Central problem in public finance: social planner wants to redistribute (or insure) but has imperfect information about "ability" (or underlying attribute along which want to redistribute (or insure))

- Concern that may transfer to people whom don’t want to, and miss people whom do
  - e.g. is DI going to people who are truly disabled, cash transfers to people who truly have no productive employment etc
- Concern about distorting incentives (e.g. distort labor supply if transfer based on earnings)

- Recall optimal income tax problem from 471
  - Diamond-Mirlees etc.
- Tagging (this unit) and screening (in Unit VI):
  - Ways to improve social planner’s ability to insure or redistribute
Tagging and screening

- Attempts to combat moral hazard
- Up until now have simply asked: empirically how to estimate the mh costs of a social insurance program and weight those against benefits
- Now want to ask: are there ways we can design programs to reduce moral hazard?
  - This brings us to: tagging and screening
- Aside on ternominoloy:
  - tagging often referred to as "targeting"
  - screening often referred to as "self-targeting"
Outline: Tagging

- Theory:
  - Akerlof (1978) Economics of Tagging
  - Diamond and Sheshinski (1995) DI as a tag

- Empirical question: how good is tag / how much distortion in behavior does it generate?
  - Application: Disability insurance
Recall from 14.471 the basic optimal income tax problem: Want to redistribute from high ability (high marginal product) to low ability (low marginal product)

- Key challenge: ability (wage) not observed therefore distribute on the basis of income (wage*hours) which creates distortion in labor supply

General vs Targeted Redistribution:
- Negative income tax: general tax system that redistributes to poor
- Targeted programs: choose an (identifiable) group to redistribute to

US has opted for targeted redistribution
- More targeted allows you to spend less to reach the people you want
- But may be more costly to administer and/or encourage adverse behavior
Akerlof Tagging Model

- Negative income tax:

\[ T = -\alpha Y_{avg} + tY \]

where \(\alpha\) is fraction of per capita avg income (\(Y_{avg}\)) received by a person with 0 gross income (i.e. minimum support); \(t\) is the marginal rate of taxation

- Summing over all individuals and dividing by total income gives:

\[ t = \alpha + g \]

where \(g\) is the ratio of net taxes collected to total income

- Key points:
  - Tradeoff: higher levels of support (\(\alpha\)) come at the cost of higher marginal tax rates (\(t\))
  - Usual distortions: \(t\) decreases incentive for labor supply
Akerlof Tagging Model (con’t)

- Suppose that we can identify (tag) a group of people that contains only the poor and this group contains only a fraction $\beta < 1$ of the population.

Give the minimum support $\alpha$ to only this fraction, funded with same marginal tax rate $t$:

$$t = \beta \alpha + g$$

vs. general negative income tax:

$$t = \alpha + g$$

- Key point: tagging allows greater support for the poor with less distortion in the tax structure.
  - For given amt of support $\alpha$, $t$ is lower with tagging.
Akerlof tagging (cont’d)

- Benefits of tagging: lower tax rate for given amount of transfers to tagged group
- Potential costs of tagging:
  - Higher administrative costs
  - Potential inequity (what if poor but not in tag?)
  - Endogenous tags / Potential behavioral distortions
- Result in paper: if tagging is costless, should always do some redistribution based on tag
  - Intuition: envelope theorem. First amt of tagging generates only second order DWL from distortion in behavior, but first order transfer gain.
- NB: Quantitative (empirical) questions still remain
  - What is the optimal level of a tag?
    - Or (a la Baily!): on the margin should we increase or decrease use of this tag?
  - Another key empirical question: endogeneity of tag
Tagging (examples)

- Akerlof example: categorical welfare
  - i.e. Cash welfare to poor in female headed households
    - Lower marginal product (i.e. child care costs etc)
    - Endogeneity of tag?

- Disability insurance can also be rationalized / understood as a potential tag...
Diamond-Sheshinski (1995)

- People have different disutilities of work
- First best outcome: only work if marginal product of work exceeds disutility from work
  - Consumption fully insured across states (work / not work)
- Issue: don’t directly observe “disutility of work”
- The disabled have higher disutility of work
  - Disability as a tag for high disutility of work / want to redistribute income to
- By adding disability insurance to existing income tax system can redistribute with less distortion (Akerlofian tag)
  - optimal disability insurance is non zero (envelope thm)
  - Again though, doesn’t tell us what optimal system is or whether on margin should expand or reduce current DI benefits...
Take optimal social insurance level problem (tradeoff between insurance and incentives) and add an imperfect tag

Key feature of their model: imperfect tag

- Observed disability is an imperfect screen of true medical condition / disutility of work
- Type I and Type II errors
- The villagers in the boy who cried wolf
Don’t Make a Type III Error

Type I error (false positive)
You’re pregnant

Type II error (false negative)
You’re not pregnant
Take optimal social insurance level problem (tradeoff between insurance and incentives) and add an imperfect tag

Key feature of their model: imperfect tag
- Observed disability is an imperfect screen of true medical condition / disutility of work
- Type I and Type II errors

Government gets an imperfect signal of disutility of work
- Standard result that larger benefits provide better insurance but with larger efficiency costs
- Main new result: optimal insurance rate increasing in how good the screening device is
  - The worse the screening device, the lower the optimal insurance rate
Empirical question: how good a tag is disability

- Type I and Type II errors in screening process
- What empirical literature discusses:
  - would those on DI be working absent DI
  - But how does this relate to optimal DI? What we really want to know is disutility of work among marginal accepted applicant
(Brief) mechanics of DI tag

- For more details see Gruber textboko, Krueger and Meyer review article, Autor and Duggan (2006 JEP)
- Risk being insured: permanent medical disability that makes you “unable to engage in substantial gainful employment”
- Instrument: Publicly provided insurance
  - Administered at national level (uniform across states)
  - Added to SS in 1956; several later expansions broadened age eligibility and type of disabilities covered
  - Largest US income replacement program for non elderly adults (and one of only a few non temporary programs)
    - 16% of SS spending
  - A large and growing program
    - In 1985: 2.2% of adults (ages 25-64) on DI
    - 2005: 4.2%
    - Predicted 6% soon...
- Financed through SS payroll tax
Eligibility

Prior work requirements
- Most non elderly adults meet these

Health requirements: subjective assessment of disability
- Must have a medically determinable physical or mental impairment that is
- Expected to result in death or to last at least a year
- Prevents the person from engaging in a "substantial gainful activity"
- Cannot be "gainfully employed" while applying

NB: Have people on boards making review decisions; highly subjective

Award process

Approximately 1/3 of applications awarded benefits at first review
- Can appeal if denied (approx 60% of rejects file at least one appeal)
- Ultimately due to appeals, about half of applicants receive DI
Cash benefits based on prior earnings history

- Amount is equivalent to SS benefits if retired at full benefit age
- Benefits a function of average annual earnings
  - 3-piece linear function (slopes (RR) of 90%, 32% and 15%)
  - Avg RR is 42% (benefits not taxed)
- Five month waiting period for benefits to begin
  - Analog to a deductible

(con’t)
Medical Benefits: Medicare

- Eligibility begins 2 years after DI (optimal? Who gets screened out by this?)

Benefit duration: permanent

- Exit DI by
  - Reaching full retirement age and shifted onto SS (44%)
  - Die (41%)
  - Recover to point where no longer meet medical standards
How good a tag is DI?

- Diamond-Sheshinski: The better we are at identifying people with high disutility of work (i.e. truly disabled) the more we can transfer at given efficiency cost
  - Question: How good is DI as a tag (i.e. what is disutility of work on those on DI vs those off DI)

- The literature has focused instead on: What are the labor force participation effects of DI?
  - NB: Not directly addressing key D-S parameter: disutility of work on those with DI
  - Even if most on DI would work absent DI might not be socially optimal for them to do so (i.e. might work despite high disutility of work bc of high cost of lack of smooth consumption and failure of private DI markets . . . )
  - Important and unexplored area of work!!
    - More generally: DI literature has more program evaluation / labor economics than public economics in it - Opportunities!
    - but see Autor et al. "Disability Benefits, Consumption Insurance and Household Labor Supply".
LFP effects of DI: Overview

- Not going to try to summarize whole literature
- Focus on several nice pieces of empirical work
  - Parsons / Bound debate
    - How to sensibly probe / investigate whether a "problematic" regression is indeed problematic
  - Maestas Mullen and Strand (AER 2013)
    - MTE's
  - Autor, Maestas, Mullen and Strand (2015)
    - Asking a new (important) question
  - Deshpande (2016)
    - Impact of SSI receipt in childhood on adult outcomes
    - Nice presentation of empirical results
Time series: 1960s and 1970s parallel increases in DI roles and decline in lfp of older men...

Parsons (JPE 1980):
- Probit of labor force participation on potential replacement rates
  - Cross – sectional variation in replacement rates (due to variation in earnings)
  - Finds large effects of replacement rates on lfp
  - "Obvious endogeneity problem"

Extrapolation suggests increase in DI generosity can explain all of the large decline in older male LFP 1948 - 1976
Extremely creative paper

Rather than just saying there’s an “obvious endogeneity problem”
  
  shows that it in practice looks important
  
  comes up with an alternative / creative approach

Showing the bias: replicates Parson’s regression on sample of workers who never applied for DI and finds similarly high elasticity of LFP wrt potential DI replacement rate

  
  But for non applicants, what would be causal connection btw high potential benefits and If withdrawal?!
Bound bounds (con’t)

- Creative alternative: Looks at behavior of rejected DI applicants
- Uses LFP behavior of denied applicants as an upper bound on effect of DI on LFP.
- Hypothesis: as long as screening procedures are even partially effective, rejected applicants should be in better health than non rejected applicants. Therefore their subsequent earnings and employment history provide an upper bound on what the LFP of successful applicants might look like in absence of DI.
  - Assumption: on average denied applicants in better health.
  - Supportive evidence for this assumption
Finding: Based on difference in LFP of denied and accepted applicants concludes that at most 34% of SSDI beneficiaries would be at work absent SSDI benefits (i.e. lots would still not be working)

- Von Wachter et al (AER 2010) find this upper bound has been relatively stable over time despite changing composition of applicants

Idea: using subsequent lfp participation of rejected applicants as upper bound on lfp of accepted applicants in absence of DI system
Bound bounds (con’t)

- Idea: using subsequent Lfp participation of rejected applicants as upper bound on Lfp of accepted applicants in absence of DI system

- Critique I: Maybe there are two types: lazy and sick. Lazy quit and apply for DI but are disproportionately rejected so their LFP understates LF potential of those accepted for DI (not lazy)
  - Although rejected and accepted have similar pre application earnings and employment

- **Critique II: program may reduce LFP among those rejected: act of applying for DI may impair LFP since have to withdraw from Lf to apply and then wait while application is pending (erosion of human capital)**
  - Therefore LFP of rejected may be lower than what it would be in absence of program
Bound bounds (conclusion)

- Methodologically: Creative and interesting paper: Well worth reading
  - We should try to do more analysis of how big the “obvious endogeneity problem” might be
  - Think about useful bounding exercises
    - e.g. welfare loss of insurance capped by premium at which you could buy it

- Substantively: Some issues
  - May not in fact be an upper bound (effects of application process – will discuss AMMS 2015)
  - LFP in “absence of DI” is not the only (or key) question
    - Marginal effects of DI on LFP (implications for reform)
    - Welfare/ optimal DI
One of several recent papers looking at how receipt of DI affects LFP

- Paper tries to get an estimate (rather than a bound) of effect of DI receipt on LFP (not getting at marginal effects of changes in benefits).
- Clear and well-written
- Key feature that I will emphasize: tracing out heterogeneous treatment effects of DI receipt as function of unobserved severity
- Will discuss: How is this done and why do we care?

Empirical strategy: Variation across DI examiners in “allowance rate” (severity / leniency).

- Provides potentially random variation in whether you get on DI
- NB: bc of appeals process think of this as ITT: examiner’s allowance propensity in initial determination state as IV for ultimate DI receipt
  - An increasingly common type of instrument (Kling and judges; Doyle and foster care case manager; Williams on patent examiners etc)

- Look at impact of DI receipt on LFP
Estimate a LATE for the marginal entrant for whom DI receipt affected by examiner

Treatment is receipt of DI

Causal effect of DI on LFP averaged over compliers: applicants whose treatment status changed as a result of examiner assignment.

Not informative for effects on “always takers” (those who would be allowed whether initially or upon appeal) even if assigned to the strictest examiner or “never takers” – those who would be denied even if assigned to the most lenient examiner
IV Assumptions

- Examiner affects probability receive DI ("first stage")
- Exclusion restriction: Examiner does not affect LFP through any mechanism other than DI receipt.
- Possible violations of exclusion restriction:
  - What if examiners of different strictness get different applicants?
    - Argue: Examiner assignment (conditionally) random
    - NB: Key distinction is not random assignment vs conditional random assignment but knowing the assignment algorithm vs. assuming it
  - What if other things about examiner affect LFP
    - look at e.g. waiting time variation by examiner leniency and conclude they are not too concerned.
    - What if lower allowance rate examiners also give a "stern speech" ?!
    - More generally, they are assuming the impact of the examiner is all via impact on receipt of DI, not other things.

- Monotonicity ("no defiers"): examiners’ award propensities affect applicants’ chances of DI receipt in the same way
  - (con’t)
Monotonicity ("no defiers"): examiners’ award propensities affect applicants’ chances of DI receipt in the same way

- Cases allowed by stricter examiners would be allowed by more lenient examiners (and cases denied by lenient examiners would also have been denied by strict examiners)
- Otherwise, w/o monotonicity if you have heterogeneous treatment effects interpretation of IV gets screwed up...

NB: monotonicity applies at the individual level (not just “on average”)

- e.g. if strict examiner flips a coin and denies 1 in 2 applicants and lenient examiner rolls a die and denies 1 in 6 applicants, that violates monotonicity
Maestas et al. (con’t)

- Data: 2005-2006 applications linked to subsequent LFP
  - Administrative data within SSA on earnings
  - Up to nine years before and 2-4 years after initial decision
This is the instrument: variation in allowance rates across examiners.

After adjustment for case mix: about one-third of examiners have allowance rates more than 7 pctg pts above or below the average.

Important to understand range of variation...
IV strategy – graphical depiction

- Left hand graph: First stage. SSDI receipt increasing in allowance rate. (Slope is <1 because of appeals)
- Right hand graph: ITT. LFP decreasing in allowance rate
Main (average) results

- First stage Table 2
  - regress DI receipt on examiner allowance rate. Coeff about 0.22.
  - Interpretation: A 10 pctg pt increase in examiner allowance rate associated with about a 2.2 percentage point increase in DI receipt

- Table 4:
  - OLS: Difference in LFP between denied and allowed after about two years is about 34 pctg pts (v similar to Bound)
  - IV estimate of LFP on DI receipt (instrumenting for DI receipt w examiner allowance rate): DI receipt reduces LFP by about 22 pctg pts after two years
  - IV estimate of impact of DI receipt on LFP lower than OLS
    - Consistent w Bound intuition that OLS estimates give you an upper bound on effect of DI receipt on LFP
    - IV allows you (presumably) to adjust for unobservable differences between accepted and rejected applicants
    - Suggests accepted applicants are in unobservably worse health
Marginal treatment effects

- With homogeneous treatment effects, LATE generalizes to ATE (causal effect of DI receipt on the LFP of all applicants) and to the average effect of treatment on the treated (ToT) – i.e. causal effect of DI receipt on beneficiaries in particular.

- With heterogeneous treatment effects, LATE may be of limited interest.
  - LATE gives the local average treatment effect of those whose treatment (DI) receipt is affected by the instrument (i.e. “the compliers”).
  - Not necessarily the population of interest. How to generalize?
  - Would like to trace out how LATEs vary.
Idea: they have variation in (continuous) instrument (examiner allowance rate)

Assume that the marginal applicant allowed by the more lenient (vs more strict) examiner is marginally less severe in terms of unobserved underlying health / disability

They estimate a series of LATEs - impact of DI on LFP – for applicants at different points in the unobserved severity distribution

- Trace out marginal treatment effects (MTEs)

MTE: treatment effect for marginal applicant (i.e. severity that just makes you a marginal accept for that examiner’s allowance rate)

- Heckman, Urzua and Vylacil (2006)
In addition to standard IV assumptions (first stage, exclusion restriction, and monotonicity) estimation of MTEs requires an additional assumption: an index model of assignment with a single unobservable

In their case they assume the unobservable is severity

Key assmpt: Unobserved severity of allowed applicants is monotonically related to examiner allowance rates
Marginal treatment effects: key additional assumption

- MTEs require an index model of assignment with a single unobservable
  - In their case: Unobserved severity of allowed applicants is monotonically related to examiner allowance rates

- This is an assumption
  - What if lenient examiner denies every 4th applicant and strict deny every second applicant? [Their response: then wouldn’t find anything]
  - What if above some severity threshold stricter ones give more of a penalty for personal appearance that day in court. Then you are still tracing out MTEs but their interpretation is different: how does treatment effect (effect of DI on LFP) vary for people with different personal appearances on their day in court?

- Can (and do) try to look at how reasonable assumption is by looking at how observable characteristics of accepted applicants change with examine allowance rate – e.g. observed impairment
  - But ultimately an additional assumption
Marginal treatment effects

- Look at impact of DI receipt on LFP for different allowance rates (transformed into predicted DI rates on X axis)

- Basically estimating a bunch of separate LATEs (as vary examiner severity) and plotting them
  
  - Find larger (in abs value) LFP effects of DI where predicted DI receipt is higher (examiner more lenient) and implicitly (by assmpt) severity is lower

- Discinentive effect of SSDI rises as allowance threshold (severity) is lowered
Why do we want MTE’s?

- In general too little attention in PF (relative to say IO) on heterogeneity and heterogeneous treatment effects
  - Here: interesting to know for design purposes how effect of DI on LFP varies (i.e. greater for those w less severe impairments)
    - Although admittedly here fact that it is on unobservables makes policy implications harder
    - Would prefer to have the MTE in a space that is inherently interesting (e.g. price vs “allowance rates”)
  - More generally, interesting economics arise when consider heterogeneous treatment effects
    - Ex in health: Heterogeneous mh and selection on it (Einav et al AER 2013)
In principle with wide enough support over which estimate MTEs can get more "general" estimates

- E.g. if variation in examiner predicted DI receipt was not 55 to 75% but say 10 to 90%, perhaps would feel comfortable trying to extrapolate to LFP when allowance rate is 0 ("no DI")
  - Although still have the erosion of human capital issue

Key: need variation in a (continuous) instrument

- Or binary instrument and more assumptions - Brinch, Mogstad, Wiswall (JPE 2017 "Beyond LATE with a discrete instrument")
DI and LFP - recap

- Bound: Using LFP of rejected applicants as upper bound for LFP of accepted applicants (Bound)
- Point estimate: Using (instrumented) award of SSDI to applicant to get estimate of impact of SSDI receipt on LFP
- What about effect of applying for SSDI (need to get out of labor force to apply)?
Another potential channel

- Literature is looking at impact of SSDI on LFP by comparing post-application behavior of those awarded vs denied benefits
- Recall one critique of Bound bound: LFP of rejected applicants may not be an upper bound on lfp of accepted applicants in absence of DI if act of applying reduces LFP
  - Key question: how does time out of LF while waiting for SSDI determination affect subsequent employment and earnings?
  - Relatedly, how does a "delay" effect bias existing estimates of a "receipt" effect?
Applying for SSDI

- Restrictions on labor force activity
  - After filing, before decision, cannot earn more than 1,000 per month (this would exceed SGA and result in denial)

- Lengthy application process
  - Average time to decision = 14.1 months
  - About half of applicants challenge initial decision → average time to decision > 2 years
Employment decay - implications

Two channels for impact of DI application on earnings and employment:
- Causal effect of DI receipt (vs not) on earnings, conditional on processing time
- Effect of processing time on earnings (delay effect)

Key implications:
- SSDI may affect employment of those not covered!
  - Policy implications for relaxing work disincentives during application process?
  - Research implications: Earnings of denied SSDI applicants understate LF potential at time of application
- Comparisons of post-determination LF participation of allowed and denied will produce biased estimates of direct effect of DI allowance if they do not account for differences in processing time between allowed and denied applicants
  - e.g. bc denied applicants often spend a lot of time appealing initial denial, impacts of benefit receipt on employment are biased downward bc confounded with effect of longer delay for denied.
Empirical model

\[ y_i = X_i \beta + \delta T_i + \gamma D I_i + s_i + \varepsilon_i \]

- \( y_i \): measure of observed labor supply at some point after initial determination
- \( X_i \): observables that influence labor supply (e.g. age, impairment type)
- \( s_i \): unobserved factors that influence labor supply (e.g. severity)
- \( T_i \): total processing time. Months from filing date to last observed DI date
- \( D I_i \): Whether applicant was ultimately awarded benefits
Figure 1. Conceptual Sketch of the Effects of SSDI Processing Time and Benefit Receipt on Labor Supply

\[ \delta \cdot T \]

\[ \gamma \]

\[ \text{slope} = -\delta \]

\( T \)

Denied Applicant

Allowed Applicant

Source: 2005 DIODS Data. Examiners with 10-900 decisions only. Caseload characteristics includes DDS, geography, body system code, age, pre-onset earnings, concurrent status, and Terminal Illness diagnosis.
Empirical challenge

\[ y_i = X_i \beta + \delta T_i + \gamma D_l_i + s_i + \epsilon_i \]

- unobserved severity \((s_i)\) correlated with processing time \((T_i)\) and/or receipt \((D_l_i)\)
- Use variation across examiners in allowance rates and processing time as instruments
Empirical strategy I: Initially allowed applicants

- Everyone receives DI in this sample ($D_i = 1$)
- Variation in processing time $T_i$ comes from differences across (conditionally randomly assigned) initial examiners
- Examiner variation in initial examiner average processing times ("productivity")
  - processing times uncorrelated with allowance decisions (so don’t control for it)
Empirical strategy II: Adding in initially denied applicants

- Longer processing time (and presumably resultant human capital deterioration) increases the chance that an initially denied applicant appeals and is ultimately awarded.
  - Examiner processing time not a valid instrument for decision time among those initially denied, because it also affects probability of ultimately receiving DI.
- To identify causal effect of processing time on sample that includes initially denied applicants, need another source of variation that affects likelihood of receiving SSDI but uncorrelated with health or initial processing time.
  - Use variation in initial examiner allowance propensity (a la MMS).
One instrument, two channels?

- Assume conditional random assignment of initial examiners
- Examiners vary on two dimensions:
  - Allowance rate
  - Processing time
- Key assumption (exclusion restriction):
  - Initial examiners affect subsequent LFP only through two channels which they have parameterized correctly: processing time and allowance rate
One instrument, two channels?

- Key assumption (exclusion restriction):
  - Initial examiners affect subsequent LFP only through two channels which they have parameterized correctly: processing time and allowance rate

- Note 1: Processing time and allowance rate do not have to be uncorrelated since they are both included in regression (although in practice they appear to be)

- Note 2: Implies original MMS results biased.
  - Initial allowance rate used to estimate direct effect of SSDI award on LFP biased downward bc applicants to stricter examiners are more likely to appeal and therefore face longer delays (as well as greater chance of not being covered).
  - MMS exclusion restriction violated. Initial examiner effect on LFP not only through channel of SSDI receipt but also through wait time.
One instrument, two channels?

- Note 3: Let’s hope the new exclusion restriction is right
  - Possible other channels?!

- Paper examines this for an over-identification test of causal pathways by which examiners affect applicants’ subsequent employment
  - How does residual sum of squares change from a "restricted" RF regression of employment compared to "unrestricted"
  - "Restricted" regression has EXALLOW and/or EXTIME
  - "Unrestricted" has examiner dummies

- Reject EXTIME or EXALLOW as sole causal pathway but fail to reject EXTIME and EXALLOW as (combined) sole pathways

- General comment: unbundling fixed effects (e.g. Doyle et al. on hospitals)
Side comment: other dimensions of examiners?

- Autor et al. (2015) consider time and allowance rate
- Sam Norris (2018 JMP) "Judicial Errors: Evidence from refugee appeals"
  - Considers two dimensions of judge: allowance rate and error rate
  - Observes two judges (if first one allows, has to get approved by a second one too).
  - Idea: if there are two first-round judges who approve the same number of first-round claimants, the more consistent judge will have a higher share of her claimants approved by the second-round judge.
  - Other possible interesting applications? (Medical second opinions? Other?)
Side comment: unbundling FEs

- Autor et al. (2015) overid tests cannot reject partition of examiner FE into EXTIME AND EXALLOW
- Can we use FEs to uncover role of different components?
  - e.g. Autor et al on examiner fixed effects
  - e.g. Doyle et al. (JPE forthcoming) on hospital FE
- If various observables not mutli-collinear (and have more FEs than observables) can do the decomposition
- Note however interpretation puts you back in realm of OLS land (correlations)
Finding: longer processing time significantly reduces subsequent employment and earnings of SSDI applicants

- Delay causes decay

Bias in estimate of impact of SSDI receipt on labor supply that ignores delay channel can be substantial

- MMS estimates understate causal effect of SSDI receipt on labor supply by about 50 percent
- Intuition: rejected applicants are initially denied and therefore more likely to appeal during which they are out of LFP so we under-estimate their LF potential (per Bound critique).
Does Welfare Inhibit Success?

- Deshpande (AER 2016) Long-Term Effects of Removing Low Income Youth from Disability Rolls (SSI)

- Debate:
  - Welfare as vital lifeline for those who face barriers to work
  - Welfare creates perverse incentives and perpetuates dependency
    - Those who receive welfare today more likely to receive it tomorrow
  - Challenge: Distinguish state dependence from serial correlation (recall same issue with inertia in health insurance)
Deshpande (2016)

- Looks at impact of being removed from SSI at age 18
- Research design: Welfare report (Personal Responsibility and Work Opportunity Act) of 1996 increased the number and strictness of medical reviews for 18-year-olds to remain on SSI
  - Applied only to children with an 18th birthday after August 22, 1996 (date of PRWORA enactment)
  - RD in birthdate of likelihood of removal from SSI at 18
Figure 2 summarizes the empirical strategy for this RD design in date of birth.

The x-axis shows the date of the child’s 18th birthday, with a vertical line at the August 22, 1996 cutoff. The graph plots the proportion of children in each birthweek bin who receive an age 18 medical review, receive an unfavorable age 18 medical review, and ever (up to age 35) receive an unfavorable medical review.

The figure confirms that the PRWORA changes were enforced: while almost no children with an 18th birthday immediately before the cutoff (hereafter, “control” group) received an age 18 medical review, nearly 90 percent of children with an 18th birthday immediately after the cutoff ("treatment" group) received one. This discontinuity in the likelihood of receiving an age 18 medical review translates into a 39 percentage point discontinuity in the likelihood of receiving an unfavorable age 18 medical review. Age 18 medical reviews are a specific type of the more general medical reviews used to verify continued eligibility for both adults and children.

As shown in Figure 2, children with an 18th birthday after the date of PRWORA enactment are 28 percentage points more likely to ever receive an unfavorable medical review until the last time I observe them at age 35. This discontinuity is smaller than the previous ones since children on the left-hand side of the graph, who do not receive an age 18 medical review, are more likely to continue on SSI as adults and receive adult medical reviews.

Notes: Figure plots proportion of SSI children in each birthweek bin who receive an age 18 medical review, receive an unfavorable age 18 medical review, and ever receive an unfavorable medical review (through 2013). Sample is SSI children with an 18th birthday within 37 weeks of the August 22, 1996 cutoff.
Empirical strategy

- Interested in impact of SSI removal on long-term outcomes:
  \[ Y_{it} = \alpha + \beta SSIStatus_{it} + X_i + \varepsilon_{it} \]

- Use cutoff of 18th birthday after August 22 1996 as instrument for SSI status.
  - Run first stage equation year by year:
    \[ SSIStatus_i = \alpha_0 + \beta_0 Post_i + \gamma_0 DOB_i^n + \kappa_0 (POST_i \times DOB_i^n) + X_i + \varepsilon_i \]
    where \( Post_i \) is a dummy for 18th birthday after cutoff and \( DOB_i^n \) is date of birth running variable of polynomial order \( n \)
First stage

August 22, 1996 cutoff as an instrument for SSI status. The first-stage equation is equation (1) with covariates and the endogenous variable on the left-hand side:

\[ SSIStatus_i = \alpha_0 + \beta_0 \text{Post}_i + \gamma_0 \text{DOB}_i + \kappa_0 (\text{Post}_i \times \text{DOB}_i) + X_i + \epsilon_i. \]

The covariates in \( X_i \) include sex, diagnosis category, age at entry, parental earnings prior to 1997, and state. Recall from Figure 2 the first-stage effect on age 18 removal: having an 18th birthday after the cutoff increases the likelihood of receiving an unfavorable age 18 review by 39 percentage points. Figure 3 plots the RD estimate for SSI enrollment estimated separately for each year using equation (4); the first-stage graph for event year 4 is given as an example in online Appendix Figure A.5.

As expected from quasi-random assignment, there is no difference between the control and treatment groups in the probability of SSI enrollment prior to age 18. The difference between the control and treatment groups does not open up until two years after the year of the child's 18th birthday as a result of lags in decision time. The first stage reaches 24 percentage points four years after age 18 and then attenuates rapidly until it plateaus at 5 percentage points.

I find that most of the attenuation is attributable to control group members leaving the program in large numbers as adults, with 47 percent not enrolled in either SSI or DI by 2013 (see online Appendix Figure A.6).

Notes: Figure plots parametric RD estimates of the effect of a child having an 18th birthday after the August 22, 1996 cutoff, using a polynomial order of 2 with covariates. Shaded region is 95 percent confidence interval. Sample is SSI children with an 18th birthday within 37 weeks of the August 22, 1996 cutoff.

Figure 3. Change in First Stage for SSI Enrollment Over Time

Years since 18th birthday

RD estimate

SSI enrollment

0

-0.05

-0.1

-0.15

-0.2

-0.25

-5 0 5 10 15

Years since 18th birthday

Figure 3. Change in First Stage for SSI Enrollment Over Time

Notes: Figure plots parametric RD estimates of the effect of a child having an 18th birthday after the August 22, 1996 cutoff, using a polynomial order of 2 with covariates. Shaded region is 95 percent confidence interval. Sample is SSI children with an 18th birthday within 37 weeks of the August 22, 1996 cutoff.
Results

SSI youth who are removed via age 18 medical review increase their earnings by an average of $830 each year, replacing approximately one-third of the $2,170 annual loss in SSI cash benefits. This earnings response represents a 20 percent increase over the control group mean of $4,200. Those who are removed increase their earnings by $9,890 in present discounted value over the 16 years following removal.

I do not observe effects on Medicaid enrollment, but I present back-of-the-envelope calculations in the online Appendix.

Figure 4. Reduced-Form Effect on Annual Earnings


Figure 5. IV Estimates of the Effect of Age 18 Removal

Notes: Figure plots parametric IV RD estimates of the effect of age 18 removal on annual SSI income, earnings, and total income, using a polynomial order of 2 with covariates. Shaded region is 95 percent confidence interval. Sample is SSI children with 18th birthday within 37 weeks of August 22, 1996 cutoff.
Youth removed from SSI at age 18 recover one-third of the lost SSI cash income in earnings
  - Removal not only decreases mean income but increases volatility of income

Implications for welfare?
  - Ideally would observe consumption (or better yet marginal utility of consumption!)

Note that increased removal probability was unanticipated
  - Question: Would effect of forseen (at say age 12) increased removal probability have different effect?
  - Why? And how to get at this?
DI and LFP - recap

- Bound: Using LFP of rejected applicants as upper bound for LFP of accepted applicants (Bound)
- Maestas et al: Point estimate - Using (instrumented) award of SSDI to applicant to get estimate of impact of SSDI receipt on LFP
- Autor et al - Effects of applying / delay for SSDI
- Deshpande - long term effects of SSI for kids
- What about marginal effect of SSDI benefit level?
Apart from impact of DI receipt on LFP (binary) would like to know marginal effect of benefits on LFP

- Getting closer to a potential input into a Baily type calculation!

Gruber (2001 JPE) provincial variation in Canada

Autor – Duggan (2003 QJE)

- National program. Where does variation in replacement rate come from?
- The denominator rather than the numerator!

- Local area changes in wage distribution over time generates variation in the replacement rate
Need for other work on DI

- Most empirical work is on moral hazard side (LFP)
- Missing work on benefits side!
  - Consumption smoothing
  - Disutility from work? Is it socially optimal for the marginal guys to be working?
- More generally, looking at normative welfare questions vs program evaluation
  - Compare to UI literature
- Recent exception: Autor, Kostol, Mogstad (NBER WP 2015) "Disability Benefits, Consumption Insurance and Household Labor Supply"
Need for other work on DI

- Program interactions / “externalities” also of interest
  - E.g. If tighten DI benefits, do we get increase in UI, SS, SSI etc
    - Missing from Baily formula?!

- Some work but not much
  - Borghans, Gielen and Luttmer (“Social support shopping”) looking at impact of DI reforms on other social services in Netherlands...
  - Impact of health insurance expansions on DI receipt (Massachusetts - Maestas et al; Oregon - Baicker et al).
Another tagging application: Categorical welfare

- Categorical welfare: restriction based on demographic characteristic (e.g. single motherhood)
  - Idea: single moms are in general low human capital, low (net) wage
  - Seems like a good potential tag

- Potential problems with categorical welfare:
  - Is tag endogenous / mutable?
  - Administrative costs of verifying tag (is there really no “man in the house”? Etc)
  - “Stigmatizing” effect of welfare
    - Could this be a good thing?...
Tagging: Akerlof; Diamond-Sheshinkin
- A way to combat moral hazard and increase constrained optimal level of insurance / transfers

Key empirical questions:
- How targeted is the tag (Type I and Type II errors)
- What is the optimal amount of the tag

What the empirical DI literature has done so far:
- Impact of DI on labor market activity
- Related but not the same

Lots of opportunities to bring some public finance to the DI literature!