Disability Benefits, Consumption Insurance, and Household Labor Supply[†]

By David Autor, Andreas Kostøl, Magne Mogstad, and Bradley Setzler*

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. (JEL D12, D14, H55, I38, J14, J22)

Over the past 50 years, disability insurance (DI) rolls have risen dramatically in many OECD countries. In the United States, Social Security Disability Insurance (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 (US Social Security Administration 2017c). In many European countries, the increases are even more striking, from 1 percent to 7 percent in the United Kingdom

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 United States, 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 gate-keeping 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 trade-off 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 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

---

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

[2] For example, the United States 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 United States, 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, Duggan, and Gruber 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.

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 9 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 and 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 and Mok (2019) and Kostøl and 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 US data to estimate labor supply impacts of DI receipt.[5] Exploiting variation in DI allowances stemming from differences in disability examiner leniency, Maestas, Mullen, and Strand (2013) and Autor et al. (2017) find that DI receipt substantially reduces earnings and employment of applicants. French and Song (2014) 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 and 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 Pistaferri 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 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

---

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

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

[6] This literature is reviewed by Blundell, Pistaferri, and Preston (2008); Meghir and Pistaferri (2011); and Blundell, Pistaferri, and Saporta-Eksten (2016).

to families. A notable exception is Blundell, Pistaferri, and Saporta-Eksten (2016), 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 matters less. In line with these results, Persson (forthcoming) finds that husbands increase their labor supply to offset household income losses following the elimination of survivors insurance for their wives, and Fadlon and Nielsen (2019) 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 and 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 I reviews the key features regarding the DI program in Norway, compares the system with the US system, and describes the research design. Section II describes the data and sample restrictions. Section III assesses the relevance and validity of our instrument. Section IV estimates the causal effect of DI allowance versus denial on applicant labor earnings and receipt of transfer income. Section V analyzes the household impact and fiscal costs of DI allowances. Section VI 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 VII 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.

I. 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 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.—We summarize the Norwegian DI program here and refer the reader to Section IIB of Dahl, Kostøl, and Mogstad (2014) for further

---

[7] A complementary exception is Finkelstein, Hendren, and Luttmer (2019), 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 (i) abstracts from labor supply considerations since labor supply appears unaffected by Medicaid provision in their setting (Baicker et al. 2014); and (ii) estimates both the transfer and ex ante insurance values of public benefits provision, whereas we estimate only the first component.

details. The Norwegian DI program provides partial earnings replacement to 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 depends upon a worker's earnings history, and the schedule is progressive, so that the replacement rate is higher for low-wage workers. 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.

Workers seeking DI benefits apply to the Norwegian Social Security Administration office. In the initial Disability Determination Stage (DDS) review, examiners check whether the applicant meets non-medical criteria, including age and prior employment requirements and, if so, use written medical evidence to evaluate the applicant's ability to work, accounting for health, age, education, work experience, and the skill transferability. Benefits are awarded to applicants assessed as unable to engage in any substantial gainful activity. Approximately three-quarters of applicants are awarded benefits at this stage, with roughly one-third of those awarded receiving partial awards. Denied applicants are often those claiming difficult to verify impairments, particularly back pain, as we discuss below.

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] In the case of appeal, ALJs are required to apply the same criteria used in the initial determination process, although 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 subsequently choose to file a new DI application. Seventy-five percent of denied appellants eventually reapply, with 65 percent of those ultimately allowed DI.[10]

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

---

[8] This definition is almost identical to the one used by the US 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. In our main analysis, we count remands, which account for only 5 percent of appeal outcomes, as rejections. Our results are unaffected if we instead code remands according to their ultimate disposition following reconsideration.

[10] If a case is denied at the ALJ level, it can also be appealed to the higher courts, but few applicants exercise this option.

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.[11]

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 on average 375 cases each. 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.)

The first column of Table 1 tests whether appellants' (predetermined) characteristics and economic conditions are predictive of case outcomes using a linear probability model. Demographic, economic, and health variables are highly predictive of whether an appealed case is allowed, as expected. Column 3 assesses whether these same case characteristics are predictive of the leniency of the judges to which cases are assigned and finds 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.

Our data do not offer insight into why some judges are more lenient than others.[12] What is critical for our analysis, however, 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 IIIA), and that cases allowed by a strict judge would also be allowed by a lenient one (consistent with the tests in Section IIIB).

Instrumental Variables Strategy.—We use variation in DI allowances induced by the random assignment of appellants to judges who differ in their leniency as an instrumental variable to estimate the economic consequences of disability receipt. Our baseline instrumental variables (IV) model is described by the following two-equation system:

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

$$(2) \qquad Y_{it} \, = \, \beta_t A_i + X_i' \theta_t + \eta_{it}.$$

---

[11] We verified these rules with the Head of the Administrative Law Court, Knut Brofoss. We have also verified our understanding with current judges and department heads.

[12] 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 online Appendix Figure A1).

TABLE 1—TESTING FOR RANDOM ASSIGNMENT OF CASES TO JUDGES

Dependent variable:	Case allowed		Judge leniency	
	Coefficient (1)	SE (2)	Coefficient (3)	SE (4)
Panel A. Predetermined 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)
Panel B. Predetermined 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]	[0.001]		[0.72]	
Observations	14,092		14,092	

Notes: Standard errors (in parentheses) are clustered at the judge level. This table reports an F-test of whether the hearing office complied with the random allocation procedure described in Section I. The baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994–2005 (see Section II 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 1 if a person has some college or has a college degree; body system codes are based on ICD-10 diagnostic codes. Predetermined 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.

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 (LATE), our baseline specification uses two-stage least squares (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 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.

II. Data and Background

A. *Data and Sample Restrictions*

Our analysis integrates data across multiple administrative registers to assess the impact of DI allowances on DI and other transfer benefits, labor income, tax payments, and consumption. Information on DI benefits is drawn from social security registers that contain complete records for all individuals who entered the DI program during 1967–2010. These data record each individual's work history and medical diagnosis, the month when DI was awarded or denied, and the level of DI benefits received. We link these data to hearing office records for 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 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, Rainwater, and Smeeding 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.[13] Finally, unique identifiers allow us to match spouses to one another and

---

[13] 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.

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 and Leth-Petersen (2003) show how one can construct such measures of consumption from longitudinal data on income and assets. Eika, Mogstad, and Vestad (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 datasets 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, Mogstad, and Vestad (2017) for more details.

Our empirical analysis studies DI applicants who appeal an initially denied DI claim.[14] Our estimation sample consists of individuals whose appeal decision was made during the period 1994–2005, which allows us to observe individuals for at least four years after the appeal decision. To reduce sampling variation in the instrumental variable, we follow Maestas, Mullen, and Strand (2013) and French and Song (2014) in excluding cases assigned to appeal judges who handled fewer than 10 cases during the 1989 through 2011 period.[15] 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. DI appellants are 20 percent more likely than the full set of DI applicants to claim musculoskeletal disorders (44 versus 37 percent), and only one-half as likely to claim circulatory system disorders (4 versus 8 percent).

B. *Institutional Background*

There are a number of similarities and some key differences between the DI systems in the United States and in Norway (see Autor and Duggan 2006, Kostøl and Mogstad 2014). DI is one of the largest transfer programs in both countries. The prevalence of DI receipt is considerably lower in the United States than in Norway, as shown in Figure 1, though both have grown five to ten times as the adult population has over the last five decades. From 1961 to 2012, DI prevalence increased

---

[14] Some individuals have several denied DI claims over the period we consider. We restrict our sample to each individual's first denied DI claim.

[15] Including these cases does not appreciably change the estimates, nor does excluding cases assigned to judges who handle fewer than 50 cases.

TABLE 2—DESCRIPTIVE STATISTICS OF APPLICANTS AND APPELLANTS

	DI applicants		DI appellants		Test of equal means
	Mean	SD	Mean	SD	*t*-statistic
Panel A. Predetermined 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
Panel B. Predetermined 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		

Notes: Standard deviations [in square brackets]. 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 II 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.

from 2.2 to 9.7 percent in Norway and from 0.8 to 5.0 percent in the United States.[16] In recent years, Norway's DI prevalence has leveled off at about 10 percent, while in the United States, SSDI prevalence rose steeply through 2013, after which time it peaked and reversed (Social Security Advisory Board 2015).[17]

In both countries, the expansion of the DI rolls 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.[18] Because these disorders have low associated

---

[16] Milligan and Wise (2011) discuss sources of differences in DI rates across countries, arguing that differences in underlying population health explain little of this variation.

[17] The US 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 US adults than does DI alone. However, the US DI program is more comparable to the Norwegian DI program than is the United States SSI program since SSI primarily provides benefits to adults with little work history. In this sense, SSI is more akin to 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.

[18] See Autor and Duggan (2006) and Liebman (2015) for discussions of this phenomenon. In the United States, 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.
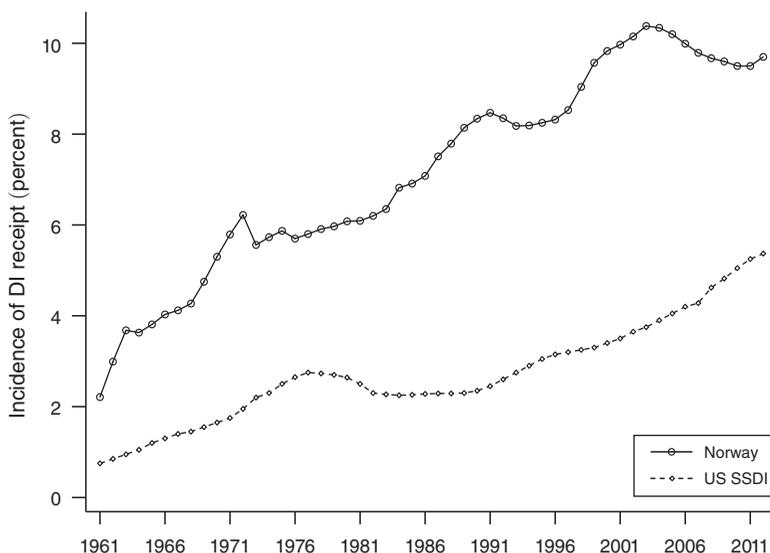
FIGURE 1. TRENDS IN DI RECEIPT IN NORWAY AND THE UNITED STATES

Notes: This figure displays trends in DI receipt in Norway and the United States (see Section I). United States trends are based on Autor and 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 is defined as the percent of 18 to 67 year-old adults receiving DI benefits in Norway and the percent of 25 to 64 year-old adults receiving DI benefits in the United States.

mortality rates, and moreover, because mental illness typically has an early onset, DI recipients with such diagnoses tend to participate in the program for relatively long periods. With a progressively smaller share of DI recipients either passing away or reaching retirement age in a given year, the DI exit rate has fallen secularly in both countries (see online Appendix Figures A2 and A3). The aging of the Baby Boom cohorts into their peak (near-elderly) disability age brackets has contributed substantially to the expansion of the US DI rolls since the mid-1990s (Liebman 2015).

There are noteworthy differences between the US and Norwegian DI programs. One difference is their income replacement rates. Kostøl and Mogstad (2014) compute the replacement rate for a typical Norwegian applicant 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 US 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 United States receive health insurance coverage through the federal Medicare program, which is a substantial in-kind benefit.[19] In Norway, by contrast, all citizens are eligible for health insurance through the Social Insurance System. Another difference concerns the appeals process. Appeals among initially rejected applicants are far more prevalent in the United States than in

---

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

Norway, 48 percent in the United States (French and Song 2014) versus 25 percent in Norway, and success rates at appeal are also considerably higher in the United States than in Norway.

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 mental and musculoskeletal disorders (see online Appendix Table A1). And in both countries, appellants are more likely than average applicants to be relatively young, have lower prior earnings, and claim mental and musculo-skeletal disorders (see online Appendix Table A2). As a further comparison among the two programs, Figure 2 uses Norwegian and US data (the latter from Maestas, Mullen, and Strand 2013) to plot earnings trajectories of DI applicants and appel-lants in Norway and the United States, before and after their DI determinations. We focus on years $t-4$ through years $t+4$ surrounding the initial DI determination (left-hand panels) and the year of the initial appeal decision (right-hand panels).

The patterns are quite similar across the two countries. DI applicants who are allowed at the initial determination have on average substantially higher prior earn-ings 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 coun-tries, and this gap is not closed over the subsequent four post-decision years. The figures in the bottom row plot the difference between denied and allowed applicants (panel E) and appellant labor earnings (panel F) over time, controlling flexibly for observable characteristics and lagged dependent variables (up to the year of the ini-tial decision, after which they are fixed as the mean over the years prior to decision).

III. Assessing the Instrument

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

A. *Instrument Relevance*

Figure 3 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 (con-trolling 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 nineti-eth percentile allowing approximately 18 percent of cases as compared to approxi-mately 8 percent for a judge at the tenth 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
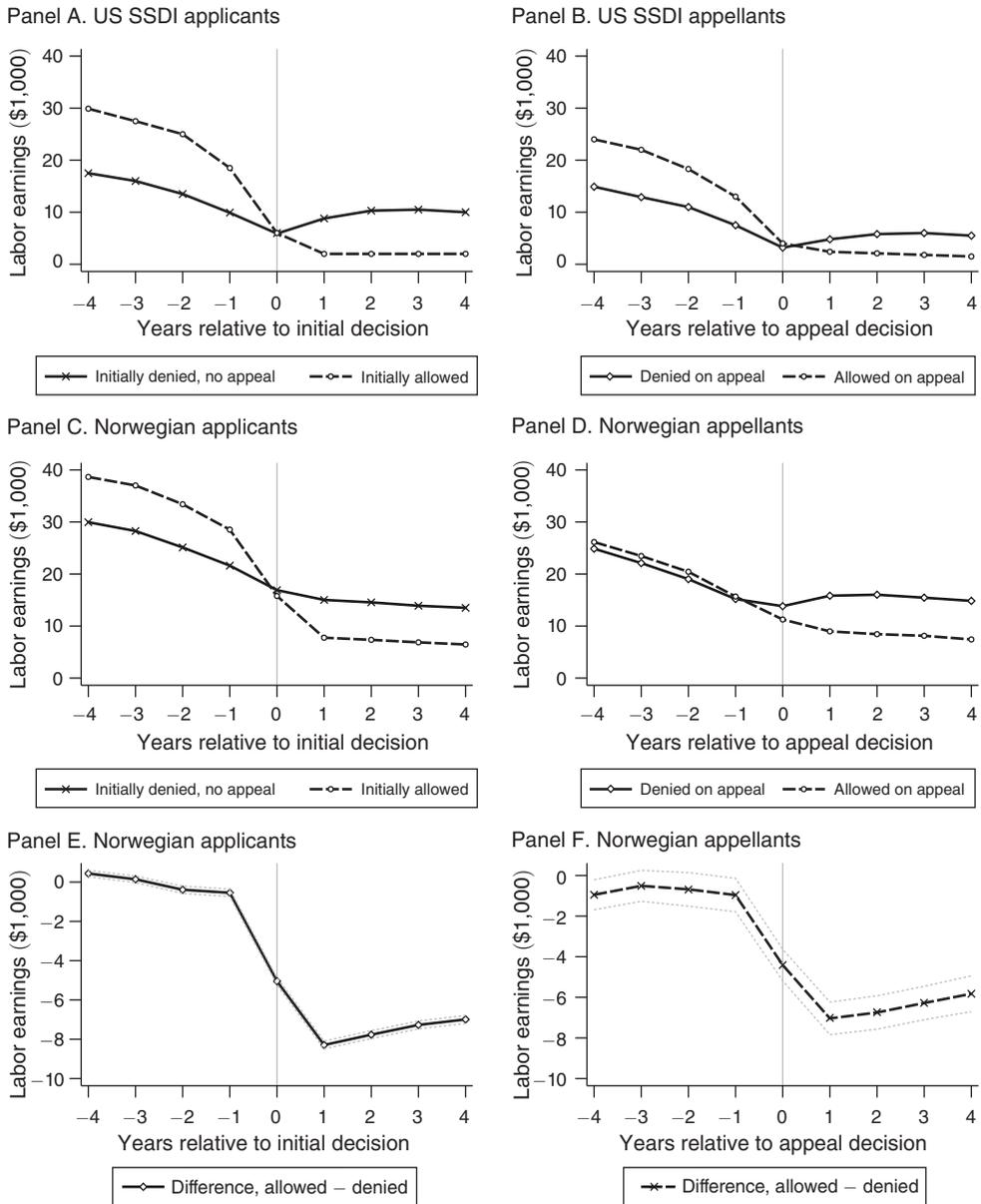
FIGURE 2. EARNINGS TRAJECTORIES OF ALLOWED AND DENIED DI APPLICANTS AND APPELLANTS

Notes: Panels A through D 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 United States (top row, sourced from Maestas, Mullen, and Strand 2013), and for Norway (middle row). 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 II for further details). Panels E and F plot the difference between denied and allowed applicants (E) and appellant labor earnings (F) 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 row represent 90 percent 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.
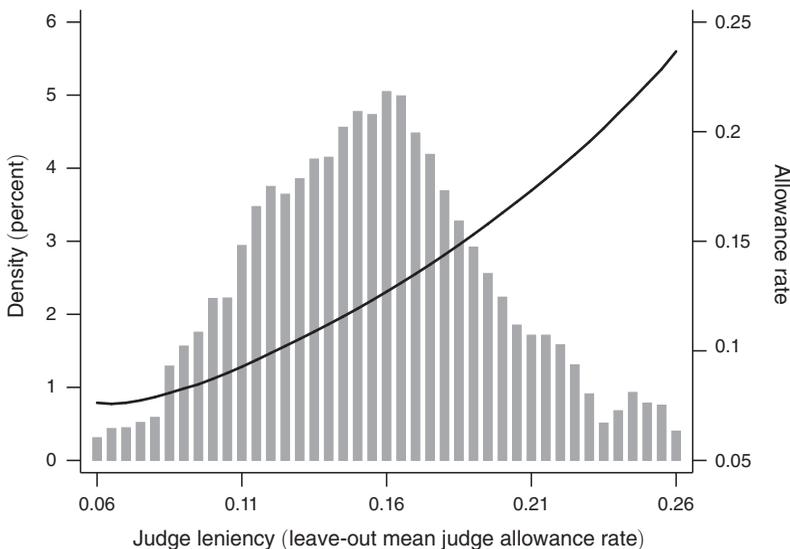
FIGURE 3. EFFECT OF JUDGE LENIENCY ON DI ALLOWANCE

Notes: This figure displays the estimated effect of judge leniency on DI allowance among appellants. Baseline estimation sample consists of individuals who appeal an initially denied DI claim during the period 1994–2005 (see Section II for further details). There are 75 unique judges. The solid line plots a local linear regression of allowances on judge leniency while including fully interacted year and department dummies. A histogram of judge leniency is plotted in the background.

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 1 through 4 following appeal. These columns are identical except for the very modest impact of sample attrition (less than 3 percent over four years) stemming from death or emigration of appellants.[20] 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.

---

[20] Column 1 of Appendix Table A5 documents that the assignment variable (judge leniency) does not affect the probability that an appellant either dies or emigrates during the outcome period.

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

Note: Standard errors (in parentheses) are clustered at the judge level. 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 II 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.

B. *Instrument Validity*

Table 1 provides strong empirical support for the 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 condition needed for a causal interpretation of the IV estimates is that the judge leniency instrument has a monotonic effect on DI allowances.[21] Monotonicity requires that, for each appellant, the probability 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 toward a subset of cases. Reassuringly, when separately estimating the first stages based on the (predetermined) observable characteristics of the individual, we find that the estimates are consistently positive and sizable, consistent with the monotonicity assumption (see online Appendix Table A3).

As a second check on this threat to validity, we directly examine whether judges who are stricter toward one subset of appellants (e.g., young appellants, those with mental disorders) are also relatively strict toward 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 leniency using only decisions in cases with female appellants. Column (2) of online Appendix Table A3 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.

IV. 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.

A. *Effects on Labor Earnings and DI Benefits*

In panels A through 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.

---

[21] If the treatment effect of the disability determination were constant among appellants, the monotonicity assumption would be unnecessary. But we do not find the constant treatment effect assumption plausible in this setting.

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

Notes: Standard errors (in parentheses) are clustered at the judge level. 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 II 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 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.[22] Over the first four years following appeal, the causal effect of being allowed at the appeal on subsequent DI receipt falls by approximately one-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.[23] Panel B displays analogous estimates for DI benefit

---

[22] Note that $0.989 = 0.784/0.793$, where 0.793 is the corresponding first-stage coefficient from Table 3, panel B, column 1.
[23] Although this pattern could also be consistent with successful appellants exiting the DI program in years 2 through 4, such exits rarely occur.

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 4. 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 40 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.

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 and 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).[24]

In online Appendix Figure A4, 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.

B. *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 toward these other programs. Conversely, DI beneficiaries may also seek other

---

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

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 online Appendix Figure A5, 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 VI, we explore whether spousal labor supply provides an additional margin through which married appellants may buffer household income in the event of DI denial.

V. 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.[25] 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

---

[25] 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 online Appendix Table A5, 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, Mogstad, and Vestad (2017).

TABLE 5—EFFECT OF DI ALLOWANCE ON HOUSEHOLD INCOME, FISCAL COSTS, AND CONSUMPTION

	Years after decision				
	1	2	3	4	Average
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: $\beta_t = \Delta$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: $\beta_t = \Delta$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: $\beta_t = \Delta$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: $\beta_t = \Delta$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

Notes: Standard errors (in parentheses) are clustered at the judge level. 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 II 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: $\beta_t = \Delta$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 consumption in outcome years $t \in \{1,2,3,4\}$. *p*-values for H_0: $\beta_t = \Delta$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 $t \in \{1,2,3,4\}$.

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 5 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 is far smaller than its gross effect due to the influence of household labor supply, reapplication behavior, and benefit substitution.[26]

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.[27] 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 A.2 suggests that DI allowances raise household income by less than $0.50 per $1 of net government expenditure, 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).[28]

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.

VI. 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, Pistaferri, and Saporta-Eksten 2016). 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.

---

[26] Specifically, the row labeled H_0: $\beta_t = \Delta \text{DI benefit}_{t1}$ reports p-values for the null hypothesis that 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 $\Delta \text{DI} > \Delta \text{HH}$ consumption are equal to one-half the p-values for two-tailed tests.)

[27] 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.

[28] 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).

A. *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.[29] Our point estimates therefore imply that $0.55 of each $1 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 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 the panel 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 income effect exceeds $2,500 (9 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.[30] 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, which differs from zero with 90 percent confidence.

---

[29] 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.

[30] 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).

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: $\beta_t = \Delta$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: $\beta_t = \Delta$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: $\beta_t = \Delta$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: $\beta_t = \Delta$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

Notes: Standard errors (in parentheses) are clustered at the judge level. 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 II 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: $\beta_t = \Delta$DI benefit$_{t1}$ in panels A.1 and B.1 correspond to tests of whether the effect of a successful DI appeal (in year *t*0) on initial DI income (in year *t*1) 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: $\beta_t = \Delta$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 *t*0 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.

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.

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: $\beta_t = \Delta$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: $\beta_t = \Delta$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: $\beta_t = \Delta$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: $\beta_t = \Delta$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

Notes: Standard errors (in parentheses) are clustered at the judge level. 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 II 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: $\beta_t = \Delta$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: $\beta_t = \Delta$HH Income$_{t1}$ 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.

B. *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

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

Notes: Standard errors (in parentheses) are clustered at the judge level. 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 II 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.

analysis since our data do not allow us to determine whether non-married appellants are single or cohabiting.[31]

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 subsample, we estimate that labor supply reductions and transfer payment increases roughly offset each other. Panels C and D consider the impact 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 2 through 4 following the award as shown in panel C. These estimated labor supply reductions in years 2 through 4, averaging $16,500, are statistically significant. Panel D shows, however, that as much as 50 percent of

---

[31] The Norwegian decennial census data allow us to observe cohabitation though unfortunately our annual administrative data do not. The Census data show that 59 percent of DI participants (applicants are not identified in the Census data) are married, 32 percent are single non-cohabitants, and only 9 percent are cohabitants. We test whether judge leniency causes endogenous selection into or out of marital status in online Appendix Table A5. Columns 2, 3, and 4 find no evidence that judge leniency affects the likelihood of a change in marital status (overall, from unmarried to married, or from married to unmarried).

the reduction in spousal labor earnings induced by a DI allowance is effectively offset by a countervailing increase in transfer payments to the spouse.[32] 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 2 and 3 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 online Appendix Figure A6, 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 2 (panel D) 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.

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 online Appendix Table A3, 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.

VII. 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.

A. *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.[33] 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).[34] As in Low and 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 and Pistaferri (2015) in the parametric specification of preferences. At time t, the instantaneous utility function of unmarried households with a given disability severity is

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

[33] 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 and Meghir (1991); Blundell, Browning, and Meghir (1994); Aguiar and Hurst (2013); and Blundell, Pistaferri, and Saporta-Eksten (2016).

[34] As in Maestas, Mullen, and Strand (2013) and Kostøl and Mogstad (2014), employment is an indicator variable that is equal to 1 if annual earnings exceed the annual substantial gainful activity threshold, set annually by the Norwegian Social Security Administration (at approximately US$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:

$$(4) \qquad U_2\big(C_t, P_{A,t}, P_{S,t}, R_t; H\big)$$

$$= \big(C_t \exp\{-P_{A,t}\phi_{2,A,H} - P_{S,t}\phi_{2,S}\} - R_t \exp\omega_2\big)^{1-\mu_2}/(1-\mu_2),$$

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 and 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 and Pistaferri (2015) in defining young disability appellants as those less than 45 of years of 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

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

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

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

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

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

where

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

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 and 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. Online Appendix Tables A7 and A9 present parameter estimates of the earnings processes and the corresponding labor supply elasticities with and without the selection correction. As shown in online Appendix Table A10, 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, Storesletten, and Violante (2014) and Blundell, Pistaferri, and Saporta-Eksten (2016), 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:

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

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

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

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

$$(9) \qquad I_{M,t} = \Phi_{M,D_t,K,t}.$$

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 and Pistaferri (2015), we model DI approval upon reapplication as

$$(10) \qquad D_{t+1} = D_t + R_t\left(1 - D_t\right)\pi_{M,H,t},$$

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

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

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:

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

where r is the real interest rate. To close the model, we follow Low and 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

$$V_{2,t}(D_t, \tau_{A,t}, \tau_{S,t}, S_t; O_2)$$

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

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 are 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, Pistaferri, and Saporta-Eksten (2016) 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 and Pistaferri (2015), as described above. In the second step, certain model parameters are set externally. We follow Low and 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.[35] 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 online 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 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). Online Appendix Figure A7 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

---

[35] 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 online Appendix B.B1. 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 online Appendix B.B2. 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 4. 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.

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 4 and online Appendix Figure A7 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 10 IV estimates) for married couples and 52 moments (46 raw data moments and 6 IV estimates) for unmarried individuals. We choose two sets of moments to match. The first set 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.[36]

---

[36] 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

B *Empirical Results*

Parameter Estimates and Model Fit.—The externally set parameters and estimated utility parameters are presented in panels A and B of online Appendix Table A8, respectively, while the estimated parameters from the earnings processes are reported in online Appendix Table A7. The parameters are precisely estimated. As anticipated, and in agreement with results by Low and 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 1 percent increase in disposable income from working, obtaining (Marshallian) labor supply elasticities (see online Appendix Table A9). 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, Pistaferri, and Saporta-Eksten (2016) obtain estimates ranging from -0.08 to 0.42, while Blundell et al. (2016) 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, Pistaferri, and Saporta-Eksten (2016) 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.[37] The model is relatively successful in replicating both consumption and income effects of DI allowance. Figure 5 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. Online Appendix Figure A8 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 moments do we observe differences that exceed one standard deviation.

Lastly, Figure 6 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.

---

equation (13) of Blundell et al. (2016) and is motivated by the finding of Altonji and 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 online Appendix B.B3 for further details.

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

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 US$1,000.
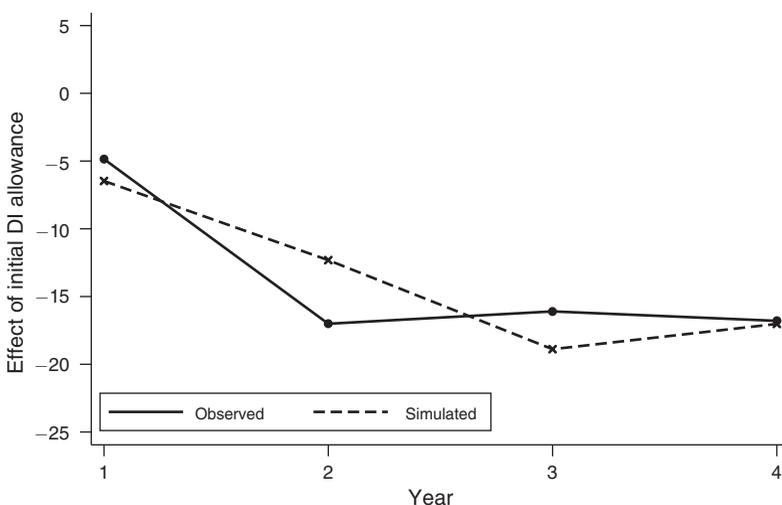

FIGURE 5. FIT OF INSTRUMENTAL VARIABLES ESTIMATES: SPOUSAL EARNINGS DYNAMICS

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 US$1,000.

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 (WTP) 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 online Appendix B.B4 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
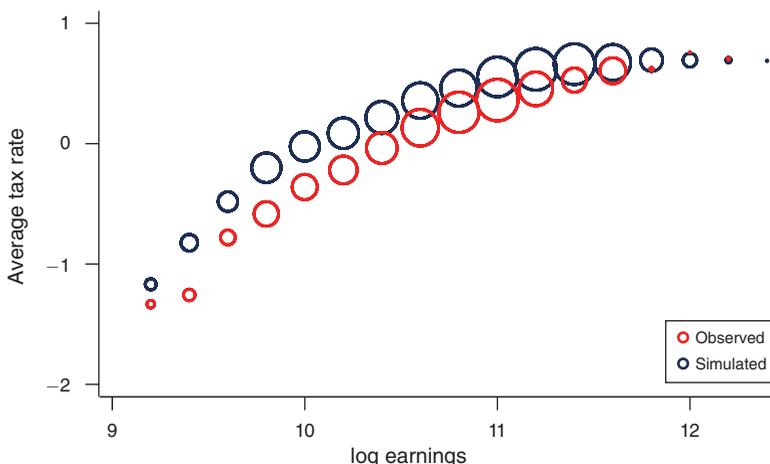
FIGURE 6. FIT OF AVERAGE TAX RATES SIMULATED FROM THE MODEL

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.

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

Notes: 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 *Net of income effect* row, we subtract the average effect of DI on disposable income from the average willingness to pay. In the *Per dollar of fiscal costs* row, we divide the average willingness to pay by the average effect of DI on fiscal costs. In the *As a share of household consumption* row, we divide the willingness to pay by the level of consumption for each household. The hypothesis test H_0: *WTP* $= 0$ corresponds to testing whether the willingness to pay for DI receipt is equal to 0. *p*-values are based on re-estimating the model on 20 block bootstrap replicates of the data (where the block corresponds to the individual).

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.

Comparing the willingness to pay for a DI allowance relative to its fiscal cost suggests that, on average, each net $1 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 $1 of fiscal costs remains substantial. Benchmarking willingness to pay against consumption levels, the average DI allowance is valued at about 43 percent of annual consumption for single and unmarried households and 9 percent 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 twenty-fifth percentile, $6,274 at the median, and $12,563 at the seventy-fifth 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

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:		
$\quad WTP(Baseline) = WTP(Counterfactual)$	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

Notes: 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 panel A, we use the estimated model to compute the welfare benefit of DI receipt. In panel B, 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 panel D, 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 0. The hypothesis H_0: $WTP(Baseline) = WTP(Counterfactual)$ correspond to testing whether the average willingness to pay in the baseline (unconstrained) model equals the average willingness to pay in the counterfactual. p-values are based on re-estimating the model on 20 block bootstrap replicates of the data (where the block corresponds to the individual).

as much among married households and 1.7 times as much among unmarried households.[38] 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 one-half of initial appellants receive a DI allowance within four years (panel A of Table 4). 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 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.

Robustness to Relaxing the Borrowing Constraint.—Following Low and 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

---

[38] 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.

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.

VIII. 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 $1 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 US 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 United States 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 US than Norwegian households, despite the lesser generosity of the US 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 United States, 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 and Duggan 2006, Kostøl and 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 reinforce spousal labor supply responses to temporary or moderate income cuts to the other spouse (see, e.g., Chetty and 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.

<div align="center">REFERENCES</div>

Aguiar, Mark, and Erik Hurst. 2013. "Deconstructing Life Cycle Expenditure." *Journal of Political Economy* 121 (3): 437–92.

Altonji, Joseph G., and Lewis M. Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." *Journal of Business & Economic Statistics* 14 (3): 353–66.

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

Autor, David H. 2011. "The Unsustainable Rise of the Disability Rolls in the United States: Causes, Consequences, and Policy Options." NBER Working Paper 17697.

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

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

Autor, David H., Mark G. Duggan, and Jonathan Gruber. 2014. "Moral Hazard and Claims Deterrence in Private Disability Insurance." *American Economic Journal: Applied Economics* 6 (4): 110–141.

Autor, David H., Andreas Kostøl, Magne Mogstad, and Bradley Setzler. 2019. "Disability Benefits, Consumption Insurance, and Household Labor Supply: Dataset." *American Economic Review.* https://doi.org/10.1257/aer.20151231.

Autor, David H., Nicole Maestas, Kathleen J. Mullen, and Alexander Strand. 2017. "Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants." NBER Working Paper 20840.

Baicker, Katherine, Amy Finkelstein, Jae Song, and Sarah Taubman. 2014. "The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment." *American Economic Review* 104 (5): 322–28.

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

Blundell, Richard, Monica Costa Dias, Costas Meghir, and Jonathan Shaw. 2016. "Female Labor Supply, Human Capital, and Welfare Reform." *Econometrica* 84 (5): 1705–53.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *American Economic Review* 98 (5): 1887–921.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. "Consumption Inequality and Family Labor Supply." *American Economic Review* 106 (2): 387–435.

Borghans, Lex, Anne C. Gielen, and Erzo F. P. Luttmer. 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, John. 1989. "The Health and Earnings of Rejected Disability Insurance Applicants." *American Economic Review* 79 (3): 482–503.

Browning, Martin, and Søren Leth-Petersen. 2003. "Imputing Consumption from Income and Wealth Information." *Economic Journal* 113 (488): F282–301.

Browning, Martin, and Costas Meghir. 1991. "The Effects of Male and Female Labor Supply on Commodity Demands." *Econometrica* 59 (4): 925–51.

Chen, Susan, and Wilbert van der Klaauw. 2008. "The Work Disincentive Effects of the Disability Insurance Program in the 1990s." *Journal of Econometrics* 142 (2): 757–84.

Chetty, Raj, and Àdàm Szeidl. 2007. "Consumption Commitments and Risk Preferences." *Quarterly Journal of Economics* 122 (2): 831–77.

Cullen, Julie Berry, and Jonathan Gruber. 2000. "Does Unemployment Insurance Crowd out Spousal Labor Supply?" *Journal of Labor Economics* 18 (3): 546–72.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics* 129 (4): 1711–52.

Eika, Lasse, Magne Mogstad, and Ola L. Vestad. 2017. "What Can We Learn about Household Consumption from Information on Income and Wealth?" Unpublished.

Fadlon, Itzik, and Torben Heien Nielsen. 2019. "Household Labor Supply and the Gains from Social Insurance." *Journal of Public Economics* 171: 18–28.

Finkelstein, Amy, Nathaniel Hendren, and Erzo F. P. Luttmer. Forthcoming. "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment." *Journal of Political Economy.*

French, Eric, and Jae Song. 2014. "The Effect of Disability Insurance Receipt on Labor Supply." *American Economic Journal: Economic Policy* 6 (2): 291–337.

Gruber, Jonathan. 2000. "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy* 108 (6): 1162–83.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2014. "Consumption and Labor Supply with Partial Insurance: An Analytical Framework." *American Economic Review* 104 (7): 2075–126.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.

Imbens, Guido W., and Donald B. Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies* 64 (4): 555–74.

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

Kostøl, Andreas Ravndal, and Magne Mogstad. 2014. "How Financial Incentives Induce Disability Insurance Recipients to Return to Work." *American Economic Review* 104 (2): 624–55.

Kostøl, Andreas Ravndal, and Magne Mogstad. 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): 123–50.

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

Lundberg, Shelly. 1985. "The Added Worker Effect." *Journal of Labor Economics* 3 (1): 11–37.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–829.

Meghir, Costas, and Luigi Pistaferri. 2011. "Earnings, Consumption and Lifecycle Choices." In *Handbook of Labor Economics,* Vol. 4B, edited by Orley Ashenfelter and David Card 773–854. Amsterdam: Elsevier.

Meyer, Bruce D., and Wallace K.C. Mok. 2019. "Disability, Earnings, Income and Consumption." *Journal of Public Economics* 171: 51–69.

Milligan, Kevin S., and David A. Wise. 2011. "Social Security and Retirement around the World: Mortality and Health, Employment, and Disability Insurance Participation and Reforms." Unpublished.

Organisation for Economic Co-operation and Development (OECD). 2010. *Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries.* Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). 2015. OECD Social Expenditure Database (SOCX).

Parsons, Donald O. 1980. "The Decline in Male Labor Force Participation." *Journal of Political Economy* 88 (1): 117–34.

Persson, Petra. Forthcoming. "Social Insurance and the Marriage Market." *Journal of Political Economy.*

Pistaferri, Luigi. 2015. "Household Consumption: Research Questions, Measurement Issues, and Data Collection Strategies." *Journal of Economic and Social Measurement* 40 (1–4): 123–49.

Social Security Advisory Board. 2015. Technical Panel on Assumptions and Methods.

US Social Security Administration (SSA). 2017a. *The 2017 Annual Report of the Board of Trustees if the Federal OId-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* https://www.ssa.gov/OACT/TR/2017/.

US Social Security Administration (SSA). 2017b. *Annual Report of the Supplementary Security Income Program.* https://www.ssa.gov/OACT/ssir/SSI17/.

US Social Security Administration (SSA). 2017c. *Annual Statistical Report on the Social Security Disability Insurance Program, 2016.* Washington, DC: Social Security Administration Office of Retirement and Disability Policy.