal (1). In each column, we regress a dummy variable for whether an individual is allowed DI at appeal on the judge leniency measure. We include fully interacted year and department dummies in panel A but otherwise include no other controls. The four columns of the table correspond to years one through four following appeal. These columns are identical except for the very modest impact of sample attrition (less than three percent over four years) stemming from death or emigration of appellants.[18] The point estimate of approximately 0.82 is essentially identical across columns, indicating that attrition exerts a negligible impact on the first stage relationship. All else equal, assignment to a judge with a 10 percentage point higher overall allowance rate increases the probability of receiving an allowance by 8.2 percentage points.

4.2 Instrument Validity

In order for judge leniency to be a valid instrument, appellants' assignment to judges must be uncorrelated with case characteristics. Table 1 above provides strong empirical support for the

[18]Column 1 of Appendix Table A.5 documents that the assignment variable (judge leniency) does not affect the probability that an appellant either dies or emigrates during the outcome period.

Figure 4: **Effect of Judge Leniency on DI Allowance**



Notes: This figure displays the effect of judge leniency on DI allowance, conditional on fully interacted year and department dummies. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. The solid line plots a local linear regression of allowances on judge leniency. The histogram of judge leniency is shown in the background of both figures (top and bottom 1 percent excluded from the graph).

claim that the DI system in Norway randomly assigns appeal judges within each department and year. Panel B of Table 3 provides a second confirmation of this fact: adding controls for appellant characteristics to the regression model has almost no effect on the point estimates, consistent with random assignment of appellants to judges.

This random assignment mechanism is sufficient for consistent estimation of the reduced form effect of judge leniency on appellant outcomes. However, to interpret the IV estimates of equations (1) and (2) as the causal effects of *DI allowances* on appellant outcomes requires two further assumptions. The first is that judge leniency affects appellant outcomes of interest only through its impact on the appellant's allowance decision. This exclusion restriction appears particularly likely to hold in Norway, where all appeals are presented in writing, individuals (and their families) observe only judges' allowance or denial decisions, and there is no personal contact between judges and appellants. One potential violation of the exclusion restriction could occur if appeals processing times differed systematically with judge leniency and, moreover, exerted an independent effect on appellant outcomes (as explored in Autor *et al.* 2017). To test this possibility, we calculated each judge's average processing time based on the residual average processing time in his or her other cases. Panel C of Table 3 shows that the first stage estimates do not change appreciably when controlling for judge processing time.

The second assumption needed for a causal interpretation of the IV estimates is the monotonicity of the judge leniency instrument.[19] Monotonicity requires that, for each appellant, the probability

[19]If the treatment effect of the disability determination were constant among appellants, the monotonicity assump-

Table 3: **First Stage: Judge Leniency and DI Allowance**

	Years after decision			
	1	**2**	**3**	**4**
Panel A.	**No covariates**			
Judge leniency	0.818***	0.819***	0.821***	0.828***
	(0.082)	(0.084)	(0.083)	(0.083)
Panel B.	**With individual covariates**			
Judge leniency	0.793***	0.792***	0.794***	0.800***
	(0.078)	(0.081)	(0.080)	(0.080)
Panel C.	**With judge characteristics**			
Judge leniency	0.808***	0.811***	0.815***	0.822***
	(0.074)	(0.075)	(0.074)	(0.075)
Dependent mean	0.13	0.13	0.13	0.13
Observations	13,972	13,842	13,709	13,607

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.
Note: This table reports the first stage coefficients of equation 1. The baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. In panel A, DI allowance is regressed on judge leniency and fully interacted year of appeal and department dummies. Panel B includes flexible controls for individual characteristics: fully interacted year and department dummies, and dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and number of medical diagnoses. Panel C also controls for judge leave-out-mean processing time.

of being allowed at appeal would be at least as high if assigned to a strict judge (low value of Z) as if assigned to a lenient judge (high value of Z). Since no individual can be assigned to two different judges at the same point in time, it is impossible to verify this assumption. There are, however, some testable implications which would allow us to reject the assumptions. The first testable implication we consider is that the first stage estimate should be non-negative for any subpopulation. If this were not the case, we would infer that the judges whom we estimate to be more lenient on average are stricter towards a subset of cases. Reassuringly, when separately estimating the first stages based on the (pre-determined) observable characteristics of the individual, we find that the estimates are consistently positive and sizable, consistent with the monotonicity assumption (see Appendix Table A.3).

As a second check on this threat to validity, we directly examine whether judges who are stricter towards one subset of appellants (e.g., young appellants, those with mental disorders) are also relatively strict towards the complementary group of appellants (e.g., older appellants, those without mental disorders). We perform this test by again partitioning the data into the subpopulations that were used in the prior test, but in this case, we recalculate the leniency instrument for each subpopulation to be the judge's leniency for cases outside of the subpopulation. For example, when assessing the effect of judge leniency on allowances for male appellants, we calculate judge

tion would be unnecessary. But we do not find the constant treatment effect assumption plausible in this setting.

leniency using only decisions in cases with female appellants. Column (2) of Appendix Table A.3 reports these results. All estimates using this redefined instrument are positive and statistically significant, consistent with the maintained assumption that leniency is a judge-specific attribute that characterizes judges' decision-making across the panoply of cases that they are assigned.

5 Impacts of DI allowances on the Appellants

This section investigates the impacts of DI allowances on the labor earnings, DI benefits, and total transfers received of the appellants. These results lay the groundwork for the analysis in the next section of the household impacts and fiscal costs of DI allowances.

5.1 Effects on Labor Earnings and DI Benefits

In Panels A-C of Table 4, we report 2SLS estimates of equations (1) and (2) with DI participation, DI benefits payments, and labor earnings as dependent variables in the second stage. As in Table 3, we separately estimate the effects of the initial appeal decision on outcomes for each of the four subsequent years. All specifications control for observable case characteristics and include a full set of year by department dummies.

The first two panels consider the impact of being allowed at the appeal on DI participation and benefits payments. Column 1 of panel A reports a 2SLS point estimate of 0.989, indicating that allowances induced by judge leniency increase DI receipt almost one-for-one in the first year following appeal.[20] Over the the first four years following appeal, the causal effect of being allowed at the appeal on subsequent DI receipt falls by approximately half, from 0.99 to 0.47, reflecting the fact that a substantial fraction of appellants who are initially denied DI benefits reapply and are ultimately allowed.[21] Panel B displays analogous estimates for DI benefit payments. Receiving a DI allowance at appeal leads to a large increase in benefit payments relative to the alternative outcome, with this increment equal to $16,240 in the first year. This contrast declines over time due to successful DI reapplications, reaching $8,167 in year four. Over the initial four years following appeal, receiving a DI allowance increases DI benefit payments by approximately $11,900 per annum.

Panel C of Table 4 displays analogous estimates for annual labor earnings of DI appellants. DI allowances have sizable negative causal effects on labor earnings. Receiving a DI allowance on appeal reduces annual earnings by approximately $6,800 in the first year after appeal, equal to approximately forty percent of the annual DI transfer benefit received. Distinct from the steeply declining causal effect of an initial allowance on DI participation and DI transfers, the causal effect of an initial DI allowance on appellant earnings declines only modestly over the four years following the initial appeal outcome. Thus, over the first four years following appeal, receiving a DI allowance reduces labor earnings by around $5,200 per annum.

[20] Note that 0.989 = 0.784/0.793, where 0.793 is the corresponding first stage coefficient from Table 3, panel B column 1.

[21] Although this pattern could also be consistent with successful appellants exiting the DI program in years two through four, such exits rarely occur.

Table 4: **Effect of DI Allowance on Labor Earnings, DI Benefits, and Transfer Payments of the Appellant**

	Years after decision				
	1	**2**	**3**	**4**	**Average**
Panel A.			**DI participation**		
Allowed DI	0.989***	0.727***	0.646***	0.470***	0.707***
	(0.071)	(0.102)	(0.098)	(0.084)	(0.078)
Dependent mean	0.305	0.432	0.519	0.577	0.456
Panel B.			**DI benefits ($1,000)**		
Allowed DI	16.240***	12.596***	10.203***	8.167***	11.883***
	(1.539)	(1.696)	(1.660)	(1.567)	(1.316)
Dependent mean	5.708	8.377	10.277	11.502	8.921
Panel C.			**Earnings ($1,000)**		
Allowed DI	-6.791**	-5.946**	-5.577*	-5.660**	-5.178**
	(2.765)	(2.877)	(2.952)	(2.706)	(2.275)
Dependent mean	14.240	14.282	13.802	13.245	13.813
Panel D.			**Total transfers ($1,000)**		
Allowed DI	10.188***	8.807***	8.148***	6.429**	8.072***
	(2.736)	(2.749)	(2.433)	(2.683)	(2.499)
Dependent mean	19.567	20.072	20.54	21.053	20.305
Panel E.			**Non-DI transfers ($1,000)**		
Allowed DI	-6.308*	-3.744	-1.884	-1.611	-3.823*
	(3.273)	(2.656)	(2.062)	(2.525)	(2.298)
Dependent mean	14.009	11.839	10.398	9.666	11.521
Observations	13,972	13,842	13,709	13,607	13,972

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Note: This table reports instrumental variables estimates of the causal effect of receiving a DI allowance at the appeal stage on DI participation (panel A), annual DI benefits (panel B), and annual labor earnings (panel C), annual total transfers inclusive of DI benefits (panel D), and annual transfers excluding DI benefits (panel E). Columns 1-4 report separate estimates for each year, whereas column 5 reports estimates for the average outcome over the four year period. The baseline sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. All control variables are measured prior to appeal.

The estimates in Table 4 can be interpreted as local average treatment effects (LATE) for appellants whose DI decisions are affected by the instrument (i.e., the compliers), meaning they could have received a different allowance decision had their case been assigned to a different judge. As shown in Imbens & Rubin (1997), we can decompose these LATEs to draw inference about what compliers would have received in DI benefits and earned in labor income if denied or, alternatively, if allowed at appeal. These *potential outcomes* for compliers may be recovered by combining (i) the shares of never-takers and compliers to the instrument with (ii) the average *observed outcomes* of individuals who were not allowed with the most lenient or strictest judges (that is, those facing the highest and lowest values of the instrument).[22]

In Appendix Figure A.4, we implement these calculations to decompose the LATE into the potential outcome of appellant compliers if denied or, alternatively, if allowed. Relative to the regression estimates in Table 4, the figure plots *levels* of potential outcomes rather than simply depicting the LATE contrast between potential outcomes in the two states. Although many denied appellants reapply for, and eventually receive, DI benefits (Table 4), we find that labor earnings of compliers who are denied at appeal change little following denial. In contrast, labor earnings of compliers allowed at initial appeal fall steeply, particularly in the year of allowance and the year immediately thereafter. This pattern suggests that among the population of denied compliers, a small but non-negligible subset persists in employment following denial, while a larger group works minimally and pursues further appeals.

5.2 Benefit Substitution in Response to DI denial

As in many European countries, DI is one of several transfer programs available to Norwegians, and those whose DI claims are denied may potentially substitute towards these other programs. Conversely, DI beneficiaries may also seek other transfer benefits following the award of DI benefits. Key transfers programs other than DI benefits are social assistance (i.e., traditional welfare benefits), housing benefits, and vocational rehabilitation benefits.

Panels D and E of Table 4 report 2SLS estimates of the impact of an initial DI allowance on total transfers (DI benefits plus all other cash transfers) and cash transfers excluding DI benefits. These estimates point to the importance of accounting for benefit substitution when considering the impacts of disability allowances on household incomes and public expenditure: the net impact of a DI allowance on total transfers received is about 20 to 40 percent smaller than its gross impact, with the largest discrepancies in the first two years following the initial appeal decision. On average, the net impact of a DI allowance on total transfers is about \$8,100 per annum, approximately \$3,800 less than the estimated gross impact on DI benefits. Both of the average increase in total transfers and the average decline in non-DI transfers are significantly different from zero at the 10 percent significance level.

In Appendix Figure A.5, we decompose the LATE estimates for benefit receipt into potential outcomes for compliers when allowed and when denied. When compliers are awarded DI benefits, we see a sizable fall in their payments from non-DI transfer programs, indicating benefit substitution. Non-DI transfer payments change little in the year following appeal when compliers are denied DI, however. As many compliers who were denied at initial appeal successfully reapply for DI, their DI payments rise and non-DI transfers fall in the years after the initial denial. The fact that the net impact of a DI allowance on appellant transfer payments is smaller than its gross impact indicates that DI and non-DI transfer programs serve as substitutes. In Section 7, we explore whether spousal

[22] Imbens & Rubin (1997) show how to derive the potential outcomes of compliers with and without treatment in settings with a binary instrument. Dahl *et al.* (2014) extend this to settings with multi-valued or continuous instruments. We follow the procedure of Dahl *et al.* (2014).

labor supply provides an additional margin through which married appellants may buffer household income in the event of DI denial.

6 Household Impacts and Fiscal Costs of DI Allowances

In Table 5, we present estimates of the income and consumption gains that households obtain from DI allowances, and compare these gains with the fiscal costs that other taxpayers bear. This table reports 2SLS estimates of the impact of allowances versus denials at appeal on disposable income and consumption as well as fiscal costs inclusive of DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts. Panel A uses the full sample of appellants, while Panel B restricts the sample for whom we have detailed measures of household consumption expenditure.[23] To facilitate comparison across households of difference size, we divide the outcomes by the size of the appellant's household (so that both income gains and fiscal costs are per capita).

Despite both countervailing behavioral responses and countervailing transfer program interactions documented above, DI allowances nevertheless yield meaningful income gains among individuals and their families at the margin of program entry. Panel A.1 of Table 5 indicates that DI allowances raise income available for consumption per household member by approximately $3,200 per annum. This effect is statistically significant at approximately the five percent level when pooling outcomes over the four years following appeal. At the same time, we readily reject the null hypothesis that the causal effect of a DI allowance on income (per household member) in each of the four post-appeal years is as large as its effect on initial DI benefits payments. Thus, the net effect of DI allowances on household incomes are far smaller than their gross effect due to the influence of household labor supply, reapplication behavior, and benefit substitution.[24]

Given that DI allowances significantly increase disposable income (per household member) among appellants while reducing household labor supply (and hence tax revenue), we can infer that DI allowances have net fiscal costs. Panel A.2 of Table 5 provides a direct accounting of these costs by summing the impact of DI allowances on DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts.[25] Our point estimates suggest that DI allowances granted on appeal increase annual net government spending (per household member) by nearly $7,000. A comparison of the point estimates in panels A.1 and A2. suggests that DI allowance raise household income by less than 50 cents per dollar of net government expenditure,

[23]There is no evidence of a significant effect of judge leniency on the likelihood of being excluded from the full sample, as shown in the fourth column of A.5, indicating that our estimates based on the restricted sample are unlikely to be biased by endogenous compositional changes. For details on the measurement of consumption expenditure, see Eika *et al.* (2017).

[24]Specifically, the row labeled $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$ reports p-values for the null hypothesis that the the causal effect of a successful DI appeal on household consumption in outcome years $t \in \{1, 2, 3, 4\}$ is equal to its immediate effect on DI benefits payments (panel B, column 1 of Table 4). We report two-tailed tests of equality. (P-values for one-tailed tests of $\Delta DI > \Delta HH$ consumption are equal to one-half the p-values for two-tailed tests.)

[25]Fiscal costs are equal to transfer income minus taxes, while household income is equal to transfer income minus taxes plus labor income and other market income (e.g. capital income). Since capital income plays a small role in the DI appellant sample, most of the offset is due to labor supply impacts.

and we reject the hypothesis that the rise in disposable income (per household member) is as large as the increase in fiscal costs per (household member) over the pooled four-year outcome period (see the final column of panel A.2).[26]

Panels B.1 and B.2 present 2SLS estimates of the effects of DI allowance on disposable income and consumption expenditure (per household member) for the restricted sample of appellants for whom we have complete consumption data. We estimate that DI allowances increase both disposable income and household consumption by roughly $16-18$ percent relative to their sample means, with both effects significantly different from zero at the 10 percent level for outcomes over the four post-appeal years. The fact that the point estimates for household income are broadly similar to those for consumption suggests that DI allowances have relatively little impact on household savings among appellants. Consistent with this observation, our data do not reject the null hypothesis that the consumption gains induced by DI allowances are equal on average to the income gains.

7 Heterogeneity in Impacts of DI Allowances by Marital Status

Recent evidence points to an important role for family labor supply in self-insuring household consumption against wage shocks (see e.g. Blundell *et al.* , 2016b). Motivated by this evidence, we examine whether DI allowances have differential impacts on household income and consumption among married and non-married households. We then explore how spousal responses to the allowance decision may help explain this heterogeneity.

7.1 Impacts on household income and consumption

Tables 6 and 7 examine how the economic consequences of DI determinations differ by household structure, and in particular between married and non-married appellants (i.e., those who are single or cohabiting). All outcomes are reported per household member in $1,000.

Focusing first on the subpopulation of *non-married* appellants, Table 6 documents that DI allowances generate large positive impacts on disposable income and consumption among non-married appellants, who comprise just over 40 percent of all appellants (Table 2). The panel A estimates indicate that a DI allowance raises the household incomes of unmarried appellants by approximately $6,600 per annum over the four years following appeal, and it generates net fiscal expenditures of approximately $12,300 per annum.[27] Our point estimates therefore imply that $0.55 cents of each dollar of public expenditure induced by a successful appeal by a non-married appellant accrues to household income, though we note that available precision does not allow us to reject the hypothesis that the effects on household incomes and fiscal expenditures are equal. Panel B focuses

[26]This test is reported in the bottom row of the panel and denoted p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$, indicating that we are comparing the net fiscal cost to the induced rise in household income in the contemporaneous year (or, in the final column, for the pooled four-year period).

[27]As above, we divide impacts on fiscal costs by household size so that both income gains and fiscal costs are scaled on a per household member basis.

Table 5: **Effect of DI Allowance on Household Income, Fiscal Costs and Consumption**

	Years after decision				Average
	1	**2**	**3**	**4**	
Panel A.	**Full sample**				
	A.1: Household income ($1,000)				
Allowed DI	1.282	5.578**	2.671	3.198	3.208*
	(1.998)	(2.249)	(2.127)	(2.008)	(1.649)
p-value for $H_0 : \beta_t = 0$	0.5212	0.0131	0.2091	0.1113	0.0518
p-value for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$	0.0000	0.0308	0.0003	0.0003	0.0000
Dependent mean	26.248	26.773	27.144	27.651	26.541
	A.2: Fiscal costs ($1,000)				
Allowed DI	3.627	8.914***	8.525***	7.010***	6.859***
	(2.286)	(2.057)	(2.213)	(2.395)	(1.756)
p-value for $H_0 : \beta_t = 0$	0.1126	0.0000	0.0001	0.0034	0.0001
p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$	0.3049	0.1049	0.0082	0.1115	0.0376
Dependent mean	7.017	7.730	8.283	9.036	7.671
Observations	13,972	13,842	13,709	13,607	14,092
Panel B.	**Restricted sample**				
	B.1: Household income ($1,000)				
Allowed DI	2.764	5.184**	2.352	4.951**	4.066**
	(2.293)	(2.063)	(2.693)	(2.386)	(2.032)
p-value for $H_0 : \beta_t = 0$	0.228	0.012	0.3825	0.038	0.0453
p-value for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$	0.0016	0.0197	0.0045	0.0345	0.0035
Dependent mean	25.318	25.86	26.222	26.768	25.634
	B.2: Household consumption ($1,000)				
Allowed DI	2.484	5.313*	1.896	4.728	4.705*
	(5.125)	(2.730)	(3.803)	(3.967)	(2.831)
p-value for $H_0 : \beta_t = 0$	0.6278	0.0517	0.6181	0.2333	0.0965
p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$	0.9565	0.9623	0.9044	0.9552	0.8214
Dependent mean	26.000	26.859	27.698	28.325	26.543
Observations	10,827	10,772	10,655	10,523	10,945

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Notes: This table reports the impact of DI allowance on household disposable income and fiscal costs for the baseline sample (panel A) and household disposable income and consumption for the restricted sample (panel B). All outcomes are reported per household member in $1,000. Baseline estimation sample consists of DI applicants who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The restricted sample excludes households with housing transactions and large financial transactions. There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and number of medical diagnoses. All control variables are measured prior to appeal. P-values for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$ in panels A.1 and B.1 correspond to tests of whether the effect of a successful DI appeal (in year $t0$) on initial DI income (in year $t1$) is equal to the effect of the successful appeal after on HH consumption in outcome years $t \in \{1, 2, 3, 4\}$. P-values for $H_0 : \beta_t = \Delta\text{HH Income}_t$ in panels A.2 and B.2 correspond to tests of whether the effects of a successful DI appeal in year $t0$ on household income and household consumption are equivalent in outcome years $t \in \{1, 2, 3, 4\}$
.

on the subset of non-married appellants for whom we have detailed consumption data. DI awards increase both income and consumption in this subpopulation, raising them by approximately $9,400 and $10,400 respectively. These are very large increments to both outcomes, equivalent to 35 to 40 percent of their baseline values. Estimated impacts on household income and consumption are highly comparable overall and in each year, and the p-values reported in the bottom of indicate that we cannot reject the hypothesis that DI allowances raise household incomes and consumption one-for-one in this subpopulation.

Table 7 reports analogous estimates for married appellants. Accounting for the effect of DI allowances on household labor supply and net payments across all public transfer programs substantially alters our picture of the income and consumption effects of disability receipt among married beneficiaries. As shown in Table 7, DI allowances are not estimated to increase household income or consumption of married applicants, and we can rule out with 95 percent confidence that any positive effect exceeds $2,500 (nine percent of baseline income). We can also strongly reject equality of the average effects of household income (per household member) on singles and unmarried versus married. These estimates imply that the combination of household labor supply and benefit substitution largely or fully offset the effects of DI benefit payments on household incomes of married appellants—though we stress that this does not mean that the welfare consequences of these transfers is nil, a point that we explore in our structural estimates below.[28] DI allowances made to married appellants do, however, incur meaningful fiscal costs through increased cash transfers and reduced payroll tax revenues. We estimate that each DI allowance to a married appellant generates a fiscal burden of approximately $4,000 per year in the four years following appeal. We readily reject the hypothesis that the average annual gain in household income among married appellants equals the average fiscal cost of a DI allowance for the marginal married appellant.

7.2 Spousal Responses

To help understand why the income and consumption effects of DI determinations differ by household structure, Table 8 extends our inquiry to consider spousal responses to DI allowance. We focus exclusively on married households in this analysis since our data do not allow us to determine whether non-married appellants are single or cohabiting.[29]

Panels A and B consider the effects of DI allowances on the labor supply and transfer payment receipt of the subset of appellants who are married. Though precision is quite limited in this

[28] In fact, DI allowances appear to weakly lower household income and consumption, plausibly reflecting discrete choices in labor supply by denied appellants' spouses (e.g., due to fixed costs associated with working).

[29] The Norwegian decennial census data allow us to observe cohabitation though unfortunately our annual administrative data do not. The Census data show that 59% of DI participants (applicants are not identified in the Census data) are married, 32% are single non-cohabitants, and only 9% are cohabitants. We test whether judge leniency causes endogenous selection into or out of marital status in Appendix Table A.5. Columns 2, 3 and 4 find no evidence that judge leniency affects the the likelihood of a change in marital status (overall, from unmarried to married, or from married to unmarried).

Table 6: **Effect of DI Allowance on Household Income, Fiscal Costs and Consumption for Non-married Appellants**

	Years after decision				
	1	**2**	**3**	**4**	**Average**
Panel A.	**Full sample**				
	A.1: Household income ($1,000)				
Allowed DI	4.637	8.669***	7.710**	6.296	6.577**
	(3.380)	(3.130)	(3.604)	(4.127)	(2.803)
p-value for $H_0 : \beta_t = 0$	0.1701	0.0056	0.0324	0.1271	0.019
p-value for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$	0.0000	0.0000	0.0001	0.0002	0.0000
Dependent mean	25.346	25.745	26.164	26.731	25.549
	A.2: Fiscal costs ($1,000)				
Allowed DI	9.027*	13.168***	13.912***	9.255**	12.352***
	(5.162)	(3.354)	(4.259)	(4.228)	(4.045)
p-value for $H_0 : \beta_t = 0$	0.0803	0.0001	0.0011	0.0286	0.0023
p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$	0.395	0.1799	0.1453	0.4841	0.1534
Dependent mean	12.893	13.302	13.66	14.137	13.108
Observations	6,128	6,102	6,061	6,059	6,147
Panel B.	**Restricted sample with data on consumption expenditure**				
	B.1: Household income ($1,000)				
Allowed DI	6.924**	9.451***	7.450*	10.438**	9.443***
	(3.330)	(3.052)	(4.376)	(4.394)	(3.160)
p-value for $H_0 : \beta_t = 0$	0.0376	0.002	0.0887	0.0175	0.0028
p-value for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$	0.0000	0.0001	0.0017	0.0142	0.0002
Dependent mean	24.669	25.214	25.695	26.203	24.979
	B.2: Household consumption ($1,000)				
Allowed DI	6.716	9.793**	9.164	13.954**	10.366**
	(6.664)	(4.203)	(7.348)	(6.426)	(5.151)
p-value for $H_0 : \beta_t = 0$	0.3136	0.0198	0.2124	0.0299	0.0442
p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$	0.9751	0.9351	0.8156	0.5843	0.8578
Dependent mean	25.907	26.178	27.760	28.099	26.135
Observations	4,979	4,946	4,891	4,880	5,001

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Notes: This table reports the impact of DI allowance on household disposable income and fiscal costs for the full set of unmarried appellants (panel A) and the subset for whom we have detailed consumption data (panel B). All outcomes are reported per household member in $1,000. Columns 1-4 report separate estimates for each year, whereas column 5 reports estimates for the average outcome from 0 to 5 years after the appeal decision. Baseline estimation sample consists of unmarried DI applicants who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The restricted sample excludes households with housing transactions and large financial transactions. There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and number of medical diagnoses. All control variables are measured prior to appeal. P-values for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$ in panels A.1 and B.1 correspond to tests of whether the effect of a successful DI appeal (in year $t0$) on initial DI income (in year $t1$) is equal to the causal effect of the successful appeal on HH income in outcome years $t \in \{1, 2, 3, 4\}$. P-values for $H_0 : \beta_t = \Delta\text{HH Income}_t$ in panels A.2 and B.2 correspond to tests of whether the causal effects of a successful DI appeal in year $t0$ on fiscal costs (A.2) and household consumption (B.2) in outcome years $t \in \{1, 2, 3, 4\}$ are equal to the corresponding effect on household income.

Table 7: **Effect of DI Allowance on Household Income, Fiscal Costs and Consumption for Married Appellants**

	Years after decision 1	2	3	4	Average
Panel A.	**Full sample**				
	A.1: Household income ($1,000)				
Allowed DI	-1.705	2.007	-2.031	-0.710	-1.230
	(2.361)	(2.858)	(2.544)	(2.882)	(1.918)
p-value for $H_0 : \beta_t = 0$	0.4702	0.4824	0.4246	0.8054	0.5214
p-value for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$	0.0248	0.5785	0.027	0.1352	0.0119
Dependent mean	26.953	27.582	27.92	28.389	27.215
	A.2: Fiscal costs ($1,000)				
Allowed DI	1.053	6.002*	5.826*	6.420**	3.975*
	(2.890)	(3.358)	(3.300)	(2.836)	(2.365)
p-value for $H_0 : \beta_t = 0$	0.7157	0.0739	0.0774	0.0236	0.0928
p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$	0.3401	0.2342	0.0173	0.0119	0.0277
Dependent mean	2.426	3.337	4.022	4.941	3.256
Observations	7,844	7,740	7,648	7,548	7,945
Panel B.	**Restricted sample**				
	B.1: Household income ($1,000)				
Allowed DI	-1.300	0.078	-2.265	-2.606	-2.119
	(2.733)	(2.781)	(2.920)	(3.086)	(2.073)
p-value for $H_0 : \beta_t = 0$	0.6343	0.9777	0.4379	0.3984	0.3068
p-value for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$	0.2038	0.4512	0.1285	0.1215	0.0384
Dependent mean	25.872	26.408	26.669	27.257	26.055
	B.2: Household consumption ($1,000)				
Allowed DI	-0.548	0.215	-5.212	-5.771	-0.165
	(6.552)	(4.107)	(4.987)	(4.416)	(3.229)
p-value for $H_0 : \beta_t = 0$	0.9333	0.9582	0.296	0.1913	0.9593
p-value for $H_0 : \beta_t = \Delta\text{HH Income}_t$	0.9086	0.9733	0.5546	0.4736	0.5451
Dependent mean	26.079	27.438	27.645	28.520	26.680
Observations	5,848	5,826	5,764	5,643	5,944

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Note: This table reports the impact of DI allowance on household disposable income and fiscal costs among the full set of married appellants (panel A) and the subset for whom we have detailed consumption data (panel b). Outcomes are reported per household member in $1,000. Baseline estimation sample consists of married DI applicants who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The restricted sample excludes households with housing transactions and large financial transactions. There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. All control variables are measured prior to appeal. P-values for $H_0 : \beta_t = \Delta\text{DI benefit}_{t1}$ in panels A.1 and B.1 correspond to tests of whether the effect of a successful DI appeal (in year $t0$) on initial DI income (in year $t1$) is equal to the effect of the successful appeal on HH income in outcome years $t \in \{1, 2, 3, 4\}$. P-values for $H_0 : \beta_t = \Delta\text{HH Income}_t$ in panels A.2 and B.2 correspond to tests of whether the causal effects of a successful DI appeal in year $t0$ on fiscal costs (A.2) and household consumption (B.2) in outcome years $t \in \{1, 2, 3, 4\}$ are equal to the corresponding effect on household income.

subsample, we estimate that labor supply reductions and transfer payment increases roughly offset each other. Panels C and D consider the impact of of allowances versus denials at appeal on potential compensatory behaviors among appellants' spouses. The 2SLS estimates show that the labor supply of appellants' spouses responds strongly to the outcomes of disability determinations. Relative to spouses of denied appellants, spousal earnings of allowed appellants fall by approximately \$5,000 in the first year after a successful appeal, and by a further \$11,000 to \$12,000 in years two through four following the award as show in panel C. These estimated labor supply reductions in years two through four, averaging \$16,500, are statistically significant. Panel D shows, however, that as much as 50 percent of the reduction in spousal labor earnings induced by a DI allowance is effectively offset by a countervailing increase in transfer payments to the spouse.[30] We note, however, that this estimated positive effect on transfer payments is statistically significant only in the fourth year following appeal. It is also positive and large in years two and three but not precisely estimated.

The estimated effects on spousal earnings are consistent with the possibility that either spouses reduce their labor supply if appellants are allowed, or that spouses increase their labor supply if appellants are denied (or potentially both). The latter would suggest that DI denials induce an added worker effect; the loss in worker earnings (due to disability) absent an offsetting gain in DI income spurs spouses to increase their labor supply. The former possibility would be consistent with DI allowances inducing a decline in labor supply among spouses due to leisure complementarities. We explore which interpretation is supported by the data by decomposing the causal effects estimates in Table 8 into potential outcomes of spouses of complier appellants if denied or, alternatively, if allowed at appeal. This decomposition, found in Appendix Figure A.6, indicates that the behavioral response found in Table 8 stems almost entirely from spousal responses to DI denials: spouses of denied appellants strongly increase earnings in the years following denial; conversely, spouses of allowed appellants exhibit little earnings adjustment. By implication, DI denial induces a powerful added-worker effect among spouses.

This result is somewhat surprising at first blush since households do not *lose* income when DI appeals are denied, they simply fail to gain it. However, recall from Figure 3 (center right panel) that average labor income of DI appellants declines by approximately 40 percent—from about $\$25,000$ to about $\$15,000$—over the four years prior to appeal, while close to 80 percent of applicants are awarded DI benefits at their initial determination (Table 2). It thus appears plausible that, from the perspective of DI appellants and their spouses, the denial of benefits at the appeal stage constitutes a substantial adverse shock to expected permanent income, potentially spurring an added-worker response.

[30] In Appendix Table A.4, we run the analysis of appellant labor income and total transfer payments from Table 4 separately for married appellants versus single and unmarried appellants. While the thinner sample size available for these estimations reduces precision, the point estimates suggest that DI allowances generate somewhat smaller reductions in labor earnings, as well as smaller increases in total (individual) transfer payments, among married appellants as compared to single and unmarried appellants. This is consistent with the hypothesis that the income effects of transfers on own labor supply are much larger for unmarried disability beneficiaries due to the absence of implicit spousal income insurance.

Table 8: **Effect of DI Allowance on Spousal Earnings and Transfer Payments**

	Years after decision				
	1	**2**	**3**	**4**	**Average**
Panel A.	**Married appellant's labor earnings ($1,000)**				
Allowed DI	-5.042	-0.444	-4.426	-3.912	-3.566
	(3.461)	(4.068)	(3.993)	(3.625)	(3.269)
Dependent mean	14.991	14.784	14.168	13.535	14.238
Panel B.	**Married appellant's total transfers ($1,000)**				
Allowed DI	9.110**	6.499	5.008	5.395	5.948
	(4.000)	(4.423)	(3.703)	(3.628)	(3.662)
Dependent mean	16.621	17.356	17.919	18.508	17.497
Panel C.	**Spouses' labor earnings ($1,000)**				
Allowed DI	-4.856	-17.009**	-16.096**	-16.794**	-10.488
	(8.102)	(8.552)	(7.828)	(8.039)	(7.345)
Dependent mean	40.965	39.565	38.777	37.487	39.025
Panel D.	**Spouses' total transfers ($1,000)**				
Allowed DI	-0.027	5.823	5.957	8.020*	4.061
	(3.334)	(3.683)	(4.152)	(4.614)	(3.609)
Dependent mean	11.196	11.938	12.622	13.349	12.4
Observations	7,844	7,740	7,648	7,548	7,844

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Note: This table reports the impact of DI allowance on earnings and total transfers of married appellants (panels A and B) and their spouses (panels C and D). Baseline estimation sample consists of married DI applicants who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and number of medical diagnoses. All control variables are measured prior to appeal.

To drill down further on this finding, we have also explored the heterogeneity of the added-worker effect among households according to the spouse's sex, education, and prior earnings. While we find suggestive evidence that the labor supply of female spouses is more responsive to DI denial than is the labor supply of male spouses, these contrasts are not typically significant due to limited statistical power. Appellants may also differ in their own labor supply responses or benefit substitution following the DI determination. To examine heterogeneity in responses according to observable characteristics, we have also performed 2SLS estimation of equations (1) and (2) for each of the subsamples reported in Appendix Table A.3, where we split appellants according to diagnosis, gender, age, education, and the size of the household. These subsample estimates are insufficiently precise to draw clear inferences.

8 Deriving Welfare Implications using a Structural Model

We now apply the data and findings above to estimate a dynamic model of household behavior that translates employment, savings and reapplication decisions of appellants and their spouses into revealed preferences for leisure and consumption. We use this estimated model to compute the welfare benefits of DI receipt—by which we mean the cash equivalent value of DI allowance at appeal—and to perform counterfactual simulations that allow us to infer the extent to which spousal labor supply and other mechanisms attenuate the loss in household welfare from DI denial at appeal. Our goal is limited to understanding the post-appeal labor supply, savings, and DI reapplication decisions of appellants and their spouses, taking as given their characteristics and economic circumstances at the time of appeal, such as savings, disability severity, education, and past labor market experience. As a consequence, our model does not speak to the full insurance value of the DI system to workers prior to disability onset; only the cash equivalent value of DI allowances at the time of appeal.

8.1 Description of model and estimation procedure

Preference specification

We consider a unitary model of the household with non-separable preferences between consumption and leisure and, for married households, interactions between the value of leisure (or equivalently disutility of working) of the spouses.[31] The household utility function depends on consumption (per household member) C, indicators for employment $P_A \in \{0,1\}$ and reapplication $R \in \{0,1\}$ for the appellant, and an employment indicator $P_S \in \{0,1\}$ for the spouse (if present).[32] As in Low & Pistaferri (2015), we allow for preference heterogeneity according to disability severity. Following their paper, we construct an index of disability severity with three levels: $H = 0$ indicates low-severity, $H = 1$ indicates mid-severity, and $H = 2$ indicates high-severity. We construct this index by estimating the probability of being initially allowed at appeal as a function of the diagnosis codes, and we assign appellants to three groups of equal size based on these predicted probabilities.

We follow Low & Pistaferri (2015) in the parametric specification of preferences. At time t, the instantaneous utility function of unmarried households with a given disability severity is:

$$U_1\left(C_t, P_{A,t}, R_t; H\right) \quad = \quad \frac{1}{1-\mu_1}\left(C_t \exp\left\{-P_{A,t}\phi_{1,A,H}\right\} - R_t \exp\omega_1\right)^{1-\mu_1}, \tag{3}$$

[31] This flexible specification of preferences accommodates non-market production and work-related expenses and allows for the possibility that spouses may enjoy leisure more when they are together. For evidence on such non-separability, we refer to Browning & Meghir (1991), Blundell *et al.* (1994), Aguiar & Hurst (2013) and Blundell *et al.* (2016b).

[32] As in Maestas *et al.* (2013) and Kostol & Mogstad (2014), employment is an indicator variable that is equal to one if annual earnings exceed the annual substantial gainful activity threshold, set annually by the Norwegian Social Security Administration (at approximately USD 10,000 per year). We are unable to measure labor supply at the intensive margin because we lack reliable data on working hours.

where $\phi_{1,A,H}$, μ_1, and ω_1 are the utility parameters. The bracketed expression reflects how the marginal utility of income changes with working; it is normalized to zero if the appellant is not working. For married households with appellants of a given disability severity, we use a similar parametric specification of preferences:

$$U_2\left(C_t, P_{A,t}, P_{S,t}, R_t; H\right) = \frac{1}{1-\mu_2}\left(C_t \exp\left\{-P_{A,t}\phi_{2,A,H} - P_{S,t}\phi_{2,S}\right\} - R_t \exp \omega_2\right)^{1-\mu_2}, \quad (4)$$

where $\phi_{2,A,H}$, $\phi_{2,S}$, μ_2, and ω_2 are the utility parameters, and the bracketed expression is normalized to zero if both spouses are not working.

Earnings Process

Like Low & Pistaferri (2015), we specify the process of (latent) earnings of appellants to depend on disability severity H and other observable characteristics Q. The vector Q includes a constant, indicators for high school drop out, high school completion, and college completion; and an indicator for young age, where we follow Low & Pistaferri (2015) in defining young disability appellants as those less than 45 of years age at time of appeal. The observable characteristics are measured prior to the appeal decision, capturing heterogeneity in experience, skills, and abilities that may affect potential earnings. In addition, we include a fixed effect f_A in earnings estimated from pre-application earnings data to allow for heterogeneity in latent ability as measured prior to the application.

We specify the annual earnings process of the appellant to be:

$$\log W_{A,t} = Q'\kappa_{M,A} + \sum_{j=0}^{2} \psi_{M,A,j} H_j + \tau_{A,t} + a_{M,A}\left(f_A\right), \quad (5)$$

where $M = 1$ denotes single households and $M = 2$ denotes married households, $a_{M,A}$ is a (third order) polynomial in the pre-application fixed effect, $H_j = \mathbf{1}\{H = j\}$ is an indicator for disability severity $j = \{0, 1, 2\}$, and the stochastic component $\tau_{A,t}$ is specified as a random walk:

$$\tau_{A,t} = \tau_{A,t-1} + \nu_{A,t}, \quad \nu_{A,t} \sim \mathcal{N}\left(0, \sigma^2_{A,M}\right).$$

Similarly, the annual earnings process of the spouse (if present) is specified as:

$$\log W_{S,t} = Q'\kappa_S + \sum_{j=0}^{2} \psi_{S,j} H_j + \tau_{S,t} + a_S\left(f_S\right), \quad (6)$$

where:

$$\tau_{S,t} = \tau_{S,t-1} + \nu_{S,t}, \quad \nu_{S,t} \sim \mathcal{N}\left(0, \sigma^2_S\right).$$

As pointed out by Heckman (1979), a potential concern with the (latent) earnings processes is that earnings are not observed for those who do not work and the decision to work depends on the earnings offer. While the observable characteristics and the pre-application fixed effects may help address this concern over selection bias, we also perform a robustness check. As in Low & Pistaferri (2015), we perform a selection correction of the earnings processes by estimating a probit regression of employment on H, Q , and $a_{M,A}$ and including the inverse Mills ratio of this estimated value in the earnings process regressions. Under the assumption that the error terms of the employment equations and the earnings processes are jointly normal, this procedure provides the appropriate parametric selection correction. Appendix Table A.7 and A.9 present parameter estimates of the earnings processes and the corresponding labor supply elasticities with and without the selection correction. As shown in Appendix Table A.10, neither the estimated cash equivalent value of DI allowance nor the results from the counterfactuals are materially affected by inclusion of the selection terms in the earnings processes.

Disposable Income

As in Heathcote *et al.* (2014) and Blundell *et al.* (2016b), we approximate the tax-transfer system by specifying flexible functions mapping household earnings into disposable household income (earnings plus transfers minus taxes). Below, we show that the chosen functions approximate well the effective tax rates implicit in the complex Norwegian tax-transfer system.

For unmarried households that supply labor, we use the following specification of the relationship between disposable income $I_{1,t}$ and appellant earnings $E_{A,t}$ in year t:

$$I_{1,t} = (1 - \Lambda_{1,D_t,K,t}) \left(E_{A,t}\right)^{\left(1-\Psi_{1,D_t,K,t}\right)} \tag{7}$$

where $D \in \{0,1\}$ denotes current DI receipt and $K \in \{0,1\}$ denotes the presence of a dependent in the household. Similarly, the specification for married households that supply labor is:

$$I_{2,t} = (1 - \Lambda_{2,D_t,K,t}) \left(E_{A,t} + E_{S,t}\right)^{\left(1-\Psi_{2,D_t,K,t}\right)} \tag{8}$$

where $E_{S,t}$ and $I_{2,t}$ denote annual spousal earnings and disposable income in year t, respectively. For households that do not supply labor, disposable income$I_{M,t}$ is only a function of transfer payments as captured by the specification:

$$I_{M,t} = \Phi_{M,D_t,K,t} \tag{9}$$

The parameters Λ, Ψ and Φ are allowed to vary over time t and by marital status M, DI receipt D_t, and presence of dependents K. In a proportional tax-transfer system, $\Psi = 0$ and Λ is the proportional effective tax rate. By contrast, if $0 < \Psi < 1$, then the marginal effective tax rate is increasing in earnings.

Process for Approval of Reapplication

Like Low & Pistaferri (2015), we model DI approval upon reapplication for single and unmarried households as:

$$D_{t+1} = D_t + R_t (1 - D_t) \pi_{M,H,t}. \tag{10}$$

where $\pi_{M,H,t} \in [0,1]$ is the probability of DI approval upon reapplication ($R = 1$), which we allow to vary with time t, marital status M, and disability severity H.

The Household's Problem

Letting S_t denote savings, ζ denote the intertemporal discount factor, and O_1 denote household heterogeneity, the dynamic optimization problem of unmarried households is:

$$V_{1,t}(D_t, \tau_{A,t}, S_t; O_1) = \max_{C_t, P_{A,t}, R_t, S_{t+1}} U_1 (C_t, P_{A,t}, R_t; H) + \zeta \mathbb{E} V_{1,t+1}(\cdot, \cdot, S_{t+1}; O_1),$$

where $O_1 = (H, K, Q, f_A)$. The expectation is taken jointly across the distribution of D_{t+1} and $\tau_{A,t+1}$, and the choices are subject to the exogenous earnings process, the exogenous DI approval process, the borrowing constraint $S_{t+1} \geq 0$, the tax-transfer system, and the intertemporal budget constraint:

$$S_{t+1} = (1 + r)(I_t + S_t - C_t) \tag{11}$$

where r is the real interest rate. To close the model, we follow Low & Pistaferri (2015) in assuming exogenous and fully anticipated retirement at age 62. Retirement is characterized by consuming out of savings and retirement benefits b_1. At the end of retirement, death occurs exogenously. The terminal condition of zero savings must be satisfied upon death.

The dynamic optimization problem of married households is analogously,

$$V_{2,t}(D_t, \tau_{A,t}, \tau_{S,t}, S_t; O_2) = \max_{C_t, P_{A,t}, P_{S,t}, R_t, S_{t+1}} U_2 (C_t, P_{A,t}, P_{S,t}, R_t; H) + \zeta \mathbb{E} V_{2,t+1}(\cdot, \cdot, \cdot, S_{t+1}; O_2),$$

where $O_2 = (H, K, Q, f_A, f_S)$. The expectation is taken jointly across the distribution of D_{t+1}, $\tau_{A,t+1}$, and $\tau_{S,t+1}$, and the choices are subject to the exogenous earnings process of each spouse, the exogenous DI approval process, the borrowing constraint, the tax-transfer system, and the intertemporal budget constraint. The model for married households is closed the same way as for single and unmarried households, with retirement benefits denoted by b_2.

The model for married households allows two sources of interdependencies between spouses. As usual in models of household labor supply, spouses depend on one another through the household budget constraint, as the earnings of each spouse is assumed to be shared in the household's consumption. This means the wages and labor supply of one spouse affect the incentives for the other spouse to work through the resources available for household consumption. In addition, the household utility function is specified such that the labor disutility of one spouse may depend on whether the other spouse is working, which accommodates leisure complementarity (or the value of caring for a non-working disabled spouse). Blundell *et al.* (2016b) find evidence of such leisure complementarity in a model of household labor supply.

Estimation and identification

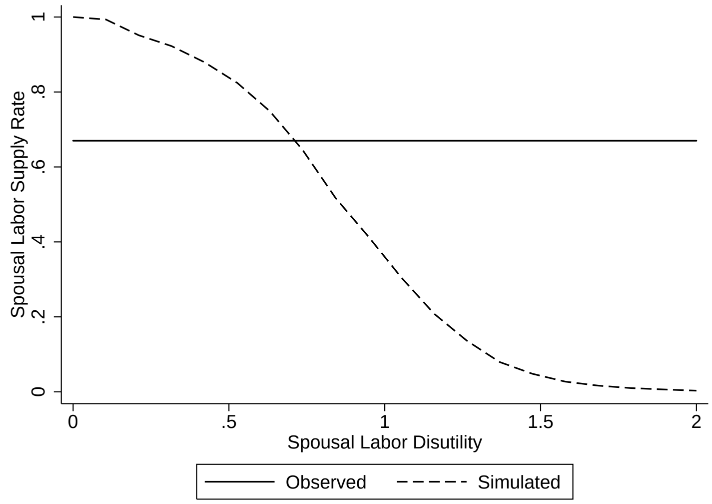
To take the model to the data, we adopt a three-step procedure. In the first step, we estimate the parameters of the earnings process similarly to Low & Pistaferri (2015), as described above. In the second step, certain model parameters are set externally. We follow Low & Pistaferri (2015) in fixing the relative risk aversion coefficient to $\mu_M = 1.5$, the intertemporal discount rate to $\zeta = 0.9756$, and the interest rate to $r = 0.016$. We specify annual retirement benefits as the minimum pension benefits provided under Norwegian law.

In the third step, we use the method of simulated moments to jointly estimate all remaining parameters, conditional on the estimated earnings parameters from the first step and the externally set parameters from the second step. For any given candidate parameters, we simulate the model recursively beginning with the terminal condition and ending with decisions made in the period after initial DI allowance.[33] The output of the model simulation is the set of optimal choices that would be made by the household given the candidate parameters. We choose the optimal parameters as those candidate parameters that minimize a weighted distance metric between observed data moments (discussed below) and corresponding moments simulated from the model. See Appendix B for further details.

All estimated model parameters except for the disutility parameters are identified directly from sample data moments given the distributional assumptions about the error terms. The parameter $\pi_{M,H,t}$ in Equation (10) is the DI approval rate upon reapplication for households with marital status M and disability severity H at time t, identified from the observed DI approval rate in the sample conditional on (M, H, t). Equation (8) can be expressed equivalently as the linear regression of log disposable income on log earnings of the household, where the terms $\log(1 - \Lambda_{M,D_t,K,t})$ and $1 - \Psi_{M,D_t,K,t}$ are the intercept and slope, respectively. Conditional on (M, D_t, K, t), the parameters are identified from the mean of log disposable income and the mean of the expectation of log disposable income conditional on log earnings among households in which neither the appellant

[33] The value function does not have a closed form solution. To solve for it, we discretize the state-space by forming a grid in each of savings, appellant potential earnings state, and spousal potential earnings state (if a spouse is present), as detailed in Appendix B.1. Each grid is formed using equally-spaced quantiles from the observed marginal distributions of savings and earnings, so that more grid points are positioned around denser regions. To compute the continuation value, bivariate Gaussian quadrature is used to integrate across the joint distribution of earnings shocks for the appellant and spouse (if present). Cubic spline interpolation is fit to map the discretized value and policy functions into continuous value and policy functions, as detailed in Appendix B.2. The value functions are computed and cubic splines fit separately for each discrete type in the state-space of the model. Finally, the cubic splines are applied to the full sample in each observed time period to simulate the optimal choices of each household as a function of their discrete and continuous state-space values.

Figure 5: **Identification: Using Single-crossing to Pin Down Utility Parameters**



Notes: The x-axis is the parameter representing the disutility of labor for spouses in married households, which we vary while fixing all other parameters to their estimated values. The y-axis is the average labor supply of spouses. The solid line indicates the observed value in the data, while the dashed line indicates the value simulated from the estimated model. The spousal disutility of labor is pinned down as the point on the x-axis that corresponds to the crossing of the solid and dashed lines.

nor the spouse (if present) is working. Analogously, $\Phi_{M,D_t,K,t}$ is identified as the mean transfer to unemployed households conditional on (M, D_t, K, t). Although the tax-transfer parameters could be estimated in a first step like the earnings process, we estimate these parameters simultaneously with the other model parameters in order to improve efficiency.

While the mapping between model parameters and sample moments is less direct for the disutility parameters, there are data moments that intuitively provide identifying information. While all parameters are estimated simultaneously, it can be instructive to focus on one parameter at a time. For instance, consider the disutility of labor supply of an unmarried appellant of a given health H, $\phi_{1,A,H}$. This parameter is pinned down by seeking the value of the disutility of labor supply that is consistent with the observed employment rate for this group of appellants given the gains in disposable income from working, which are in turn determined by the wage equation and the tax transfer function (which are identified separately). Appendix Figure A.7 illustrates this exercise by using the model to simulate or predict the employment rate for any given value of $\phi_{1,A,H}$. Holding all other parameters fixed, the simulated employment rate is monotonically decreasing in $\phi_{1,A,H}$. The disutility of labor is pinned down as the parameter value that equates the simulated employment rate to its observed value. The other disutility parameters are recovered by similar revealed preference arguments. Figure 5 and Appendix Figure A.7 demonstrate that each disutility parameter has a unique value that matches the simulated moment to the observed data moment.

We estimate the model separately for married couples (47 parameters) and unmarried individuals (46 parameters). To pin down these parameters, we match 57 moments (47 raw data moments and ten IV estimates) for married couples and 52 moments (46 raw data moments and six IV estimates) for unmarried individuals. We choose two sets of moments to match. The first set

Table 9: **Fit of Instrumental Variables Estimates: Income and Consumption**

	Income (per capita)	**Consumption (per capita)**
Panel A.	**Married**	
IV estimate of effect of DI participation	-2.119	-0.165
Simulated effect of DI participation	-1.610	-2.631
Panel B.	**Single and Unmarried**	
IV estimate of effect of DI participation	9.443	10.366
Simulated effect of DI participation	9.815	8.894

Notes: This table compares the IV estimates of initial DI allowance on consumption and disposable income to the model simulated effects of these variables. All units are in $1,000 USD.

consists of raw data moments, chosen based on the identification arguments above. These moments are mean log disposable income and expected log disposable income conditional on log earnings among households that supply labor; mean disposable income among households that do not supply labor; and employment rates and reapplication rates among those not receiving DI. Each of these moments is matched conditional on observable types over which the parameters vary in order to pin down all of the type-specific model parameters. The second set of moments is the IV results for consumption, disposable income, and earnings among appellants and spouses. These are included to discipline the model to recover our estimates of the causal effects of DI allowance.[34]

8.2 Empirical results

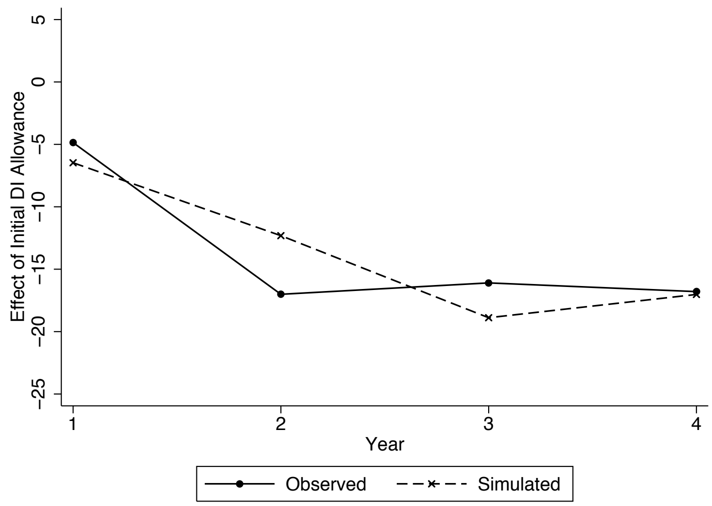
Parameter estimates and model fit

The externally set parameters and estimated utility parameters are presented in panels A and B of Appendix Table A.8, respectively, while the estimated parameters from the earnings processes are reported in Appendix Table A.7. The parameters are precisely estimated. As anticipated, and in agreement with results by Low & Pistaferri (2015), we find that the disutility of labor for appellants is strictly increasing in the severity of disability for married as well as single and unmarried households.

To interpret the magnitude of the utility parameters, we use the fitted models to simulate how employment rates of appellants and spouses change with a permanent one percent increase in

[34]We equally weight the two sets of moments. Within each set, we use the diagonal weighting matrix to form the objective function, which is equivalent to weighting each deviation between an observed and simulated moment by the inverse of the standard deviation of the observed moment. This is the form of the objective function in Equation (13) of Blundell *et al.* (2016a) and is motivated by the finding of Altonji & Segal (1996) that the asymptotically efficient weighting matrix has poor small-sample properties. We use a particle swarm numerical optimization algorithm to solve for the globally optimal parameters, and we validate the optimum locally using the BFGS algorithm. We use the block bootstrap to perform inference. See Appendix B.3 for further details.

Figure 6: **Fit of Instrumental Variables Estimates: Spousal Earnings Dynamics**



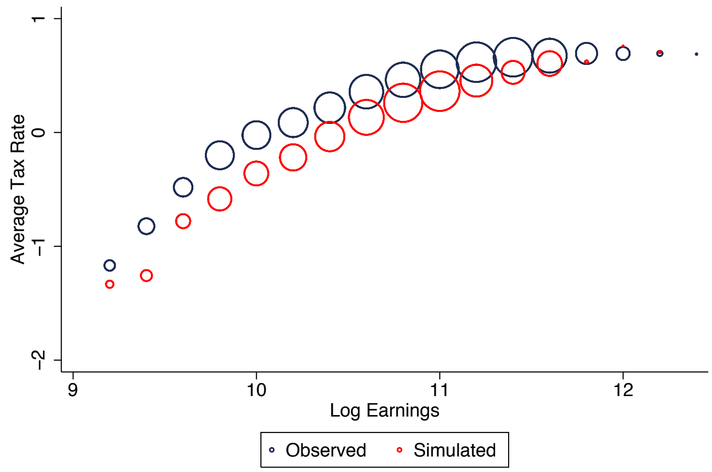
Notes: This figure compares the IV estimates of the effects of DI allowance on the earnings of spouses to those simulated from the estimated model. All units are in $1,000 USD.

disposable income from working, obtaining (Marshallian) labor supply elasticities (see Appendix Table A.9). Because few appellants on DI are working, we focus on employment responses in the non-DI state. We obtain plausible labor supply elasticities. Our own-wage labor elasticities range from 0.20 to 0.36. Keane (2011) provides a survey of own wage Marshallian labor elasticities, which range from -0.47 to 0.51 in papers published over the past two decades. In more recent work, Blundell *et al.* (2016b) obtain estimates ranging from -0.08 to 0.42, while Blundell *et al.* (2016a) obtain estimates ranging from 0.22 to 1.36 for females only. Our cross spouse wage labor elasticities range from -0.35 to -0.30. By comparison, Blundell *et al.* (2016b) reports estimates ranging from -0.75 to -0.22.

Table 9 compares the IV estimates of the causal effects of DI allowance on disposable income and consumption (repeated from Tables 6 and 7) to the effects simulated from the model.[35] The model is relatively successful in replicating both consumption and income effects of DI allowance. Figure 6 compares the IV estimates of the effect of DI allowance on spousal labor earnings over time in the data versus the simulation. Importantly, the model is able to replicate the time trend in the effects of DI participation on spousal employment. When we consider the other targeted moments (which vary much less over time), the model also performs well. Appendix Figure A.8 summarizes how the models fit all 109 moments (including the IV estimates and the raw data moments). Since the variables behind the moments are measured in different units (e.g. income versus employment), we divide the difference between an actual and a simulated moment by the standard deviation of the respective variable. The distribution is centered around zero, and only for a small fraction of

[35] The complete set of moment fits—targets, weights, and simulated values—is omitted for brevity and is available from the authors.

Figure 7: **Fit of Average Tax Rates Simulated from the Model**



Notes: This figure compares the simulated to the observed average tax rates across the distribution of earnings. The circle width represents the relative density of observed and simulated earnings, respectively.

moments do we observe differences that exceed one standard deviation.

Lastly, Figure 7 shows that we approximate well the average effective tax rates implicit in the tax-transfer system. The similar widths of the circles in this figure illustrate that the density of observed earnings is well-replicated by the model, even though the distributions of earnings and average effective tax rates are not directly targeted in the model estimation.

Household valuation of DI receipt

Building on the successful fit of the structural estimates to the IV estimates of the effects of DI allowance on earnings and total household income and consumption—and the plausible labor supply elasticities implied by the model—we now apply the estimated structural model to explore households' willingness to pay for an initial DI allowance, by which we mean the yearly disposable income that appellants would be willing to give up to be allowed DI at appeal. We refer to Appendix B.4 for details on how this willingness to pay value is computed.

The results of this estimation are shown in Table 10. Panel A presents the average willingness to pay for initial DI allowance and compares it to the average income and fiscal cost effects presented above. We find that non-married households have statistically significant and relatively high average willingness to pay for a DI allowance ($11,316 yearly, per household member). By comparison, the average willingness to pay is relatively low for married households ($2,300 yearly, per household member). Note that this willingness to pay value is annualized for comparability with the annual income and fiscal cost effects; the total lifetime value of an initial DI allowance is of course many times greater than the annualized value. Comparing the willingness to pay to the effect of DI allowance on disposable income, we estimate that average willingness to pay among married households is about \$4,400 greater (per year) than the net effect of an initial DI allowance on household income. For unmarried households, estimated willingness to pay is about \$1,900 greater than the net effect of the allowance on household income. This pattern of results indicates that married households primarily value receiving a DI allowance because it enables a reduction in labor supply whereas unmarried households primarily value receipt of a DI allowance because it raises household disposable income.

Table 10: **Estimated Household Valuation of DI Receipt**

	Married	**Single and Unmarried**
Panel A.	**Average Willingness to Pay (\$1,000)**	
Average WTP:	2.300	11.316
p-value for $H_0 : WTP = 0$	0.700	0.000
Average WTP net of income effect:	4.419	1.873
Average WTP per dollar of fiscal costs:	0.579	0.916
Average WTP as a share of household consumption	0.086	0.433
Panel B.	**Distribution of Willingness to Pay**	
Percentiles		
25th:	1.161	4.327
50th:	2.216	8.490
75th:	3.031	15.594

Note: This table shows estimates of the average welfare benefit (\$1,000, per household member, annuitized over the four years after initial DI allowance) of DI receipt for married people and single and unmarried individuals. In the row titled, "Net of income effect," we subtract the average effect of DI on disposable income from the average willingness to pay. In the row titled "per dollar of fiscal costs," we divide the average willingness to pay by the average effect of DI on fiscal costs. In the row titled "as a share of household consumption," we divide the willingness to pay by the level of consumption for each household. The hypothesis test $H_0 : WTP = 0$ correspond to testing whether the willingness to pay for DI receipt is equal to zero. P-values are based on reestimating the model on 20 block bootstrap replicates of the data (where the block corresponds to the individual).

Comparing the willingness to pay for a DI allowance relative to its fiscal cost suggests that, on average, each net dollar in public expenditure induced by a DI allowance raises the (money metric) welfare of single and unmarried awardees by nearly \$0.92. While the fiscal costs of DI allowance are nearly twice as large as the money-metric welfare benefits accruing to married households, the estimated willingness to pay of \$0.58 per dollar of fiscal costs remains substantial. Benchmarking willingness to pay against consumption levels, the average DI allowance is valued at about about 43% of annual consumption for single and unmarried households and 9% of annual consumption for married households. Panel B displays several moments of the estimated willingness to pay for DI allowances across households. The difference in willingness to pay between unmarried households and married households is $3,166 at the 25th percentile, $6,274 at the median, and $12,563 at the 75th percentile.

While the results in Table 10 suggest that valuation of DI receipt is substantially lower among married than unmarried households, we note that these estimates do not account for the ex ante insurance value of DI, and hence may understate total household valuation of the DI *system*. It is therefore important to bear in mind that these estimates do not preclude the possibility that both unmarried and married households value the DI system at more than its cost.

Quantifying the importance of spousal labor and other insurance mechanisms

In table 11, we report results from counterfactual analyses that help us assess the extent to which spousal labor supply, savings, and reapplication buffer the household welfare consequences of a DI denial versus allowance at appeal.

In the first counterfactual exercise, reported in panel B, we set each spouse's labor supply to be equal to his or her labor supply in the year prior to the appeal decision. In effect, the spouse is prevented from adjusting labor supply in response to whether the appellant is initially allowed or denied DI. Eliminating the option for a spousal labor supply response substantially increases the willingness of married households to pay for DI; indeed, this restriction substantially eliminates the difference in the willingness to pay of married versus unmarried households. This result underscores the importance of spousal labor supply as an alternative household-level mechanism for buffering income losses from disability that are not compensated by the DI program.

In the next counterfactual exercise (reported in panel C), we set the savings of each household equal to zero at the time of the appeal decision. This has little effect on willingness to pay for DI receipt since appellant households tend to have little savings, and so savings provides little self-insurance against disability in this population. By contrast, the possibility of reapplying has important implications for households' valuations of an initial DI allowance. When we impose the constraint in panel D that denied appellants cannot reapply for DI benefits, we find that households would be willing to pay far more for an *initial* allowance on appeal: 6.7 times as much among married households and 1.7 times as much among unmarried households.[36] This substantial increase in willingness to pay underscores that a key mechanism that insures households against the financial costs of an initial unsuccessful appeal is the opportunity to reapply for benefits—where more than half of initially appellants receive a DI allowance within four years (Table 4 panel A). Stated differently, appellant households are willing to pay far less for an initial successful appeal than they would be willing to pay to ultimately receive DI benefits. By the same logic, the marginal fiscal cost

[36]Formally, we set the probability of being allowed DI upon reapplication equal to zero—so appellants who are denied at the appeal never reapply—and then compare appellants' willingness to pay in the constrained and unconstrained settings.

of granting one additional DI allowance is far higher than the marginal fiscal cost of granting an *initial* appeal since the majority of initially denied appellants will be subsequently granted benefits.

Table 11: **Counterfactual analyses**

	Married	**Single and Unmarried**
Panel A.	**Baseline**	
Willingness to pay ($1,000)	2.300	11.316
p-value for $H_0 : WTP = 0$	0.700	0.000
Panel B.	**Constraining Spousal Labor Supply**	
Willingness to pay ($1,000)	9.852	
p-value for $H_0 : WTP = 0$	0.000	
p-value for H_0 : $WTP(Baseline) = WTP(Counterfactual)$	0.000	
Panel C.	**No Initial Savings Available**	
Willingness to pay ($1,000)	3.319	13.740
p-value for $H_0 : WTP = 0$	0.550	0.000
p-value for H_0 : $WTP(Baseline) = WTP(Counterfactual)$	0.000	0.000
Panel D.	**No Reapplication Available**	
Willingness to pay ($1,000)	15.506	19.490
p-value for $H_0 : WTP = 0$	0.000	0.050
p-value for H_0 : $WTP(Baseline) = WTP(Counterfactual)$	0.000	0.200

Note: This table shows estimates of the average welfare benefit ($1,000, per household member, annuitized over the four years after initial DI allowance) of DI allowance at appeal for married households and single and unmarried households. In the rows titled "Baseline", we use the estimated model to compute the welfare benefit of DI receipt. In the row titled, "Constrained spousal labor," we compute the willingness to pay for DI receipt while constraining the spousal labor supply to the observed labor supply during the year before DI allowance is announced. In the rows titled "No reapplication," we compute the willingness to pay for DI receipt while constraining denied appellants from reapplying for benefits by setting the probability of transitioning into DI equal to zero. The hypotheses tests $H_0 : WTP(Baseline) = WTP(Counterfactual)$ correspond to testing whether the average willingness to pay in the baseline (unconstrainted) model equals the average willingness to pay in the counterfactual. P-values are based on reestimating the model on 20 block bootstrap replicates of the data (where the block corresponds to the individual).

Robustness to relaxing the borrowing constraint

Following Low & Pistaferri (2015), we have so far assumed that households can save but not borrow—meaning that they cannot have negative net wealth. To examine the sensitivity of our findings to this assumption, we relax the borrowing constraint when computing willingness to pay for DI.

This is done by simulating optimal consumption and employment profiles with a modified budget constraint that allows households to borrow up to \$5,000. We find that the mean WTP for single households falls from about \$11,300 to about \$10,400, while the mean WTP for married households is unaffected at around \$2,300.

Why does allowing for borrowing have such a small effect on our WTP estimates? Three factors disincentivize borrowing in our setting. First, DI appellant households tend to be relatively old and do not anticipate high future income against which to borrow. Indeed, income after retirement is often less than working income, so households will tend to save rather than borrow so as to maintain consumption in retirement. Second, there are several sources of uncertainty in our model, including the risk of lost wages and the risk of DI denial. This uncertainty disincentivizes borrowing, instead leading households to desire a buffer stock of savings. Third, while borrowing may help households smooth the effects of modest, transitory shocks, DI denials typically generate relatively large and persistent changes in income. As a consequence, forward–looking households will not seek to smooth these income changes through borrowing.

9 Conclusion

While a mature literature finds that DI receipt discourages work, the welfare implications of this finding depend on two rarely studied economic quantities: the value that individuals and families place on receipt of disability benefits; and the full cost of DI allowances to taxpayers, summing over DI transfer payments, benefit substitution to or from other transfer programs, and induced changes in tax receipts. We assess these missing margins in the context of Norway's DI system, drawing on two strengths of the Norwegian environment: Norwegian register data, which allow us to characterize the household impacts and fiscal costs of disability receipt by linking employment, taxation, benefits receipt, and assets at the person and household level; and random assignment of DI applicants to Norwegian judges who differ systematically in their leniency, allowing us to recover the causal effects of DI allowance on individuals at the margin of program entry.

Accounting for the total effect of DI allowances on both household labor supply and net payments across all public transfer programs substantially alters our picture of the consumption benefits and fiscal costs of disability receipt. While DI denial causes a significant drop in household income and consumption on average, it has little impact on income or consumption of married appellants; spousal earnings and benefit substitution counteract the effect of denial of DI benefit payments on household income.

To explore the welfare implications of these findings, we estimate a dynamic model of household behavior that translates employment, reapplication and savings decisions into revealed preferences for leisure and consumption. We use the estimated model to compute the welfare benefits of DI receipt, and to perform counterfactual exercises that help us infer the extent to which spousal labor supply, savings, and reapplication attenuate the welfare loss from DI denial at appeal. The model estimates suggest that the cash-equivalent value of DI benefits is positive and sizable: on average, each net dollar in public expenditure induced by a DI allowance raises the (money metric) welfare of single and unmarried awardees by nearly \$0.92 and of married households by \$0.58. The value of DI receipt is smaller for married than unmarried DI appellants in substantial part because spousal labor supply of married DI appellants strongly buffers the household income and consumption consequences of DI allowance or denial.

When considering the interpretation and generality of our study, we emphasize four caveats. First, our structural model permits us to estimate the economic value of the transfer component of DI benefits—that is, the cash equivalent value of a DI award—but does not encompass the ex ante insurance value of the DI system for potential applicants. Since this insurance value is doubtless positive and potentially large, our estimates should not be interpreted as a full accounting of the welfare value of the DI system.

Second, in considering the implications of our findings for the U.S. Social Security Disability Insurance system (SSDI), it is worth noting that the SSDI program features a lower income replacement rate than the Norwegian system, and hence allowances and denials might be expected to have less pronounced effects on spousal labor supply. Conversely, the cash and in-kind transfers available to non-SSDI households in the U.S. are surely less comprehensive than in Norway, so the marginal impact of a DI allowance on individual and household consumption may be as large or larger for U.S. than Norwegian households, despite the lesser generosity of the U.S. program.

Third, we emphasize that the estimates we obtain from quasi-experimental variation in judicial disability determinations correspond to the average effect of DI allowance for individuals who could potentially have received a different allowance decision in the appeal process had their case been assigned to a different judge. Since the work capacity of individuals at the margin of program entry is likely to differ from that of inframarginal individuals, one must be cautious in extrapolating the causal estimates obtained here to the broader population at large as well as other programmatic settings. Nevertheless, the economic consequences of DI receipt for marginal DI claimants are relevant for policy. In both Norway and the U.S., the rise in DI rolls in recent decades is driven in significant part by de jure or de facto changes in the screening criteria applied to claimants reporting difficult-to-verify disorders, such as back pain or mental disorders (Autor & Duggan, 2006; Kostol & Mogstad, 2014). Logically, reforms aimed at altering DI screening criteria are likely to have the largest impacts on applicants on the margin of program entry, a substantial share of whom are applicants with difficult-to-verify disorders. These observations suggest that while the estimates provided by this paper are not directly generalizable to the full DI population, they are likely to be informative for policymaking.

Fourth, our structural model makes several strong assumptions. For example, we ignore that many households allocate a considerable portion of their budgets to consumption items that are not easily adjustable (e.g. durable goods and housing). The presence of such so-called consumption commitments is likely to re-enforce spousal labor supply responses to temporary or moderate income cuts to the other spouse (see e.g. Chetty & Szeidl, 2007). In particular, by increasing labor supply after a negative income shock, the household can maintain current consumption and avoid a costly adjustment to a shock which may prove transitory. If, however, shocks are large and persistent—as may be true for DI denials—then consumption commitments will matter less for spousal labor supply response since shocks of this type will induce households to optimally abandon their previously committed expenditures. Note finally that the presence of consumption commitments should lead to *larger* spousal labor supply responses in the short than long run, since these commitments can normally be unwound over time. In contrast, we find that spousal labor supply responses to DI denials build over time, suggesting that the adjustment cost of adjusting hours, accumulating skills, or seeking new employment may play a more important role than consumption commitments in determining the trajectory of spousal labor supply.

References

Administration, U.S. Social Security. 2017. *Annual Report of the Supplemental Security Income Program*. Tech. rept.

Aguiar, Mark, & Hurst, Erik. 2013. Deconstructing Life Cycle Expenditure. *Journal of Political Economy*, **121**(3), 437 – 492.

Altonji, Joseph G, & Segal, Lewis M. 1996. Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics*, **14**(3), 353–366.

Atkinson, Anthony B., Rainwater, Lee, & Smeeding, Timothy M. 1995. Income Distributions in OECD Countries: Evidence from the Luxembourg Income Study. *OECD Publications and Information Center*.

Autor, D. G., Maestas, N., Mullen, K., & Strand, A. 2017. Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability. *NBER Working Paper No. 23466*.

Autor, D. H. 2011. The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options. *NBER Working Paper*.

Autor, D. H., & Duggan, M. G. 2003. The Rise in the Disability Rolls and the Decline in Unemployment. *Quarterly Journal of Economics*, **118**(1), 157–205.

Autor, D. H., & Duggan, M. G. 2006. The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding. *Journal of Economic Perspectives*, **20**(3), 71–96.

Autor, D. H., Duggan, M. G., & Gruber, J. 2014. Moral Hazard and Claims Deterrence in Private Disability Insurance. *American Economic Journal: Applied Economics*, **6**(4), 110–141.

Baicker, Katherine, Finkelstein, Amy, Song, Jae, & Taubman, Sarah. 2014. The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment. *American Economic Review*, **104**(5), 322–328.

Blundell, R., Pistaferri, L., & Preston, I. 2008. Consumption Inequality and Partial Insurance. *The American Economic Review*, **98**(5), pp. 1887–1921.

Blundell, R., Costa Dias, M., Meghir, C., & Shaw, J. 2016a. Female Labor Supply, Human Capital and Welfare Reform. *Econometrica*, **85**(5), 1701–1753.

Blundell, Richard, Browning, Martin, & Meghir, Costas. 1994. Consumer Demand and the Life-Cycle Allocation of Household Expenditures. *The Review of Economic Studies*, **61**(1), pp. 57–80.

Blundell, Richard, Pistaferri, Luigi, & Saporta-Eksten, Itay. 2016b. Consumption inequality and family labor supply. *The American Economic Review*, **106**(2), 387–435.

Board, Social Security Advisory. 2015 (September). *Technical Panel on Assumptions and Methods*. Tech. rept.

Borghans, L, Gielen, A., & Luttmer, E. 2014. Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform. *American Economic Journal-Economic Policy*, **6**(4), 34–70.

Bound, J. 1989. The Health and Earnings of Rejected Disability Insurance Applicants. *American Economic Review*, **79**(3), 482–503.

Browning, M., & Leth-Petersen, S. 2003. Imputing consumption from income and wealth information. *The Economic Journal*, **113**(488), F282–F301.

Browning, Martin, & Meghir, Costas. 1991. The Effects of Male and Female Labor Supply on Commodity Demands. *Econometrica*, **59**(4), pp. 925–951.

Chen, S., & van der Klaauw, W. 2008. The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics*, **142**(2), 757–784.

Chetty, Raj, & Szeidl, Adam. 2007. Consumption commitments and risk preferences. *The Quarterly Journal of Economics*, **122**(2), 831–877.

Cullen, J., & Gruber, J. 2000. Does Unemployment Insurance Crowd out Spousal Labor Supply? *Journal of Labor Economics*, **18**(3), pp. 546–572.

Dahl, G., Kostøl, A., & Mogstad, M. 2014. Family Welfare Cultures. *The Quarterly Journal of Economics*, **129**(4), 1711–1752.

Eika, L., Mogstad, M., & Vestad, O. 2017. What can we learn about household consumption from information on income and wealth? *Mimeo*.

Fadlon, I., & Nielsen, T. H. 2015. Household Responses To Severe Health Shocks And The Design of Social Insurance. *Mimeo*.

Finkelstein, A., Hendren, N., & Luttmer, E. 2015. The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment. *NBER Working Paper No. 21308*, **June**.

French, E., & Song, J. 2013. The Effect of Disability Insurance Receipt on Labor Supply, Appeals and Re-Applications: Evidence from Hearing Office Level Decisions. *American Economic Journal: Applied*, **6**(2), 291–337.

Gruber, J. 2000. Disability insurance benefits and labor supply. *Journal of Political Economy*, **108**(6), 1162–1183.

Heathcote, Jonathan, Storesletten, Kjetil, & Violante, Giovanni L. 2014. Consumption and labor supply with partial insurance: An analytical framework. *The American Economic Review*, **104**(7), 2075–2126.

Imbens, Guido W., & Rubin, Donald B. 1997. Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *The Review of Economic Studies*, **64**(4), 555–574.

Keane, Michael P. 2011. Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, **49**(4), 961–1075.

Kostol, A., & Mogstad, M. 2014. How Financial Incentives Induce Disability Insurance Recipients to Return to Work. *American Economic Review*, **104**(2), 624–55.

Kostol, Andreas, & Mogstad, Magne. 2015. Earnings, Disposable Income, and Consumption of Allowed and Rejected Disability Insurance Applicants. *American Economic Review*, **105**(5), 137–41.

Liebman, Jeffrey B. 2015. Understanding the Increase in Disability Insurance Benefit Receipt in the United States. *Journal of Economic Perspectives*, **29**(2).

Low, Hamish, & Pistaferri, Luigi. 2015. Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off. *The American Economic Review*, **105**(10), 2986–3029.

Lundberg, S. 1985. The added worker effect. *Journal of Labor Economics*, 11–37.

Maestas, N., Mullen, K. J., & Strand, A. 2013. Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economic Review*, **103**(5), 1797–1829.

Meghir, C., & Pistaferri, L. 2011. Earnings, Consumption and Life Cycle Choices. Handbook of Labor Economics, vol. 4, Part B. Elsevier.

Meyer, B. D., & Mok, W. 2013. Disability, Earnings, Income and Consumption. *NBER Working Paper*, March.

Milligan, K. S., & Wise, D. A. 2011. Social Security and Retirement around the World: Mortality, Health, Employment and Disability Insurance Participation and Reforms. *Harvard Kennedy School Working paper.*

OECD. 2010. *Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries.* OECD Publishing.

OECD. 2015. *OECD Social Expenditure Database (SOCX).*

of Trustees of the Federal Hospital Insurance, Board, & Funds, Federal Supplememntary Medical Insurance Trust. 2017. *2017 Medicare Trustees Report.* Tech. rept.

of Trustees of the Federal OASDI Trust Funds, Board. 2017. *The 2017 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* Tech. rept. U.S. Social Security Administration.

Parsons, Donald O. 1980. The Decline in Male Labor Force Participation. *The Journal of Political Economy*, **88**(1), 117–134.

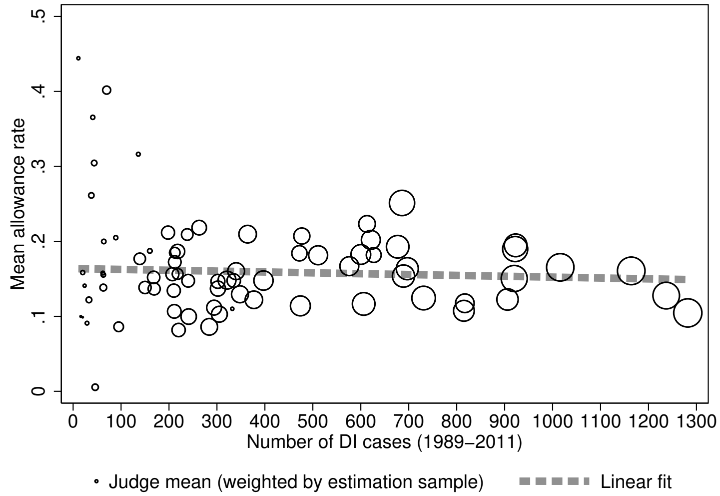
Persson, Petra. 2015 (Mar.). *Social insurance and the marriage market.* Working Paper Series 2015:6. IFAU - Institute for Evaluation of Labour Market and Education Policy.

Pistaferri, L. 2015. Household consumption: Research questions, measurement issues, and data collection. *Mimeo.*

U.S. Social Security Administration. 2017. *Annual Statistical Report on the Social Security Disability Insurance Program, 2016.* Office of Retirement and Disability Policy.

A Appendix:Additional Tables and Figures

Figure A.1: **Judge Leniency versus Number of Cases Handled**

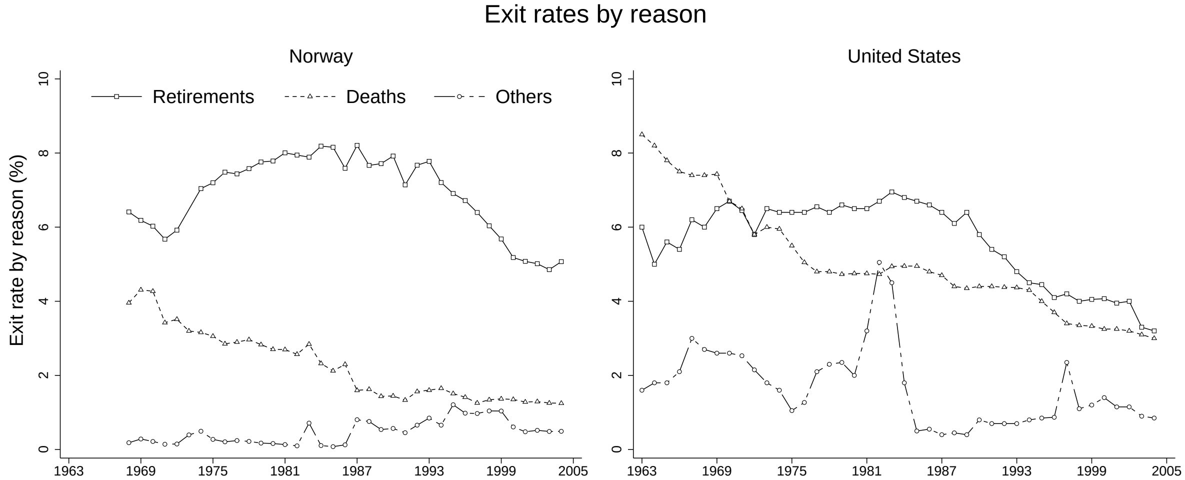


Notes: The figure plots judge allowance rate against the total number of cases handled by each judge. There are 75 unique judges, and on average, each judge has handled a total of 325 cases. Allowance rates are normalized by subtracting off year × department deviations from the overall mean. Cases are restricted to claimants appealing their first denied case during the period 1994-2005. Dot size is proportional to the number of cases a judge handles in the estimation sample (which is weakly smaller than the number of cases they have ever handled, as plotted on the x-axis).

Figure A.2: **DI Awards and DI Exits in Norway and the U.S.**

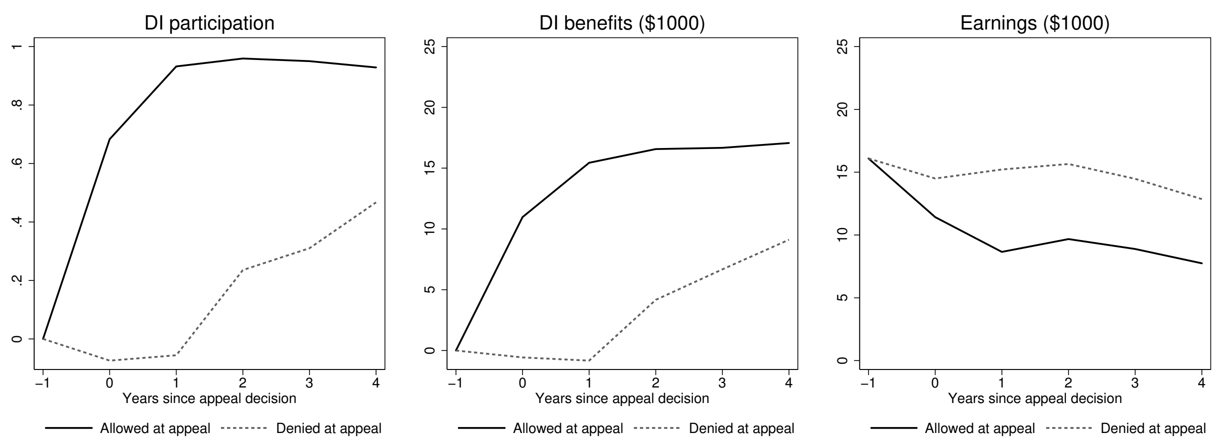
Notes: The U.S. trends are based on Autor & Duggan (2006), while the Norwegian trends are collected from various issues of the SSA Supplement. The graphs show award rates in the insured population and exit rates from the DI program in both countries.

Figure A.3: **DI Exits by Reason in Norway and the United States**



Notes: The U.S. trends are based on Autor & Duggan (2006), while the Norwegian trends are collected from various issues of the SSA Supplement. The graphs show exit rates because of death, retirement or other reasons (including eligibility-based exits).

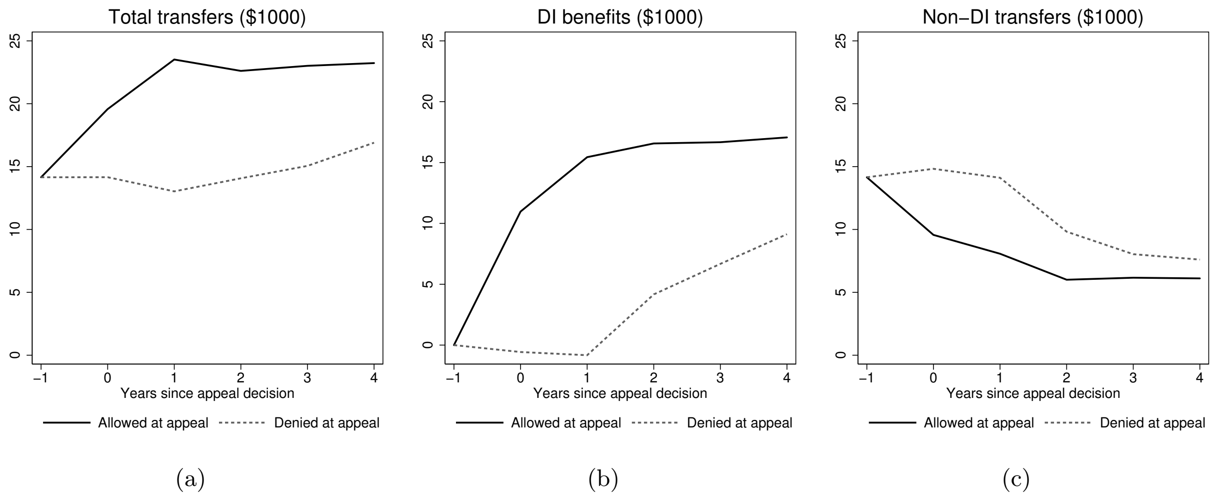
Figure A.4: **Potential Outcomes: Labor Earnings and DI benefits**



(a) (b) (c)

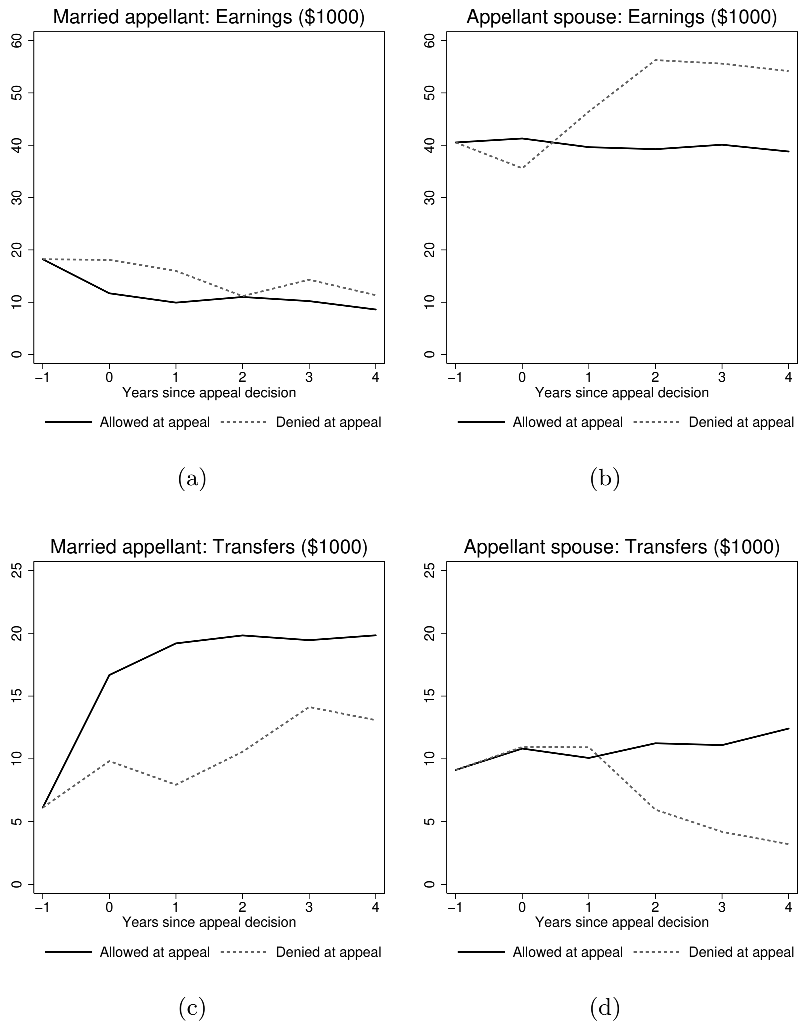
Notes: These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* 2014 for details).

Figure A.5: **Potential Outcomes: Benefit Substitution**



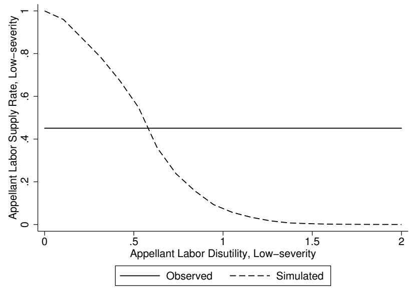
Notes: These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* 2014 for details).

Figure A.6: **Potential Outcomes: Married appellants and Spouses**

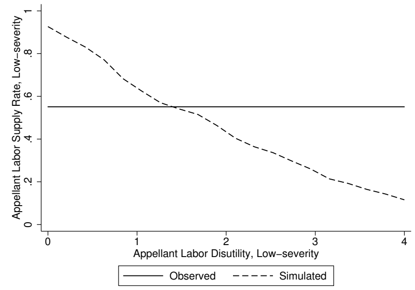


Notes: These figures display the decomposition of our LATE estimates into potential outcomes for allowed and denied complier appellants (see Dahl *et al.* (2014) for details).

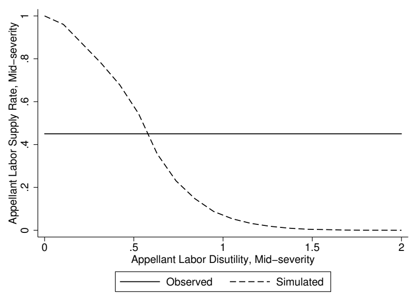
Figure A.7: **Local Identification of Disutility Parameters**



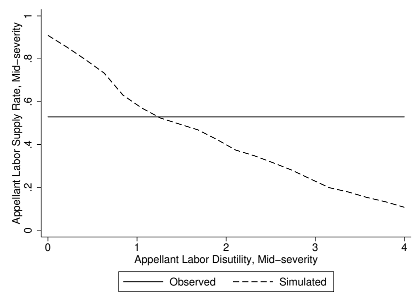
(a) Married: Labor disutility, low-severity



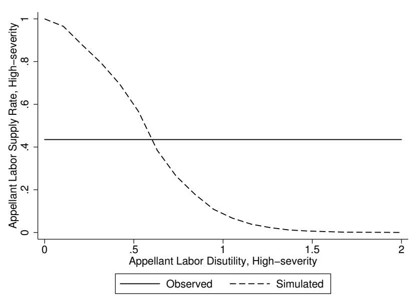
(b) Unmarried: Labor disutility, low-severity



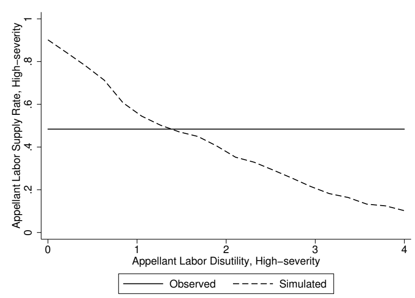
(c) Married: Labor disutility, mid-severity



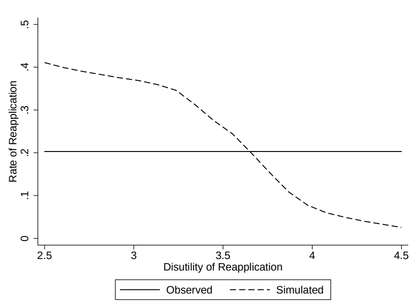
(d) Unmarried: Labor disutility, mid-severity



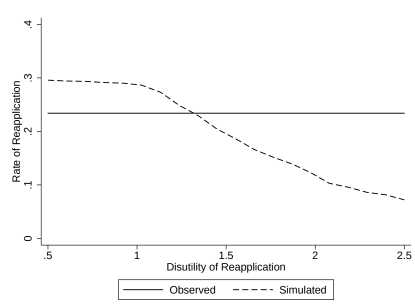
(e) Married: Labor disutility, high-severity



(f) Unmarried: Labor disutility, high-severity



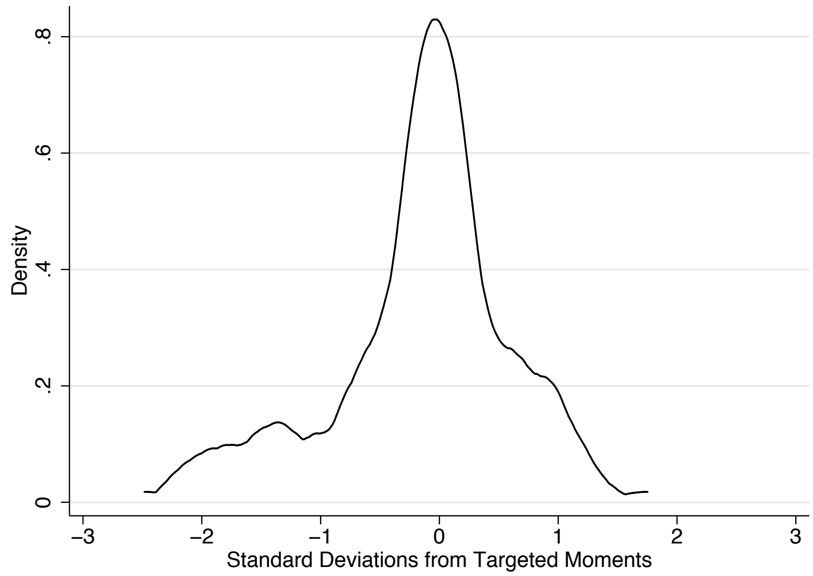
(g) Married: Reapplication disutility



(h) Unmarried: Reapplication disutility

Notes: These figures illustrate local identification of the six labor disutility parameters for appellants and two disutility parameters associated with reapplication. In each figure, the x-axis is the parameter representing the disutility. The y-axis is the corresponding moment, which is the average labor supply rate by severity type for labor disutility parameters and the average reapplication rate for reapplication disutility parameters. The solid line indicates the observed value in the data, while the dashed line indicates the value simulated from the estimated model, holding all other parameters fixed to their estimated values. The disutility is pinned down as the point on the x-axis that corresponds to the crossing of the solid and dashed lines.

Figure A.8: **Distribution of Standardized Deviations from Targeted Moments**



Notes: This figure displays the distribution of standardized deviations of each simulated moment from its corresponding observed moment for the full set of targeted moments in the method of simulated moments model estimation.

Table A.1: **Characteristics of DI recipients in Norway and the U.S.**

Characteristic	**Norway DI Recipients**	**U.S. SSDI Recipients**
Difficult to verify disorder	59.2 %	57.3 %
Age (at decision on initial application)	52.2	49.1
Prior earnings relative to the median	71.0 %	69.9 %

Notes: The U.S. numbers are from Maestas *et al.* (2013), and the Norwegian numbers are drawn from the sample of DI applicants during the years 2000-2003. Difficult to verify disorders include musculoskeletal and mental diagnoses. Prior earnings are measured in years three through five prior to application or appeal.

Table A.2: **Characteristics of DI Applicants and Appellants in Norway and the U.S.**

	Norway		**U.S.**	
Characteristic	**Applicants**	**Appellants**	**Applicants**	**Appellants**
Difficult to verify disorder	60.9%	69.7%	58.5%	62.2%
Age (at decision on initial application)	51.1	47.1	47.1	46.1
Prior earnings relative to the median	66.5%	50.4%	60.5%	56.3%

Notes: This table reports the key characteristics of DI applicants and appellants discussed in Section 2. The U.S. numbers are from Maestas *et al.* (2013), and the Norwegian numbers are drawn from the sample of DI applicants during the years 2000-2003. Difficult to verify disorder comprise musculoskeletal and mental diagnoses. Prior earnings are measured during years three through five prior to application or appeal.

Table A.3: **Sub-Sample First Stage Estimates**

	Baseline instrument		**Reverse-sample instrument**	
	(1)		(2)	
Dependent variable	$Pr(Allow)$		$Pr(Allow)$	
Younger appellants (age$\leq$ 48)	0.777	Dep. mean: 0.093	0.613	Dep. mean: 0.093
	(0.077)	N: 7,458	(.082)	N: 7,392
Older appellants (age >48)	0.838	Dep. mean: 0.165	0.838	Dep. mean: 0.165
	(0.106)	N: 6,634	(.124)	N: 6,563
Small households ($N \leq 3$)	0.921	Dep. mean: 0.140	0.710	Dep. mean: 0.139
	(0.091)	N: 9,532	(.168)	N: 9,329
Large households ($N > 3$)	0.589	Dep. mean: 0.100	0.474	Dep. mean: 0.099
	(0.097)	N: 4,560	(.090)	N: 4,522
Female appellants	0.837	Dep. mean: 0.134	0.606	Dep. mean: 0.133
	(0.083)	N: 8,851	(.078)	N: 8,700
Male appellants	0.774	Dep. mean: 0.115	0.668	Dep. mean: 0.115
	(0.115)	N: 5,241	(.139)	N: 5,184
Married appellants	0.830	Dep. mean: 0.133	0.666	Dep. mean: 0.133
	(.095)	N: 8,061	(.096)	N: 7,950
Unmarried and single appellants	0.775	Dep. mean: 0.119	0.685	Dep. mean: 0.118
	(0.097)	N: 6,031	(.091)	N: 5,978
Foreign born	0.425	Dep. mean: 0.091	0.373	Dep. mean: 0.090
	(.155)	N: 2,534	(.141)	N: 2,509
Less than high school degree	0.899	Dep. mean: 0.116	0.778	Dep. mean: 0.116
	(0.089)	N: 7,097	(.115)	N: 7,044
At least a high school degree	0.725	Dep. mean: 0.139	0.547	Dep. mean: 0.137
	(.092)	N: 6,995	(.100)	N: 6,897
At least one child below age 18	0.727	Dep. mean: 0.102	0.495	Dep. mean: 0.101
	(0.062)	N: 8,140	(.079)	N: 8,029
No children below age 18	0.927	Dep. mean: 0.162	0.976	Dep. mean: 0.161
	(0.105)	N: 5,952	(.127)	N: 5,888
Musculoskeletal disorders	0.823	Dep. mean: 0.118	0.732	Dep. mean: 0.118
	(0.112)	N: 6,149	(.119)	N: 6,102
Mental disorders	0.810	Dep. mean: 0.134	0.605	Dep. mean: 0.133
	(0.120)	N: 3,666	(.125)	N: 3,624
Circulatory system	0.754	Dep. mean: 0.150	0.829	Dep. mean: 0.150
	(0.367)	N: 512	(.347)	N: 510

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Notes: This table reports heterogeneity in first stage estimates using the baseline instrument (1) and the reverse-sample instrument (2). The first stage specification in (1) corresponds to panel B in Table 3. The reverse-sample instrument (2) is constructed by calculating judge leniency using all cases *except* for those in the specified subsample (e.g., judge leniency for the subsample of older applicants is constructed using judges' decisions for younger applicants). We exclude appellants whose judges handled fewer than ten cases in the reverse sample.

Table A.4: **Effect of DI Allowance on Earnings and Transfer Payments Among Married and Unmarried**

	Years after decision			
	1	**2**	**3**	**4**
Panel A.	**Married appellant labor earnings ($1,000)**			
Allowed DI	-5.042	-0.444	-4.426	-3.912
	(3.461)	(4.068)	(3.993)	(3.625)
Dependent mean	14.991	14.784	14.168	13.535
Panel B.	**Married appellant total transfers ($1,000)**			
Allowed DI	9.110**	6.499	5.008	5.395
	(4.000)	(4.423)	(3.703)	(3.628)
Dependent mean	16.621	17.356	17.919	18.508
Observations	7,844	7,740	7,648	7,548
Panel C.	**Unmarried appellant labor earnings ($1,000)**			
Allowed DI	-5.099	-10.939	-4.589	-6.475
	(7.402)	(6.932)	(7.018)	(5.686)
Dependent mean	13.279	13.646	13.34	12.883
Panel D.	**Unmarried appellant total transfers ($1,000)**			
Allowed DI	15.811***	14.466***	17.152***	10.714***
	(5.054)	(4.131)	(4.497)	(4.084)
Dependent mean	23.336	23.518	23.848	24.224
Observations	6,128	6,102	6,061	6,059

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Note: This table reports the impact of DI allowance on earnings and total transfers among married (panel A and B) and unmarried appellants (panel C and D). Baseline estimation sample consists of unmarried DI applicants who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and number of medical diagnoses. All control variables are measured prior to appeal.

Table A.5: **Specification Checks**

Dependent variable	**Died or migrated**	**Change in marital status**			**In restricted sample**
		Overall	Initially Unmarried	InitiallyMarried	
Judge leniency	0.017	-0.045	-0.023	-0.030	-0.015
	(0.048)	(0.058)	(0.074)	(0.071)	(0.020)
Dependent mean	0.092	0.141	0.166	0.109	0.981
Observations	14,359	14,092	6,031	8,061	14,092

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Note: This table reports the impact of judge leniency on the probability of death or migration, the probability of a change in marital status, or membership in the restricted sample. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). The second to fourth columns exclude those who die or migrate during the year of the appeal. The second column tests whether DI allowance affects the likelihood of a change in marital status (married to non-married or vice versa) for the baseline sample, and the third and fourth columns test whether DI allowance affected marriage entry and exit rates respectively. There are 75 unique judges. All regressions mirror the reduced form specification of Table 4.

Table A.6: **Effect of DI Allowance on Types of Transfer Payments of the Appellant**

	Years after decision				
	1	**2**	**3**	**4**	**Average**
Panel A.		**DI benefits ($1,000)**			
Allowed DI	16.240***	12.596***	10.203***	8.167***	11.883***
	(1.539)	(1.696)	(1.660)	(1.567)	(1.316)
Dependent mean	5.708	8.377	10.277	11.502	8.921
Panel B.		Total transfers ($1,000)			
Allowed DI	10.188***	8.807***	8.148***	6.429**	8.072***
	(2.736)	(2.749)	(2.433)	(2.683)	(2.499)
Dependent mean	19.567	20.072	20.54	21.053	20.305
Panel C.		**Non-DI transfers ($1,000)**			
Allowed DI	-6.308*	-3.744	-1.884	-1.611	-3.823*
	(3.273)	(2.656)	(2.062)	(2.525)	(2.298)
Dependent mean	14.009	11.839	10.398	9.666	11.521
Panel D.		**Social assistance ($1,000)**			
Allowed DI	-1.524	-1.169	-1.315*	-0.395	-0.964
	(1.123)	(1.031)	(0.783)	(0.677)	(0.778)
Dependent mean	2.852	2.182	1.78	1.464	2.103
Observations	13,972	13,842	13,709	13,607	13,972

***p<.01, **p<.05, *p<.10. Standard errors (in parentheses) are clustered at the judge level.

Note: This table reports instrumental variables estimates of the causal effect of receiving a DI allowance at the appeal stage on DI participation (panel A), annual DI benefits (panel B), and annual labor earnings (panel C), annual total transfers inclusive of DI benefits (panel D), and annual transfers excluding DI benefits (panel E). Columns 1-4 report separate estimates for each year, whereas column 5 reports estimates for the average outcome over the four year period. The baseline sample consists of individuals who appeal an initially denied DI claim during the period 1994-2005 (see Section 3 for further details). There are 75 unique judges. All regressions include fully interacted year and department dummies, dummy variables for month of appeal, county of residence, age at appeal, household size, gender, foreign born, marital status, children below age 18, education, and a number of medical diagnoses. All control variables are measured prior to appeal.

Table A.7: **Model Parameters: Log Earnings Regressions**

	Main Estimation			**Robustness to Control Function**		
	Married		**Single and Unmarried**	**Married**		**Single and Unmarried**
	Appellant	**Spouse**	**Appellant**	**Appellant**	**Spouse**	**Appellant**
Mid-severity of Disability	-0.015*	-0.000	-0.021***	-0.012	0.002	-0.021***
	(0.008)	(0.005)	(0.002)	(0.011)	(0.006)	(0.005)
High-severity of Disability	-0.031***	-0.032***	-0.041***	-0.045***	-0.033***	-0.027***
	(0.008)	(0.005)	(0.002)	(0.012)	(0.006)	(0.010)
High School	0.096***	0.072***	0.145***	0.180***	0.086***	0.100***
	(0.007)	(0.005)	(0.002)	(0.040)	(0.008)	(0.030)
Some College	0.249***	0.144***	0.251***	0.399***	0.167***	0.178***
	(0.010)	(0.007)	(0.002)	(0.071)	(0.013)	(0.048)
Old Age	-0.064***	-0.066***	0.117***	-0.193***	-0.105***	0.176***
	(0.007)	(0.005)	(0.002)	(0.061)	(0.019)	(0.038)
Pre-App. Earnings	0.847***	0.947***	1.003***	0.991***	0.993***	0.934***
	(0.014)	(0.009)	(0.003)	(0.069)	(0.023)	(0.046)
Pre-App. Earnings, Squared	0.112***	0.056***	-0.017***	-0.202	0.045***	0.027
	(0.018)	(0.009)	(0.004)	(0.147)	(0.011)	(0.030)
Pre-App. Earnings, Cubed	0.016	0.068***	-0.029***	0.046	0.085***	-0.004
	(0.029)	(0.013)	(0.006)	(0.038)	(0.015)	(0.023)
Constant	10.133***	10.773***	10.186***	9.790***	10.731***	10.375***
	(0.009)	(0.006)	(0.002)	(0.158)	(0.020)	(0.122)
Inverse Mills				0.547**	0.152**	-0.161
				(0.251)	(0.069)	(0.303)

Notes: This table presents estimated parameters from the post-application log earnings regression described in the text. "Pre-app. Earnings" refers to the average earnings constructed by forming the individual-specific average log earnings over the 10 years prior to application and residualizing on the other pre-application covariates.

Table A.8: **Model Parameters: Calibrations and Labor Disutility Estimates**

	Married	**Single and Unmarried**
Panel A.	**Externally set parameters**	
Interest rate:	0.016	0.016
Discount rate:	0.976	0.976
Coefficient of relative risk aversion:	1.5	1.5
Panel B.	**Estimated disutility parameters**	
Labor Disutility		
Low-severity:	0.702	0.485
	(0.014)	(0.010)
Mid-severity:	0.781	0.877
	(0.033)	(0.011)
High-severity:	0.829	1.076
	(0.007)	(0.013)
Spouse:	0.792	
	(0.008)	
Reapplication Disutility:	1.392	2.612
	(0.008)	(0.020)

Notes: This table summarizes the calibrated model parameters described in the text as well as the estimated model parameters representing disutility. Inference is based on reestimating the model on 20 block bootstrap replicates of the data (where the block corresponds to the individual).

Table A.9: **Simulated Labor Supply Elasticities**

	Main Estimation		**Robustness to Control Function**	
	Own-wage	**Cross-wage**	**Own-wage**	**Cross-wage**
Single Appellant:	0.201		0.184	
Married Appellant:	0.349	-0.331	0.324	-0.299
Spouse:	0.358	-0.345	0.309	-0.319

Notes: This table compares labor supply elasticities for married households and single and unmarried households by severity of disability. Because few appellants allowed DI are working, we consider in the immediately following year the subsample initially denied DI. These are Marshallian elasticities, that is, labor supply responses to permanent wage shocks. We compute the elasticity using the finite-difference evaluated at a one standard deviation permanent shock to the log wage.

Table A.10: **Robustness of Willingness to Pay**

	Married			Single and Unmarried		
	Control function:		Test of equality:	Control function:		Test of equality:
	With	Without	P-value	With	Without	P-value
Panel A.	**Baseline: Average Willingness to Pay**					
Average	2.327	2.300	0.750	10.816	11.316	0.999
Panel B.	**Counterfactuals: Average Willingness to Pay**					
Constraining Spousal Labor Supply:	10.036	9.852	0.100			
No Initial Savings Available:	3.349	3.319	0.750	13.247	13.740	0.999
No Reapplication Available:	15.589	15.506	0.200	19.085	19.490	0.999

Note: This table shows estimates of the average welfare benefit ($1,000, per household member, annuitized over the four years after initial DI allowance) of DI allowance at appeal for married households and single and unmarried households. In the rows titled "Unconstrained", we use the estimated model to compute the welfare benefit of DI receipt. In the row titled, "Constrained spousal labor," we compute the willingness to pay for DI receipt while constraining the spousal labor supply to the observed labor supply during the year before DI allowance is announced. In the rows titled "No reapplication," we compute the willingness to pay for DI receipt while constraining denied appellants from reapplying for benefits by setting the probability of transitioning into DI equal to zero. The hypotheses tests correspond to testing equality in the average willingness with and without correcting for selectivity bias in the estimation of the earnings processes. P-values are based on reestimating the model on 20 block bootstrap replicates of the data (where the block corresponds to the individual).

B Appendix: Estimation and Computation Details

B.1 Solving the Model, given a Discrete State Space

Here, we detail the algorithm used to compute the value function at each time period for each type of household, given a discrete state space. In particular, we present the algorithm for value functions after retirement (both single and married households), before retirement with DI (for single households only), and before retirement without DI (for single households only). The algorithms before retirement for married households are identical to those for single households, except that there is an additional choice (spousal labor supply) and an additional source of uncertainty that must be integrated out (spousal wage shocks), so we omit the algorithms for married households for brevity. The algorithms rely on a discretized state space in the continuous state variables, savings S_t and log wages $\log W_t$; denote the associated grids by $\mathcal{S}$ for savings and $\mathcal{W}$ for log wages.

B.1.1 Solution Algorithm after Retirement

We begin the solution method by solving for the value function at the years during retirement, which is simpler than the working age model because it does not involve labor supply (and the associated wage uncertain) or disability insurance (and the associated reapplication process).

1. Year of death ($t = T + 10$, where T is retirement year). In the final year of life, the household optimally consumes all remaining savings S_{T+10} plus retirement benefits b, so consumption

is optimally given by $C_{T+10} = S_{T+10} + b$. The value function is then $V_{M,T+10}(S_{T+10}) = \frac{1}{1-\mu_M}(S_{T+10}+b)^{1-\mu_M}$, so given parameters (b, μ_M), $V_{M,T+10}$ is known for all S_{T+10}. We evaluate $V_{M,T+10}$ for all $S_{T+10} \in \mathcal{S}$.

2. Year prior to death ($t = T + 9$): The value function is simply, $V_{M,T+9}(S_{T+9}) = \max_{C_{T+9},S_{T+10}} \frac{1}{1-\mu_M}(C_{T+9})^{1-\mu_M} + \zeta V_{T+10}(S_{T+10})$ subject to the budget constraint $S_{T+10} \leq (1+r)(S_{T+9} + b - C_{T+9})$. Given (b, μ_M, r, ζ) and $V_{M,T+10}$ from 1., we find $S_{T+10} \in \mathcal{S}$ such that $V_{M,T+9}$ is maximized, ruling out those S_{T+10} that do not satisfy the budget constraint. This is done for each $S_{T+9} \in \mathcal{S}$.

3. We then repeat the procedure in 2. for $t = T + 8, t = T + 7, \ldots, t = T + 1$. Note also that the terminal condition is always satisfied, since all remaining savings are consumed at $t = T + 10$ and $S_{T+11} = 0$. Lifetime utility would be infinitely negative if $C_{T+10} = S_{T+10} + b$ were negative with positive probability, so households will use precautionary savings across the lifecycle to ensure that $S_{T+10} \geq -b$.

This procedure yields $V_{M,t}(S_t)$ for each $t = T + 1, \ldots, T + 10$, for each S_t in the grid of possible values, given only the parameters (b, μ_M, r, ζ).

B.1.2 Solution Algorithm before Retirement for a Single Household with Disability Insurance

A single household ($M = 1$) with DI ($D_t = 1$) chooses appellant labor supply ($P_{A,t}$) and savings, which also determines consumption through the budget constraint. This is only more complicated than the retired household's problem due to the need to account for the labor supply decision and associated wage uncertainty, as well as the tax-transfer system that maps earnings, DI status, and household characteristics into disposable income.

1. One year prior to retirement ($t = T$): The household's problem is, $V_{1,T}(D_t = 1, \log W_{A,T}, S_T; O_1) = \max_{P_{A,T},C_T,S_{T+1}} \frac{1}{1-\mu_1}(C_T \exp(-\phi_{1,A,H}P_{A,T}))^{1-\mu_1} + \zeta V_{T+1}(S_{T+1})$ subject to the budget constraint $S_{T+1} \leq (1+r)(S_T + I_T - C_T)$. Note that we express the value function in terms of the observed log wage $\log W_{A,t}$ rather than the shock $\tau_{A,t}$, as each is known from the other given O_1 and t (see the wage equation in the main text). Furthermore, disposable income is determined by $I_T = (1 - \Lambda_{1,1,K,T})(E_{A,t})^{(1-\Psi_{1,1,K,T})}$ if $P_{A,t} = 1$ and $I_T = \Phi_{1,1,K,T}$ if $P_{A,T} = 0$, where $E_T = W_{A,T}P_{A,T}$ is earnings. Recall that $V_{M,T+1}$ is known from the retirement solution. Given $V_{1,T+1}(S_{T+1})$, we evaluate the objective $\frac{1}{1-\mu_1}(C_T \exp(-\phi_{1,A,H}P_{A,T}))^{1-\mu_1} + \zeta V_{T+1}(S_{T+1})$ for each $P_{A,T} \in \{0, 1\}$, $S_{T+1} \in \mathcal{S}$, choosing the objective-maximizing combination as the optimal household solution, which yields $V_{1,T}(D_t = 1, \log W_{A,T}, S_T; O_1)$, given each state space combination $S_T \in \mathcal{S}$, $\log W_T \in \mathcal{W}$, O_1.

2. Two years prior to retirement ($t = T-1$): The household's problem is $V_{1,T-1}(D_{T-1} = 1, \log W_{A,T-1}, S_{T-1}; O_1)$ $\max_{P_{A,T-1},C_{T-1},S_T} \frac{1}{1-\mu_1}(C_{T-1}\exp(-\phi_{1,A,H}P_{A,T-1}))^{1-\mu_1} + \zeta \mathbb{E} V_{1,T}(D_T = 1, \cdot, S_T; O_1)$. Since we know from 1. how to compute $V_{1,T}(D_t = 1, \log W_{A,t}, S_T; O_1)$ for each $\log W_{A,T} \in \mathcal{W}$, we can

integrate across the distribution of $\log W_{A,T}$ to compute the expectation $\mathbb{E}V_{1,T}(D_T = 1, \cdot, S_T; O_1)$. In particular, since we have assumed $\log W_{A,t}$ follows a random walk process, then $\log W_{A,T}$ is Normally distributed with mean $\log W_{A,T-1}$, so we can use Gaussian quadrature to approximate the integral numerically. Given $\mathbb{E}V_{1,T}(D_T = 1, \cdot, S_T; O_1)$, we evaluate the objective $\frac{1}{1-\mu_1}(C_{T-1}\exp(-\phi_{1,A,H}P_{A,T-1}))^{1-\mu_1} + \zeta\mathbb{E}V_{1,T}(D_T = 1, \cdot, S_T; O_1)$ for each $P_{A,T-1} \in \{0,1\}$, $S_T \in \mathcal{S}$, choosing the objective-maximizing combination as the optimal household solution, which yields $V_{1,T-1}(D_{T-1} = 1, \log W_{A,T-1}, S_{T-1}; O_1)$, given each state space combination $S_{T-1} \in \mathcal{S}$, $\log W_{T-1} \in \mathcal{W}$, O_1.

3. We then repeat the procedure in 2. for $t = T-2, t = T-3, \ldots, t = 1$. Recall that $T = 27$ for single households that are young at the time of appeal and $T = 11$ for households that are old at the time of appeal, so we must compute the algorithm separately for young and old households.

This procedure yields $V_{1,t}(D_t = 1, \tau_{A,t}, S_t; O_1)$ for each $t = 1, \ldots, T$, for each discretized $(S_t, \log W_{A,t}, O_1)$ combination, given the model parameters.

B.1.3 Solution Algorithm before Retirement for a Single Household without Disability Insurance

A single household without DI chooses appellant labor supply, reapplication (R_t), and savings, which also determines consumption through the budget constraint. This is only more complicated than the solution algorithm with DI because the household must choose DI reapplication and we must account for the probability of receiving DI approval in the next period.

1. Year prior to retirement ($t = T$): The household's problem is, $V_{1,T}(D_t = 0, \log W_{A,T}, S_T; O_1) = \max_{P_{A,T}, C_T, R_T, S_{T+1}} \frac{1}{1-\mu_1}(C_T \exp(-\phi_{1,A,H}P_{A,T}) - R_T\exp(\omega_1))^{1-\mu_1} + \zeta V_{T+1}(S_{T+1})$ subject to the budget constraint $S_{T+1} \leq (1+r)(S_T + I_T - C_T)$, where earnings and disposable income are determined analogously to the case with DI. Recall that $V_{M,T+1}$ is known from the retirement solution, so for each O_1, we can directly compute the objective $\frac{1}{1-\mu_1}(C_T\exp(-\phi_{1,A,H}P_{A,T}))^{1-\mu_1} + \zeta V_{T+1}(S_{T+1})$ for each combination of $P_{A,T} \in \{0,1\}$, $R_T \in \{0,1\}$, $S_{T+1} \in \mathcal{S}$, $\log W_T \in \mathcal{W}$, choosing the maximizing combination as the optimal household solution, which yields $V_{1,T}(D_t = 1, \log W_{A,T}, S_T; O_1)$, given each state space combination $S_{T-1} \in \mathcal{S}$, $\log W_{T-1} \in \mathcal{W}$, O_1. Note that $R_T = 0$ is always optimal, since reapplication incurs a cost but no benefits are received due to retirement in the next period.

2. One year earlier ($t = T-1$): The household's problem is $V_{1,T-1}(D_{T-1} = 0, \log W_{A,T-1}, S_{T-1}; O_1) = \max_{P_{A,T-1}, R_{T-1}, C_{T-1}, S_T} \frac{1}{1-\mu_1}(C_{T-1}\exp(-\phi_{1,A,H}P_{A,T-1}) - R_{T-1}\exp(\omega_1))^{1-\mu_1} + \zeta\mathbb{E}V_{1,T}(\cdot, \cdot, S_T; O_1)$. Note that we can write, $\mathbb{E}V_{1,T}(\cdot, \cdot, S_T; O_1) = (1 - \pi_{1,H,T-1})\mathbb{E}V_{1,T}(D_T = 0, \cdot, S_T; O_1) + \pi_{1,H,T-1}\mathbb{E}V_{1,T}(D_T = 1, \cdot, S_T; O_1)$ using the DI approval rate $\pi_{1,H,T-1}$. Since we know from 1. how to compute $V_{1,T}(D_t = 0, \log W_{A,t}, S_T; O_1)$ for each $\log W_{A,T} \in \mathcal{W}$, we can integrate across the distribution of $\log W_{A,T}$ to compute the expectation $\mathbb{E}V_{1,T}(D_T = 0, \cdot, S_T; O_1)$. Given $\mathbb{E}V_{1,T}(D_T = 1, \cdot, S_T; O_1)$ and $\mathbb{E}V_{1,T}(D_T = 0, \cdot, S_T; O_1)$,

we evaluate the objective of $V_{1,T-1}(D_{T-1}=1, \log W_{A,T-1}, S_{T-1}; O_1)$ at each combination of $P_{A,T-1} \in \{0,1\}$, $R_{T-1} \in \{0,1\}$, $S_T \in \mathcal{S}$, choosing the objective-maximizing combination as the optimal household solution, which yields $V_{1,T-1}(D_t=1, \log W_{A,T-1}, S_{T-1}; O_1)$, given each state space combination $S_{T-1} \in \mathcal{S}$, $\log W_{T-1} \in \mathcal{W}$, O_1.

3. We then repeat the procedure in 2. for $t = T-2, t = T-3, \ldots, t = 1$. Again, we must compute the algorithm separately for young and old households.

This procedure yields $V_{1,t}(D_t=0, \tau_{A,t}, S_t; O_1)$ for each $t = 1, \ldots, T$, for each discretized $(S_t, \log W_{A,t}, O_1)$ combination, given the model parameters.

B.1.4 Feasible Discretization of the State Space

The algorithms described above rely on the discretized state spaces $\mathcal{S}$ and $\mathcal{W}$. We construct a grid in each using equally spaced quantiles of the observed marginal distributions of savings and wages, respectively. In practice, we use ten points to represent the state space for S_t and ten points to represent the state space for $\log W_t$. We investigate the robustness of the model results to this approximation for single households by allowing for 100 points in the state space for S_t and 100 points in the state space for $\log W_{A,t}$, so that there are 100 times as many grid points and the model requires approximately 100 times longer to compute. We find that the model fit is similar, suggesting that the model solution is not very sensitive to additional fineness of the grid. When integrating across the distribution of $\log W_{A,t}$, we construct the 10×10 transition matrix representing the probability of transitioning to any point on the grid from any other point on the grid using the conditional Normal probability distribution function evaluated at each point.

The implied computational burden is substantial. For single households before retirement without DI that are young at the time of appeal, one simulation of the model requires millions of distinct numerical evaluations, with even more for married households.[1] We must then also perform these evaluations for young single and married households with DI, old single and married households without DI, and old single and married households with DI, as well as compute the single and married household value functions after retirement. As discussed below, estimating the unknown model parameters will require that we repeat these solution algorithms thousands of times, while bootstrapping the estimates requires thousands more times.

One other point to note: This approach does not simulate or approximate the initial distribution of the state space. Instead, it uses the exact observed values. That is, for each household, our approach computes its predicted optimal choice variables conditional on its observed characteristics, preserving the exact initial distribution of observed characteristics.

[1] This is because we must evaluate the value function objective at (10 current savings state points) $\times$ (10 next period savings choice points) $\times$ (10 current appellant wage state points) $\times$ (2 appellant labor supply choice points) $\times$ (2 DI reapplication choice points) $\times$ (12 static household types) $\times$ (27 time periods), for a total of over a million distinct numerical evaluations. For married households, there are also 10 current spouse wage states and 2 spousal labor supply choice points to consider, for a total of over 20 million distinct numerical evaluations.

B.2 Interpolating the Discretized Model to a Continuous State Space and Simulating Sample Moments

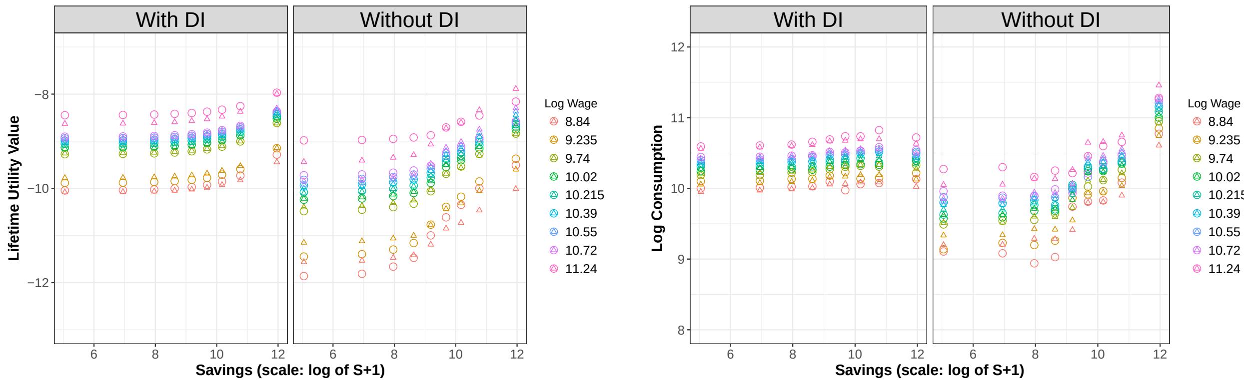
Our aim is to use the value functions to infer optimal choices of consumption, labor supply, and DI reapplication for the households in our sample, given the model parameters. The computational algorithms described above provide an approximate mapping from the state space to the value function for a discrete grid of points in the state space. For example, for single households, it provides a numerical solution to the value function, $V_{1,t}$, when given a time period t, the vector of static household characteristics O_1, and the current values of savings in $\mathcal{S}$ and log wages in $\mathcal{W}$. However, it does not provide the value function $V_{1,t}$ or associated optimal choices for $S_t \notin \mathcal{S}$ or $\log W_t \notin \mathcal{W}$, which is nearly the entire sample of households.

In order to approximate optimal household choices (i.e., consumption, labor supply, and DI reapplication) for all $(S_t, W_{A,t})$ pairs in the observed sample, we use interpolation. In particular, for each of the choice outcomes, interpolation uses the state space points for which we know the optimal choices, $\mathcal{S} \times \mathcal{W}$, to approximate the optimal choices for other state spaces. Our chosen interpolation method is cubic spline interpolation in the consumption choice and cubic spline interpolation of the underlying index functions for labor supply and DI reapplication choices. This is implemented using the "gam" function from the published R package *mgcv*, version 1.8-12. To allow full flexibility in the discrete components of the state space, we interpolate separately for each (observed time period $t = 1, 2, 3, 4$) $\times$ (current DI state 0 or 1) $\times$ (household static characteristic O_M) $\times$ (marital status M). This way, the interpolation method only requires smoothing approximations on the continuous components of the state space $(S_t, W_{A,t})$, but is unrestricted across discrete components of the state space.

For each individual in our sample at each time period, the interpolation fit on the discretized model provides a prediction of consumption, labor supply, and DI reapplication at times $t = 1, 2, 3, 4$. Using these predictions, we then construct the simulated moments from these predictions. For example, average spousal labor supply is simulated as the mean predicted labor supply of spouses provided by the interpolation.

Figure 1 (of this appendix) demonstrates the performance of the interpolation in out-of-grid prediction for the value function as well as the optimal choice of consumption. In particular, we first compute $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ for $t = 1$ on the grid $\mathcal{G} = \mathcal{S} \times \mathcal{W}$, separately for each of the 12 O_1 types. Second, we also compute $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ for an alternate grid $\mathcal{H}$, where $\mathcal{G} \cap \mathcal{H} = \emptyset$. We then interpolate $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ computed on $\mathcal{G}$ onto the points $\mathcal{H}$, which is an out-of-grid prediction. We choose the points in $\mathcal{H}$ to be particularly difficult to match by selecting the midpoints between any two grid points in $\mathcal{G}$ (midpoints maximize the distance between $\mathcal{H}$ points and $\mathcal{G}$ points within the same interval).

The figure shows the out-of-grid prediction of $\mathcal{H}$ as triangles, and the in-grid prediction of $\mathcal{H}$ from directly-computing $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ on $\mathcal{H}$ as circles, where each is formed from averaging across the 12 O_1 types. Visually, the goal of the interpolation is to approximate the circles with the triangles. This allows us to approximate the value function and optimal choices at thousands of



(a) Value Function (b) Consumption

Figure A.9: Interpolation Fit for Out-of-Grid State Space Values

Notes: This figure demonstrates the performance of the interpolation in out-of-grid prediction for the value function as well as the optimal choice of consumption. In particular, we first compute $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ for $t = 1$ on the grid $\mathcal{G} = \mathcal{S} \times \mathcal{W}$, separately for each of the 12 O_1 types. Second, we also compute $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ for an alternate grid $\mathcal{H}$, where $\mathcal{G} \cap \mathcal{H} = \emptyset$. We then interpolate $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ computed on $\mathcal{G}$ onto the points $\mathcal{H}$, which is an out-of-grid prediction. We choose the points in $\mathcal{H}$ to be particularly difficult to match by selecting the midpoints between any two grid points in $\mathcal{G}$ (midpoints maximize the distance between $\mathcal{H}$ points and $\mathcal{G}$ points within the same interval). The figure shows the out-of-grid prediction of $\mathcal{H}$ as triangles, and the in-grid prediction of $\mathcal{H}$ from directly-computing $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1\right)$ on $\mathcal{H}$ as circles, where each is formed from averaging across the 12 O_1 types.

out-of-grid points in the sample using only a small grid. We see that the out-of-grid interpolation predictions track the in-grid directly-computed circles. It performs especially well across the savings grid and at interior values of the log wage grid. The approximation is less precise at the end points of the log wages grid, but these points represent only a small sample of outlier observations in the data.

B.3 Solving for the Optimal Parameters

We choose two sets of moments to match. The first set consists of raw data moments, chosen based on the identification arguments in the text. These moments are mean log disposable income and expected log disposable income conditional on log earnings among households that supply labor, mean disposable income among households that do not supply labor, and employment rates and reapplication rates among those not receiving DI. Each of these moments is matched conditional on observable types over which the parameters vary in order to pin down all of the type-specific model parameters. The second set of moments is the IV results for consumption, disposable income, and earnings among appellants and spouses, included to discipline the model to recover our estimates of the causal effects of DI allowance.

To simulate the moments, we solve the value function, estimate the interpolation splines and predict the choice of each household, then compute the moment on the predicted household choices. We compare each simulated moment to the same moment computed on the observed household choices from the data. We form the objective function by forming the difference between the observed and simulated moment, and divide by the standard deviation corresponding to the observed moment. This weighting is equivalent to using the diagonal weighting matrix to form the objective function, as in Equation (13) of Blundell, et al. (2016) and motivated by the finding of Altonji and Segal (1996) that the asymptotically efficient weighting matrix has poor small-sample properties. We weight up the IV moments so that the sum of their weights is equal to that of the raw data moments.

We solve numerically for the parameters that minimize this objective function. For each vector of candidate parameters, we compute the value function on the discrete state space conditional on these parameters, interpolate to form the model prediction of optimal choices for each household, then evaluate the objective function for the simulated moments. To minimize the objective function, we apply two approaches. First, we use a particle swarm optimization algorithm to search for the globally optimal parameter vector, utilizing the "psoptim" function from the published R package *psoptim*, version 1.0. Second, we use the standard BFGS optimization algorithm, initialized at the optimal parameters found by psoptim, to verify that psoptim has found the locally optimal parameters. Together, these optimization algorithms require over a thousand complete evaluations of the model. We perform inference using the block bootstrap, where each bootstrap also requires estimation using these optimization algorithms. In particular, we randomly draw block replicates of the sample, where each "block" is a household's four-year history, then repeat the approach described above to find the optimal parameter vector for this replication sample. The distribution of each parameter across replication samples is then used to compute block bootstrap p-values for inference.

B.4 Extracting Willingness to Pay

Once the optimal parameter estimates are obtained, we can use the estimated model to perform counterfactual exercises. The counterfactual exercise of interest is to solve for the amount of income a household would be willing to give up each year across the remainder of the working life in order to be initially approved for DI. In particular, for single households, we parameterize this by modifying the budget constraint to include a *Cost* parameter: $S_{t+1} \leq (1+r)\left(S_t + I_t - Cost - C_t\right)$. Denote the value function with this budget constraint by $V_{1,t}\left(D_t, \log W_{A,t}, S_t; O_1, Cost\right)$. Then, the willingness to pay from time $t = 1$, denoted WTP, solves this equation:

$$V_{1,1}\left(D_1 = 0, \log W_{A,t}, S_t; O_1, Cost = 0\right) = V_{1,1}\left(D_1 = 1, \log W_{A,t}, S_t; O_1, Cost = WTP\right)$$

In words, WTP is the value of *Cost* that makes the household indifferent between being initially denied DI ($D_1 = 0$) but paying $Cost = 0$, and being initially approved DI ($D_1 = 1$) but paying $Cost = WTP$ annually.

In practice, we cannot solve the above equation exactly because we do not have a closed-form representation for $V_{1,1}$, so we instead use numerical optimization. For each household in our sample, we express WTP as the solution to a one-dimensional optimization problem in which we search for the value of *Cost* that minimizes the squared deviation from the above equality:

$$WTP = \arg\min_c \left\{V_{1,1}\left(D_1 = 0, \log W_{A,t}, S_t; O_1, Cost = 0\right) - V_{1,1}\left(D_1 = 1, \log W_{A,t}, S_t; O_1, Cost = c\right)\right\}^2$$

We then report the average WTP across the sample of single households, and perform the analogous procedure for married households.