Appendix to "Early-Onset Disability, Education Investments, and Social Insurance"

By Robert Millard

April 7, 2025

1 Model Parameters

Table 1: Summary of Model Parameters.

Parameter	Description
Individual Heterogeneity	
$\bar{a}^{d_0}$	mean of endowed ability distribution
$\sigma^2_{a^{d_0}}$	variance of endowed ability distribution
$\bar{\psi}^{d_0}$	mean of idiosyncratic cost distribution
$\sigma^2_{\psi^{d_0}}$	variance of idiosyncratic cost distribution
Earnings Process	
ϕ^{d_0}	direct effect of disability on earnings
$\mu_1^{d_0,s}$	return to potential experience
$\mu_2^{d_0,s}$	return to potential experience squared
$\sigma^2_{\xi^{s,d_0}}$	variance of productivity shock distribution
$\bar{\xi}_0$	mean of initial productivity shock distribution
$\sigma^2_{\xi_0}$	variance of initial productivity shock distribution
$h^s_{d_0}$	return to post-secondary
Utility Parameters	
β	discount factor
κ	coefficient of relative risk aversion
θ	utility cost of disability
η_1	utility cost of working
η_2	utility cost of working with a disability
F^d	cost of working with a disability at old age
Policy Parameters	
π^s	probability of DI acceptance
π^{SA}	probability of SA-D acceptance
C^{s,d_0}_{App}	utility cost of DI application
Labour Market Environment	
$\delta_t^{d,s}$	exogenous job destruction rate
$\lambda_t^{s,d}$	exogenous job arrival rate
$\gamma_{i,j}^{d_0,t}$	disability transition probability

2 Measuring Disability in the Data

Disability in the model is measured by reported limitations to activities of daily living (LADL). The set of LADLs is derived from a short version of a module called "the disability screening questions" developed by Statistics Canada for identifying individuals with disabilities in general population surveys (Grondin, 2016). This model distinguishes five main areas of activity limitation: Seeing, Hearing, Physical, Cognitive, and Mental Health. Sample survey questions used to identify disability status in the data are reported in Table 2

Table 2: Survey Questions on Limitations to Daily Activities

<u>Physical limitation</u> -How much difficulty do you have walking on a flat surface for 15 minutes without resting? -How much difficulty do you have walking up or down a flight of stairs, about 12 steps without resting? -How much difficulty do you have reaching in any direction, for example, above your head? -How much difficulty do you have using your fingers to grasp small objects like a pencil or scissors? -Do you have pain that is always present?
<u>Cognitive limitation</u> -Do you think you have a condition that makes it difficult in general for you to learn? This may include learning disabilities such as dyslexia, hyperactivity, attention problems, etc.. -Has a teacher, doctor or other health care professional ever said that you had a learning disability? -Has a doctor, psychologist or other health care professional ever said that you had a developmental disability or disorder? This may include Down syndrome, autism, Asperger syndrome, mental impairment due to lack of oxygen at birth, etc.. -Do you have any ongoing memory problems or periods of confusion? Please exclude occasional forgetfulness such as not remembering where you put your keys.
<u>Mental Health limitation</u> -Do you have any emotional, psychological or mental health conditions? These may include anxiety, depression, bipolar disorder, substance abuse, anorexia, etc..
<u>Sensory Limitation</u> - How often does this difficulty seeing limit your daily activities? - How often does this difficulty hearing limit your daily activities?

Note: Provides the disability-related survey questions in LISA used to construct disability status.

2.1 Validity of Disability Measures

Much research in health economics has focused on the validity of self-reported measures of one's health. One concern relates to the inherent subjectivity of how one assesses one's own health. For example, two otherwise identical individuals may differ in the reported severity of their disability. Additionally, critics of self-reported health measures argue that individuals may exaggerate the existence or severity of their health condition to justify poor economic outcomes or attachment to government programs, a phenomenon referred to as justification bias. The evidence on the endogeneity of self-reported health measures and the extent of measurement error are mixed (Black et al., 2017). Although, it is important to note that recent articles tend to find evidence for state-dependent reporting.[1]

My disability measure is derived from a respondent reporting any positive limitations to a specified activity and abstracts from the degree of impairment. This approach mitigates concerns related to subjectivity in the scale of impairment from a self-reported activity limitation, as I do not distinguish conditions along the severity margin. Moreover, much of the evidence on justification bias is based on broad questions about one's health or disability, such as "do you have a medical or physiological condition that impairs the type or amount of work you can do." The questions about activity limitations in this survey are linked to specific tasks, such as walking on a flat surface for fifteen minutes, grasping a small object like scissors, or experiencing ongoing memory problems or periods of confusion. Additionally, the presence of some activity limitations is elicited based on whether the respondent has been diagnosed with a specific condition, such as a learning or developmental disorder, by a healthcare professional.[2] Last, mental health is identified using specific examples of diagnoses, such as anxiety, depression, bipolar disorder, or anorexia. These approaches narrow the scope of justification bias to be anchored to the activities in question, base the existence of a limiting condition on the diagnosis of a medical professional, or frame limitations related to mental health with specific examples of diagnoses. I follow much of the related literature and take the responses to questions on limitations to daily activities as given. However, I acknowledge the empirical concerns that are inherent to any self-reported measures of health.

[1]It has been found that self-reported disability is close to exogenous, may actually under-represent the extent disabled population, and may even underestimate the true impact of disability on relevant labour market outcomes (Stern, 1989; Bound and Burkhauser, 1999; Burkhauser et al., 2002). Others have found evidence of justification bias related to labour market states inflating the prevalence of health conditions (Benítez-Silva et al., 2004; Baker et al., 2004; Black et al., 2017). Moreover, alternate approaches to identify individuals with disabilities, for instance, by using disability insurance beneficiaries to define the population with a disability, have been found to under-represent the population of individuals who are limited enough in the labour market to be classified as "disabled" (Bound, 1989)

[2]This type of question has been used to assess the validity of self-reported health measures in Baker et al. (2004)

3 Value Functions and Numerical Solution to Structural Model

There is no analytical solution to the model so it is solved numerically. For a given set of the structural parameters, the solution algorithm is straightforward, as each period's decisions and policy functions are conditional discrete choices. In the following, I suppress the individual's subscript, i, to simplify notation. Beginning with the terminal condition in T (retirement at age 65), I iterate backward, numerically approximating the value functions, characterizing the work decision and Disability Insurance (DI) application decision at each age after eighteen as a function of $S_t = \{d_t, \epsilon_t, e_{t-1}, \rho_t\}$. Given the solution to the individual's labour market decisions, I solve the policy function for the education choice at age eighteen as a function of initial heterogeneity, $\{a, d_0, \psi\}$.

Retirement

Solving the model starts with the terminal condition, retirement. The value of the terminal period is deterministic for a given set of the state variables. I assume that state variables remain fixed as soon as an individual retires, $S_t = S_{t+1} = \bar{S} = \{\bar{d}, \bar{\epsilon}, \bar{e}, \bar{\rho}\}$. Individuals make no decisions in retirement. They receive utility from consuming their retirement income, which is known with certainty given their earnings index at the end of their working life.[3] I assume individuals expect retirement to last until age 75, after which they die with certainty. The value of retirement is

$$V_t^R(\bar{S}) = u^N(c_t; \bar{d}\ ) + \beta V_{t+1}^R(\bar{S}) \tag{1}$$

$$= u^N(\bar{c}; \bar{d}\ ) + \sum_{\tau=1}^{T^L} \beta^\tau u^N(\bar{c}; \bar{d}\ ) \tag{2}$$

$$\text{s.t. } c_t = 5500 + 0.25\bar{e}. \tag{3}$$

Before retirement, individuals can find themselves in one of three states in the labour market; working, not working and receiving SA, or not working and receiving DI. I consider the value functions and timing of choices for each state in turn, for ages less than 60 when individuals do not have the option to retire.

Value of Working

Given S_t, employed individuals earn flow utility from consuming after-tax employment income and from SA at the beginning of the period. Shocks to productivity and disability then update to ϵ_{t+1} and d_{t+1} and the earnings index updates given their labour earnings. Individuals then face the job destruction rate, $\delta_t^{d_0,s}$, which places them out of work in the next period. If their job is not destroyed, individuals may choose to continue working or leave work. The value function for employed individuals is

$$V_t^E(S_t) = u^W(c_t; d_t) + \beta E_t\Big[\delta_t^{d_0,s} V_{t+1}^U(S_{t+1}) + (1-\delta_t^{d_0,s})\max\big\{V_{t+1}^U(S_{t+1}), V_{t+1}^E(S_{t+1})\big\}\Big] \tag{4}$$

$$\text{s.t. } c_t = \tau\big(W_t L_t,\ 0\big) + SA_t(\tau\big(W_t L_t,\ 0\big), d_t), \tag{5}$$

$$e_t = f(e_{t-1}, W_t, t). \tag{6}$$

[3] The individual's contribution period ends at T^L so their earnings index remains constant after this time.

Value of Not Working and Receiving Social Assistance (SA)

While out of work, an individual receives flow utility from consuming SA income. Then, if eligible, they choose to apply for DI, $m_t = 1$, to become a beneficiary at the beginning of the next period. If applying, they are accepted with probability π^s. If accepted, their disability and productivity shocks update and their earnings index becomes fixed. If rejected, they do not receive a job offer and remain out of work for the next period. If the agent does not apply, $m_t = 0$, then their productivity and disability status update, and they receive a job offer with probability $\lambda_t^{d_0,s}$. If offered, they choose to accept and enter work the next period or to reject and remain out of work the next period. If the individual does not receive a job offer, they remain out of work for the next period. The value function for an unemployed individual at age t is

$$V_t^U(S_t) = u^N(c_t; d_t) + \beta \, \mathrm{E}_t \max_{m_t \in \{0,1\}} \Big[m_t \big(\pi^s V_{t+1}^{DI}(S_{t+1}) + (1-\pi^s) V_{t+1}^U(S_{t+1}) - C_{app}^{d_0,s}\big) \tag{7}$$

$$+ (1-m_t)\Big(\lambda_t^{d_0,s} \max\big\{V_{t+1}^U(S_{t+1}), V_{t+1}^E(S_{t+1})\big\} + (1-\lambda_t^{d_0,s}) V_{t+1}^U(S_{t+1})\Big)\Big] \tag{8}$$

$$\text{s.t. } c_t = SA(0, d_t), \tag{9}$$

$$e_t = f(e_{t-1}, 0, t). \tag{10}$$

DI Beneficiary

I assume that individuals cannot work when receiving DI but can receive SA benefits simultaneously. Periods that the individual receives DI are not included in their contribution period. Therefore, their earnings index does not change when on DI. DI beneficiaries face the risk of reassessment of benefits, ρ. If benefits are not reassessed, the individual may or may not receive a job offer. If they receive an offer, work is added to their choice set. The value function for a DI recipient is

$$V_t^{DI}(S_t) = \; u^N(c_t; d_t) + \beta E_t \Big[(1-\lambda_t^{d_0,s}) \max\{V^U(S_{t+1}), V^{DI}(S_{t+1})\} \tag{11}$$

$$+ \lambda_t^{d_0,s} \max\{V^E(S_{t+1}), V^U(S_{t+1}), V^{DI}(S_{t+1})\}\Big] \tag{12}$$

$$\text{s.t. } c_t = \tau\big(0, \; DI_t\big) + SA_t(\tau\big(0, \; DI_t, \big), d_t) \tag{13}$$

$$e_t = e_{t-1}. \tag{14}$$

In each period t, for every possible combination of the discrete state variables—both those that are time-varying and those that are fixed—I evaluate the continuation value (Emax) on a discretized grid of the continuous state variables. The continuous state variables are initially $(a_i, \epsilon_{it}, e_{i,t-1})$, where a_i is endowed ability, ϵ_{it} is the accumulated productivity shock, and $e_{i,t-1}$ is the earnings index. Because a_i and ϵ_{it} only affect future earnings, I can reduce the problem to tracking $(W_{it}, e_{i,t-1})$ where W_{it} is the current earnings.

To compute the expected continuation value, I integrate out the next period's productivity shock, ξ_{it+1}^{s,d_0} . Assuming this shock is normally distributed, I use Gauss–Hermite quadrature to numerically approximate the integral over its distribution. For each realization of the discrete state variables, I construct an approximation of the continuation value by evaluating the expected payoff on a discrete grid for $(W_{it}, e_{i,t-1})$. Finally, to handle points that lie between the grid values in the continuous space, I apply bilinear interpolation. This approach ensures a smooth approximation of the continuation value function while keeping the computational burden tractable.

3.1 Smoothing

Applying indirect inference to a discrete choice model presents challenges due to the discontinuous nature of the mapping from structural parameters to simulated data. Small changes in the structural parameters can lead to abrupt shifts in the simulated outcomes, causing the auxiliary model's parameter estimates to change discontinuously. These discrete jumps introduce discontinuities in the objective function, complicating optimization. Additionally, some parameter changes may not affect the discrete choices at all, resulting in flat regions in the objective function.

To address these issues, I adopt a Generalized Indirect Inference (GII) procedure, which smooths the objective function and mitigates both flat spots and discontinuities (Bruins et al., 2018), (Keane and Smith, 2003). The key idea is to apply distinct auxiliary models to the simulated and observed data. In particular, the auxiliary model for the simulated data is designed to fit the continuous latent variables that underlie the observed discrete outcomes. Provided that both auxiliary models yield asymptotically equivalent vectors of pseudo-true parameters, the GII estimator—defined by minimizing the distance between the two models—remains consistent and asymptotically normal.

To implement GII, I introduce an i.i.d. taste shock, $\zeta_t^k = (\zeta_t^E, \zeta_t^U, \zeta_t^{DI})$, into the utility associated with each labor market state. These shocks are interpreted structurally as unobserved state variables known to the agents but not to the econometrician. The shocks follow a multivariate extreme value distribution with scale parameter λ. Their inclusion necessitates modifications to both the model's solution method and the estimation algorithm.

In solving the model, I follow a similar procedure as previously described, with the key distinction that I now account for the newly introduced state variables when computing the expected maximum (Emax) functions within the continuation values at each decision point. To illustrate, the value function for the employed state becomes:

$$V_t^E(S_t) = \; u^W(c_t; d_t) + \lambda\zeta_t^E + \beta E_t\Big[\delta_t^{d_0,s}(V_{t+1}^U(S_{t+1}) + \lambda\zeta_{t+1}^U) \tag{15}$$

$$+ (1-\delta_t^{d_0,s})\max\big\{V_{t+1}^U(S_{t+1}) + \lambda\zeta_{t+1}^U, V_{t+1}^E(S_{t+1}) + \lambda\zeta_{t+1}^E\big\}\Big] \tag{16}$$

$$V_t^E(S_t) = \; u^W(c_t; d_t) + \lambda\zeta_t^E + \beta\Big[\delta_t^{d_0,s}E_t(V_{t+1}^U(S_{t+1}) + \lambda\zeta_{t+1}^U) \tag{17}$$

$$+ (1-\delta_t^{d_0,s})\underset{k\in E,U}{\mathcal{LS}}\big(V_{t+1}^U(S_{t+1}), V_{t+1}^E(S_{t+1})\big)\Big] \tag{18}$$

where $\mathcal{LS}$ is the log-sum function

$$\underset{k\in E,U}{\mathcal{LS}}\big(V_{t+1}^U(S_{t+1}), V_{t+1}^E(S_{t+1})\big) = \lambda\log\big(exp(V_{t+1}^U(S_{t+1})/\lambda) + exp(V_{t+1}^E(S_{t+1})/\lambda)\big). \tag{19}$$

Now, the conditional choice probability of each labor market state at period t is given by,

$$Pr(V = V_t^j | S_t) = \frac{exp(V_t^j/\lambda)}{exp(V_t^U/\lambda) + exp(V_t^E/\lambda)} \tag{20}$$

The estimation procedure for the model with taste shocks follows similar steps as before, with one key modification: moments in the auxiliary model are now calculated using choice probabilities rather than observed outcomes. For example, the auxiliary model includes conditional employment rates computed from the observed data. I estimate the model's parameters by matching these observed rates to the corresponding conditional employment probabilities generated by the simulated model. The fundamental principle of Generalized Indirect Inference (GII) is that the estimation procedures applied to the observed and simulated data need not be identical, so long as both yield consistent estimates of the same vector of pseudo-true parameter values.

4 Censoring

The data used in the analysis is an unbalanced panel, as such there is considerable censorship present when calculating the moments making up the auxillary model for estimation. To address this, I replicate censoring observed in the data and impose it when calculating the moments using data simulated from the model. I calculate the probability of an observation being censored conditional on age, a lag for censorship in the previous period (L.1), censorship in the previous two periods (L.2), and censorship in the previous three periods (L.3). I estimate a separate linear probability model conditional on d_0 and s, giving four sets of estimates. The results from the estimation are reported in Table 3 below.

Table 3: Censoring

	Not Early Low Educ	Not Early PS	Early Low Educ	Early PS
age	-0.001	0	0	-0.001
	(0)	(0)	(0)	(0)
L.1	0.358	0.384	0.333	0.302
	(0.021)	(0.014)	(0.065)	(0.05)
L.2	0.154	0.137	0.122	0.183
	(0.041)	(0.027)	(0.127)	(0.103)
L.3	0.142	0.159	0.228	0.214
	(0.043)	(0.028)	(0.138)	(0.108)
Intercept	0.059	0.039	0.046	0.079
	(0.005)	(0.003)	(0.016)	(0.012)

Note: Reports point estimates used to construct probabilities of censoring in the panel data across disability and education subgroups, accounting for lag structures. Standard errors of estimates reported in brackets below point estimates.

5 Descriptive Statistics

Table 4: Likelihood of Post-Secondary Attainment by Early Disability Status

	Data
Early-Onset	0.460
	(0.037)
Not Early Disabled	0.640
	(0.012)

Notes: Survey weights applied to LISA data to represent the of Canada population in 2012. Post-secondary education equals one if the individuals has completed any post-secondary, which includes college certificates, university degrees below a bachelors, a bachelors degree, and degrees above a bachelors. Individuals who complete high school or drop out are grouped into the low schooling category. Standard errors are reported in parenthesis below.

Table 4 shows that the likelihood of completing post-secondary is 18 percentage-points lower for early-onset individuals. Less than half of individuals affected by an early-onset disability complete a post-secondary degree.

Table 5: Employment and Earnings by Education Level and Early Disability Status.

	Not Early Disabled		Early-Onset	
	Low Education	Post-Secondary	Low Education	Post-Secondary
All Years in Labour Market				
Annual Earnings($)	32300	50900	26000	40400
	(21300)	(31600)	(19900)	(27400)
Employment Rate	0.740	0.846	0.508	0.753
First 3 years in Labour Market				
Annual Earnings ($)	15100	20700	12900	18200
	(10800)	(14300)	(10000)	(13300)
Employment Rate	0.810	0.862	0.579	0.815

Notes: Estimates are from T1FF years 1989-2016 and survey weights applied to represent the of Canada population in 2012. Standard deviations are reported in parenthesis below.

Table 5 presents statistics on lifetime earnings and employment by early disability status and education level. Individuals with early-onset disabilities and low education who are employed earn approximately 20% less than their counterparts without early-onset disabilities, increasing their risk of applying for Social Insurance (SI). These lower returns to work are reflected in significantly reduced lifetime employment rates—23 percentage points lower than those of similarly educated individuals without early disabilities. The third and fourth rows of Table 5 report average earnings and employment in the first three years following labor market entry. The observed differences by early disability status within the low education group are smaller in this early period, suggesting that early-onset disabilities may hinder the accumulation of skills over time. Among those with post-secondary education, the earnings gap by early disability status is comparable in magnitude to that observed in the low education group. Early-onset individuals with post-secondary education earn about 20% less than their non-disabled peers, potentially reflecting lower average ability, reduced financial returns to education, or both. However, their average employment rates are much closer to those of non-disabled individuals, indicating relatively higher returns to work within this subgroup.

Table 6: Average Rate and Transfer Amount From Social Assistance (SA) and Disability Insurance (DI) by Education Level and Early Disability Status

	Not Early Disabled		Early-Onset	
	Low Education	Post-Secondary	Low Education	Post-Secondary
SA Rate				
Age $<$ 45	0.0773	0.0252	0.3702	0.0772
	(0.003)	(0.001)	(0.014)	(0.006)
Age $\geq$ 45	0.0785	0.0262	0.2963	0.1309
	(0.003)	(0.001)	(0.019)	(0.013)
Average Transfer from SA				
Age $<$ 45	6100	5600	8200	6800
	(100)	(200)	(200)	(300)
Age $\geq$ 45	7200	6600	8700	6100
	(100)	(200)	(300)	(300)
All Labour Market Years				
DI Rate	0.0238	0.0085	0.0396	0.0407
	(0.001)	(0.000)	(0.005)	(0.004)
Average Transfer from DI	9100	9300	7600	7800
	(100)	(100)	(200)	(200)

Notes: Estimates are from T1FF years 1989-2016 and survey weights applied to represent the of Canada population in 2012. Standard deviations are reported in parenthesis below.

Table 6 presents statistics on the likelihood of receiving transfers and the average benefit amounts from Disability Insurance (DI) and Social Assistance (SA), disaggregated by early disability status and education level. The first two rows indicate that individuals with early-onset disabilities are substantially more likely to receive SA benefits early in life and, on average, receive larger transfers. Across all education levels, the proportion of individuals who ever become SA recipients is more than double for the early-onset group. Notably, over 30% of early-onset individuals with low education depend on SA at some point during their lives. Rows 3 and 4 show that the difference in average SA benefits received between early-onset and non-disabled individuals decreases with education. Early-onset individuals with low education receive approximately $2,000 more per year in SA benefits, compared to a difference of around $800 for those with post-secondary education

Rows 5 and 6 report that the likelihood of receiving Canada Pension Plan Disability (CPP-D) benefits is relatively low, with approximately 4% of early-onset individuals eventually becoming beneficiaries. It is important to interpret this figure as representing only those who both applied for and were accepted into the CPP-D program. In practice, many more individuals may apply but are denied; for instance, in 2014–2015, only 43% of CPP-D applications were approved (Office of the Auditor General of Canada, 2015). Lastly, the average size of DI benefits increases with age, reflecting growth in lifetime earnings.

6 Model Policy Environment

6.1 Tax Environment

Parameters for the income tax brackets and marginal tax rates were derived from the Canadian Tax and Transfer Simulator (Milligan, 2016). In each province and calendar year, I cap the upper threshold to tax brackets to give me 5 distinct tax brackets. I then calculate the economy's average income brackets and marginal tax rates across all years and provinces in the support of my data. I each province-year tax regime based on the joint density of calendar year and province in my sample. Table 7 reports the resulting tax system used in the model.

Table 7: Tax Brackets and Marginal Tax Rates.

Income Bracket	Tax Rate
[0, 30805]	0.2280
[30805, 46586]	0.2944
[46586, 64178]	0.3433
[64178, 68066]	0.3621
[68066, ∞]	0.3833

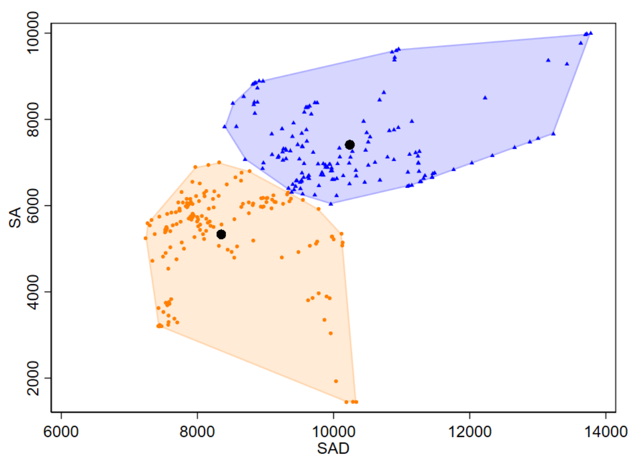
Shows the tax schedule implemented in the model based on bracketed income thresholds and associated rates. Tax rates and income brackets derived using parameters of the Canadian Tax and Transfer Simulator (Milligan, 2016).

6.2 Social Assistance Regimes

In Canada, SA policies vary across provinces and over calendar time. For each province and each time period, I represent the SA policy as a two-element "couplet" showing the maximum benefit available under SA and under SA-D. Given 10 provinces observed over 29 periods, this yields 290 distinct couplets. Accommodating three hundred different SA policies is computationally intractable. To simplify, I group "similar" couplets using a k-means clustering algorithm (Hartigan–Wong), which clusters observations based on Euclidean distances (Hartigan and Wong, 1979). The algorithm partitions the 290 couplets into clusters by minimizing the sum of squared distances between points and their assigned cluster centers.

Hartigan–Wong algorithm proceeds by trying to place each data point into the "best" cluster, which loosely translates to minimizing the overall within-cluster variance (the total euclidean distance of points to their cluster centers). Given a set of data points (290 two-dimensional "couplets" in my application). I also decide on a number of clusters k=2. The algorithm begins by assigning each point to one of k clusters in some initial way (often randomly). After the initial grouping, the algorithm checks whether moving each point it from its current cluster to a different cluster would reduce the overall distance within all clusters. If it finds that moving a point to a different cluster yields a lower overall sum of squared distances, it makes that move. Each time a point is reassigned, the center (mean) of both the old cluster and the new cluster is updated to reflect the change. The algorithm continues through the points and reassigning them whenever a beneficial move is found. Once no further improvements can be found (in terms of reducing overall distance), the algorithm has converged.

Figure 1: SA Regimes and Clusters by Province and Year



Note: Graph illustrates the k-means clustering of social assistance policies across province-time pairs. The generosity of regular SA-D is reported on the horizontal axis and generosity of SA on the vertical axis. Each point represents a province-year SA regime. Regimes are grouped into low generosity (circles) and high generosity (triangles) regimes.

Figure 1 illustrates these clusters: each point corresponds to a province–time couplet, shaded regions show which couplets are grouped together, and each black dot marks the cluster center (a weighted average of its members). The cluster centers are the SA regimes used in the model. I choose two clusters: one representing less generous SA policies and another representing more generous ones.

7 Estimation of search frictions

To calculate job arrival rates, I first estimate parameters estimates from the following probit model:

$$UE_i = \gamma_0^{s,d_0} + \gamma_1^{s,d_0} LS_i + \gamma_2^{s,d_0} age_i + LS_i * age_i \gamma_3^{s,d_0} + \epsilon_i, \tag{21}$$

where the dependent variable, UE_i is an indicator equal to one if an i is employed (part-time or full-time). The variable LS_i indicates whether the individual was actively searching for a job in the previous month. Probit regressions are estimated separately by schooling level (s) and early-onset disability status (d_0). Using the estimated coefficients, I calculate the marginal effect of job search on employment probability across age and convert monthly arrival rates into annual equivalents.

To calculate job destruction rate, I first obtain parameter estimates from the following model,

$$EU_i = \beta_0^{s,d0} + \beta_1^{s_d 0} age_i + \epsilon_i. \tag{22}$$

where the dependent variable, EU_i, equals one if the individual was fired or laid off since the last survey wave. As before, the model is estimated separately by s and d_0, and I use the resulting estimates to predict age-specific separation probabilities.

Table 8: Models for Job Arrival Rate and Destruction Rate

	d0s0	d0s1	d1s0	d1s1
LS	-0.728	-1.001	-0.619	-0.578
	(0.104)	(0.107)	(0.233)	(0.301)
age	-0.001	-0.006	-0.006	-0.011
	(0)	(0)	(0.001)	(0.001)
LS*age	-0.014	-0.012	-0.017	-0.02
	(0.003)	(0.002)	(0.009)	(0.008)
Intercept	0.082	0.507	-0.128	0.527
	(0.01)	(0.01)	(0.026)	(0.033)
age	-0.006	-0.011	-0.031	-0.016
	(0.003)	(0.003)	(0.008)	(0.009)
Intercept	-0.958	-0.984	-0.129	-0.514
	(0.137)	(0.12)	(0.277)	(0.358)

Note: Table provides estimated coefficients from probit models of job arrival and separation, by education level and disability status. Standard errors are reported in brackets below point estimates.

Table 8 presents the estimates from estimating (17) and (18). The estimation results indicate that individuals with post-secondary education ($s = 1$) receive job offers at a higher rate. Conditional on schooling, individuals with early-onset disabilities are less likely to receive job offers—consistent with employer perceptions of lower productivity, higher accommodation costs, or bias against hiring individuals with disabilities (Dixon et al., 2003). In contrast, job separation rates are lower for individuals with higher education and higher for those with early-onset disabilities, conditional on education. This aligns with the interpretation that more stable, permanent jobs are available to individuals with post-secondary credentials.

8 Description of Decomposition

I follow the framework of Shorrocks et al. (2013) in decomposing the education gap into its contributing components. Consider a statistical indicator, I, which can be fully expressed as a function of m contributory factors,

$$I = f(X_1, ..., X_m). \tag{23}$$

In this application, I is the education differential between individuals with and without an early onset disability. The contributing factors, X_m, are sets of structural parameters that differ by initial disability status, and the function f is the mapping from the structural model to the education differential. Let $F(S)$ be the value of I when a set of factors $X_k, k \notin S$, have been shut off. A decomposition of the model structure $\{K, F\}$ is defined as a set of real values $C_k, k \in K$, representing the contribution of each factor. That is, the contribution of a factor corresponds to the change in I when that factor was shut off. A decomposition rule is a function that generates these factor contributions:

$$C_k = C_k(K, F) \tag{24}$$

The first decomposition I implement calculates the marginal impact on the education gap when shutting down a single factor, with all other factors on. This is given by:

$$C_k(K, F) = F(K) - F(K/\{k\}), k \in K \tag{25}$$

This decomposition represents the *ceteris paribus* effect of each contributing factor on the gap, holding all other model features constant. However, the individual contributions derived from this method do not, in general, sum to reproduce the entire baseline education gap.

As an alternative, I employ the Shapley decomposition, which ensures that the sum of the factor contributions equals the total baseline gap. The Shapley decomposition is calculated based on calculating the marginal impact of each factor across all $m!$ possible ordered sequences in which the factors could be eliminated:

$$C_j = \sum_{k-0}^{n-1} \frac{(n-k-1)!k!}{n!} \left(\sum_{s \subset S_k/\{X_j\}:|s|=k} \left[f(s \cup X_j) - f(s) \right] \right) \tag{26}$$

where n is the total number of arguments in the original function, and $S_k/\{X_j\}$ is the set of all "*submodels*" of size k that exclude factor X_j. The weighting term, $\frac{(n-k-1)!k!}{n!}$ reflects the probability that a particular submodel of size k is randomly selected under uniform permutation.

The Shapley decomposition has three desirable properties. First, it is exact—the contributions of all factors sum to match the total education gap. Second, it satisfies symmetry: if two factors have identical marginal effects across all permutations, their contributions will be equal. This property ensures path independence in the estimation of each factor's contribution.[4] Third, the method accommodates hierarchical structures, allowing for decomposition into both primary and secondary contributing factors.

[4]In contrast, sequential shutdown methods can be path-dependent.

9 Auxiliary Model and Fit of Moments

Tables 9 - 19 display the full set of auxiliary moments used in estimation. Each table reports the moments calculated in the observed data, the moments calculated using data simulated with the model, and the standard error of the observed data moment. Estimation consists of 217 moments, including education rates and regressions (Table 9), coefficients from DI rate and DI flow regressions (Tables 10, 11, and 12), employment rates, flows, and regressions (Tables 13 and 14), and earnings distributions and regressions (Tables 15, 16, 17, 18, and and 19).

Table 9: Education Rates and Regressions

Moment	Data	Simulation	Standard Error .
	Education Rate		
$Frac(s=1\|d_0=1)$	0.466	0.457	0.037
$Frac(s=1\|d_0=0)$	0.658	0.638	0.012
	Linear Probability Models		
d_0	-0.125	-0.094	0.033
$\hat{v}$	0.146	0.254	0.015
Intercept	-0.673	-1.736	0.143
σ^2_ψ	0.214	0.220	0.003
d_0	-0.894	1.058	0.384
$\hat{v}$	0.136	0.336	0.016
$pr \times \hat{v}$	0.085	-0.125	0.042
Intercept	-0.584	-2.499	0.154
σ^2_ψ	0.213	0.219	0.003
	Conditional Linear Probability Models		
Conditional on $d_0=0$			
pr	-0.886	-0.215	0.307
$\hat{v}$	0.098	0.326	0.022
$pr \times \hat{v}$	0.094	0.014	0.033
Intercept	-0.230	-2.359	0.201
σ^2_ψ	0.211	0.206	0.004
Conditional on $d_0=1$			
pr	-1.131	-0.183	0.737
$\hat{v}$	0.169	0.205	0.049
$pr \times \hat{v}$	0.121	0.009	0.081
Intercept	-0.989	-1.330	0.450
σ^2_ψ	0.224	0.228	0.009

Table 10: OLS Regression Coefficients: DI Rate

Moment	Data	Simulation	Standard Error
	$d_0 = 0, s = 0$		
age	0.007	0.011	0.002
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	-0.063	-0.105	0.027
	$d_0 = 0, s = 1$		
age	0.006	0.005	0.002
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	-0.068	-0.053	0.018
	$d_0 = 1, s = 0$		
age	0.021	0.016	0.014
age^2	-0.001	-0.001	0.000
age^3	0.000	0.000	0.000
Intercept	-0.193	-0.141	0.153
	$d_0 = 1, s = 1$		
age	-0.003	0.006	0.014
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	0.038	-0.045	0.167

Table 11: OLS Regression Coefficients: DI Flow

Moment	Data	Simulation	Standard Error
	$d_0 = 0, s = 0$		
age	-0.002	0.000	0.001
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	0.027	0.007	0.010
	$d_0 = 0, s = 1$		
age	0.001	0.001	0.001
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	-0.007	-0.013	0.008
	$d_0 = 1, s = 0$		
age	-0.004	-0.008	0.004
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	0.045	0.106	0.042
	$d_0 = 1, s = 1$		
age	0.005	-0.006	0.006
age^2	0.000	0.000	0.000
age^3	0.000	0.000	0.000
Intercept	-0.058	0.069	0.070

Table 12: OLS Regression Coefficients: Pre-DI Ln(Average Earnings) and Employment

Moment	Data	Simulation	Standard Error
	Dependent Variable : *Ln(Average Earnings)*		
d_0	-0.119	-0.216	0.300
s	0.209	-0.038	0.106
$s \times d_0$	-0.316	0.348	0.390
intercept	10.141	9.795	0.074
	Dependent Variable: Average Employment		
d_0	-0.280	-0.407	0.173
s	0.104	0.169	0.059
$s \times d_0$	0.034	0.153	0.213
intercept	0.706	0.583	0.044

Notes: dependent Variables are calculated as the average over the 5 periods prior to applying for DI.

Table 13: Conditional Employment Rates and Flows

Moment	Data	Simulation	Standard Error
Employment Rates			
$Fr(L_{it}=1\|d_0=0,d^*_{it}=0,s_i=0,t<45)$	0.874	0.785	0.003
$Fr(L_{it}=1\|d_0=0,d^*_{it}=0,s_i=0,t\geq 45)$	0.797	0.773	0.004
$Fr(L_{it}=1\|d_0=0,d^*_{it}=0,s_i=1,t<45)$	0.908	0.879	0.002
$Fr(L_{it}=1\|d_0=0,d^*_{it}=0,s_i=1,t\geq 45)$	0.850	0.892	0.003
$Fr(L_{it}=1\|d_0=0,d^*_{it}=1,s_i=0,t<45)$	0.670	0.819	0.015
$Fr(L_{it}=1\|d_0=0,d^*_{it}=1,s_i=0,t\geq 45)$	0.479	0.678	0.009
$Fr(L_{it}=1\|d_0=0,d^*_{it}=1,s_i=1,t<45)$	0.831	0.892	0.008
$Fr(L_{it}=1\|d_0=0,d^*_{it}=1,s_i=1,t\geq 45)$	0.638	0.863	0.007
$Fr(L_{it}=1\|d_0=1,d^*_{it}=1,s_i=0,t<45)$	0.521	0.507	0.014
$Fr(L_{it}=1\|d_0=1,d^*_{it}=1,s_i=0,t\geq 45)$	0.480	0.425	0.020
$Fr(L_{it}=1\|d_0=1,d^*_{it}=1,s_i=1,t<45)$	0.815	0.754	0.009
$Fr(L_{it}=1\|d_0=1,d^*_{it}=1,s_i=1,t\geq 45)$	0.611	0.655	0.018
Employment Transition Rates			
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=0,s_i=0,t<45)$	0.046	0.123	0.002
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=0,s_i=0,t\geq 45)$	0.040	0.089	0.002
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=0,s_i=1,t<45)$	0.037	0.076	0.001
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=0,s_i=1,t\geq 45)$	0.039	0.060	0.001
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=0,s_i=0,t<45)$	0.050	0.132	0.002
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=0,s_i=0,t\geq 45)$	0.028	0.074	0.002
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=0,s_i=1,t<45)$	0.047	0.113	0.001
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=0,s_i=1,t\geq 45)$	0.024	0.053	0.001
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=1,s_i=0,t<45)$	0.080	0.139	0.007
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=1,s_i=0,t\geq 45)$	0.029	0.049	0.006
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=1,s_i=1,t<45)$	0.050	0.127	0.005
$Fr(L_{it}=0\|L_{it}=1,d_{i0}=1,s_i=1,t\geq 45)$	0.045	0.081	0.008
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=1,s_i=0,t<45)$	0.067	0.150	0.007
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=1,s_i=0,t\geq 45)$	0.019	0.035	0.005
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=1,s_i=1,t<45)$	0.058	0.162	0.005
$Fr(L_{it}=1\|L_{it}=0,d_{i0}=1,s_i=1,t\geq 45)$	0.027	0.061	0.005
Employment Rate at Labour Market Entry			
$Fr(L_{it}=1\|d_{i0}=0,s_i=0,t=1)$	0.809	0.648	0.009
$Fr(L_{it}=1\|d_{i0}=0,s_i=1,t=4)$	0.862	0.822	0.006
$Fr(L_{it}=1\|d_{i0}=1,s_i=0,t=1)$	0.579	0.363	0.028
$Fr(L_{it}=1\|d_{i0}=1,s_i=1,t=4)$	0.815	0.552	0.022

Table 14: OLS Regression Coefficients: Employment

Moment	Data	Simulation	Standard Error
		$d_0 = 0, s = 0$	
age	-0.034	0.040	0.007
$age^2/100$	0.001	0.000	0.000
$age^3/100$	0.000	0.000	0.000
d^*	-0.263	-0.019	0.009
pr	-0.026	-0.073	0.005
intercept	1.157	0.107	0.084
		$d_0 = 0, s = 1$	
age	-0.077	0.001	0.006
$age^2/100$	0.002	0.000	0.000
$age^3/100$	0.000	0.000	0.000
d^*	-0.140	-0.006	0.006
pr	-0.018	-0.025	0.003
intercept	1.723	0.750	0.074
		$d_0 = 1, s = 0$	
age	-0.043	0.042	0.029
$age^2/100$	0.001	0.000	0.001
$age^3/100$	0.000	0.000	0.000
pr	-0.130	-0.080	0.022
intercept	1.205	-0.261	0.338
		$d_0 = 1, s = 1$	
age	0.043	0.034	0.031
$age^2/100$	-0.001	0.000	0.001
$age^3/100$	0.000	0.000	0.000
pr	-0.109	-0.033	0.016
intercept	0.338	-0.137	0.383

Table 15: Annual Earnings Distribution

Moment	Data	Simulation	Standard Error
	Mean of Annual Earnings Over All Years		
$E(W\|d_0 = 0, s = 0)$	32300.000	36151.366	100.000
$E(W\|d_0 = 0, s = 1)$	50900.000	51881.700	100.000
$E(W\|d_0 = 1, s = 0)$	26000.000	29658.624	500.000
$E(W\|d_0 = 1, s = 1)$	40400.000	47926.220	600.000
	Mean of Annual Earnings in First Three Years of Labour Market		
$E(LnW\|s = 0, d_0 = 0, 1 \leq t \leq 3)$	9.390	9.406	0.018
$E(LnW\|s = 1, d_0 = 0, 4 \leq t \leq 6)$	9.700	9.712	0.012
$E(LnW\|s = 0, d_0 = 1, 1 \leq t \leq 3)$	9.200	9.167	0.052
$E(LnW\|s = 1, d_0 = 1, 4 \leq t \leq 6)$	9.530	9.606	0.047
	Variance of Initial Earnings		
$Var(LnW\|s = 0, d_0 = 0, 1 \leq t \leq 3)$	0.486	0.032	0.015
$Var(LnW\|s = 1, d_0 = 0, 4 \leq t \leq 6)$	0.526	0.033	0.010
$Var(LnW\|s = 0, d_0 = 1, 1 \leq t \leq 3)$	0.532	0.316	0.036
$Var(LnW\|s = 1, d_0 = 1, 4 \leq t \leq 6)$	0.618	0.224	0.039

Table 16: Annual Earnings Quantiles

Moment	Data	Simulation	Standard Error
		$d_0 = 0, s = 0$	
Q10	9.148	9.518	0.014
Q25	9.857	9.833	0.009
Q50	10.389	10.253	0.005
Q75	10.771	10.708	0.004
Q90	11.060	11.126	0.004
		$d_0 = 0, s = 1$	
Q10	9.525	9.805	0.011
Q25	10.240	10.185	0.006
Q50	10.751	10.652	0.003
Q75	11.126	11.115	0.003
Q90	11.416	11.509	0.004
		$d_0 = 1, s = 0$	
Q10	8.556	9.193	0.072
Q25	9.278	9.638	0.044
Q50	9.971	10.035	0.030
Q75	10.454	10.488	0.022
Q90	10.919	10.950	0.032
		$d_0 = 1, s = 1$	
Q10	9.127	9.751	0.062
Q25	9.868	10.103	0.033
Q50	10.494	10.552	0.023
Q75	10.910	11.034	0.017
Q90	11.242	11.451	0.018

Table 17: OLS Regression Coefficients: Annual Earnings

Moment	Data	Simulation	Standard Error
		$d_0 = 0, s = 0$	
age	0.123	0.155	0.003
$age^2/100$	-0.001	-0.002	0.000
d^*	-0.129	0.006	0.017
Intercept	7.625	7.051	0.051
		$d_0 = 0, s = 1$	
age	0.172	0.214	0.002
$age^2/100$	-0.002	-0.002	0.000
d^*	-0.120	-0.001	0.011
Intercept	6.859	6.033	0.044
		$d_0 = 1, s = 0$	
age	0.096	0.086	0.011
$age^2/100$	-0.001	-0.001	0.000
Intercept	7.713	7.858	0.197
		$d_0 = 1, s = 1$	
age	0.190	0.197	0.012
$age^2/100$	-0.002	-0.002	0.000
Intercept	6.331	6.254	0.221

Table 18: First-Difference Regression on Annual Earnings

Moment	Data	Simulation	Standard Error
	$d_0 = 0, s = 0$		
potential experience	0.110	0.112	0.006
potential experience^2/100	-0.204	-0.213	0.012
$E(v\|s = 0, d_0 = 0)$	9.037	9.156	0.017
$Var(v\|s = 0, d_0 = 0)$	0.345	0.192	0.014
$Var(\xi\|s = 0, d_0 = 0)$	0.209	0.099	0.004
$Cov(\epsilon_t, \epsilon_{t-1}\|s = 0, d_0 = 0)$	0.087	0.085	0.002
$Cov(\epsilon_t, \epsilon_{t-2}\|s = 0, d_0 = 0)$	0.053	0.075	0.002
	$d_0 = 0, s = 1$		
potential experience	0.145	0.126	0.004
potential experience^2/100	-0.285	-0.244	0.010
$E(v\|s = 1, d_0 = 0)$	9.211	9.447	0.013
$Var(v\|s = 1, d_0 = 0)$	0.364	0.181	0.014
$Var(\xi\|s = 1, d_0 = 0)$	0.230	0.090	0.003
$Cov(\epsilon_t, \epsilon_{t-1}\|s = 1, d_0 = 1)$	0.109	0.078	0.002
$Cov(\epsilon_t, \epsilon_{t-2}\|s = 1, d_0 = 1)$	0.067	0.068	0.001
	$d_0 = 1, s = 0$		
potential experience	0.104	0.082	0.027
potential experience^2/100	-0.197	-0.142	0.064
$E(v\|s = 0, d_0 = 1)$	8.761	8.830	0.061
$Var(v\|s = 0, d_0 = 1)$	0.471	0.398	0.053
$Var(\xi\|s = 0, d_0 = 1)$	0.255	0.095	0.017
$Cov(\epsilon_t, \epsilon_{t-1}\|s = 0, d_0 = 0)$	0.098	0.073	0.011
$Cov(\epsilon_t, \epsilon_{t-2}\|s = 0, d_0 = 0)$	0.074	0.063	0.012
	$d_0 = 1, s = 1$		
potential experience	0.133	0.144	0.018
potential experience^2/100	-0.323	-0.295	0.036
$E(v\|s = 1, , d_0 = 1)$	9.245	9.158	0.053
$Var(v\|s = 1, d_0 = 1)$	0.388	0.313	0.044
$Var(\xi\|s = 1, d_0 = 1)$	0.263	0.061	0.015
$Cov(\epsilon_t, \epsilon_{t-1}\|s = 1, d_0 = 1)$	0.127	0.050	0.009
$Cov(\epsilon_t, \epsilon_{t-2}\|s = 1, d_0 = 1)$	0.075	0.044	0.008

Table 19: Fixed Effect Quantiles

Moment	Data	Simulation	Standard Error
		$d_0 = 0, s = 0$	
Q10	8.182	8.609	0.025
Q25	8.655	8.834	0.016
Q50	9.094	9.116	0.014
Q75	9.457	9.446	0.012
Q90	9.776	9.756	0.016
		$d_0 = 0, s = 1$	
Q10	8.390	8.892	0.019
Q25	8.816	9.145	0.012
Q50	9.236	9.443	0.010
Q75	9.579	9.763	0.010
Q90	9.892	10.022	0.012
		$d_0 = 1, s = 0$	
Q10	7.953	8.033	0.084
Q25	8.396	8.423	0.049
Q50	8.825	8.816	0.045
Q75	9.329	9.232	0.054
Q90	9.716	9.648	0.049
		$d_0 = 1, s = 1$	
Q10	8.386	8.438	0.085
Q25	8.861	8.760	0.057
Q50	9.292	9.149	0.059
Q75	9.700	9.565	0.043
Q90	10.002	9.894	0.080

References

Baker, M., Stabile, M., and Deri, C. (2004). What do self-reported, objective, measures of health measure? *Journal of human Resources*, 39(4):1067–1093.

Benítez-Silva, H., Buchinsky, M., Man Chan, H., Cheidvasser, S., and Rust, J. (2004). How large is the bias in self-reported disability? *Journal of Applied Econometrics*, 19(6):649–670.

Black, N., Johnston, D. W., Shields, M. A., and Suziedelyte, A. (2017). Who provides inconsistent reports of their health status? the importance of age, cognitive ability and socioeconomic status. *Social Science and Medicine*, 191:9–18.

Bound, J. (1989). The health and earnings of rejected disability insurance applicants. *The American Economic Review*, pages 482–503.

Bound, J. and Burkhauser, R. V. (1999). Economic analysis of transfer programs targeted on people with disabilities. *Handbook of Labor Economics*, 3:3417–3528.

Bruins, M., Duffy, J. A., Keane, M. P., and Smith Jr, A. A. (2018). Generalized indirect inference for discrete choice models. *Journal of Econometrics*, 205(1):177–203.

Burkhauser, R. V., Daly, M. C., Houtenville, A. J., and Nargis, N. (2002). Self-reported work-limitation data: What they can and cannot tell us. *Demography*, 39(3):541–555.

Dixon, K. A., Kruse, D., and Van Horn, C. E. (2003). *Restricted access: A survey of employers about people with disabilities and lowering barriers to work.* John J. Heldrich Center for Workforce Development, Rutgers.

Grondin, C. (2016). *A new survey measure of disability: The Disability Screening Questions (DSQ).* Statistics Canada.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

Keane, M. and Smith, A. A. (2003). Generalized indirect inference for discrete choice models. *Yale University*.

Milligan, K. (2016). Canadian tax and credit simulator. Database, software and documentation, Version 2016-2.

Office of the Auditor General of Canada (2015). Report 6-canada pension plan disability program. https://www.oag-bvg.gc.ca/internet/english/parl_oag_201602_06_e_41063.html.

Shorrocks, A. F. et al. (2013). Decomposition procedures for distributional analysis: a unified framework based on the shapley value. *Journal of Economic Inequality*, 11(1):99–126.

Stern, S. (1989). Measuring the effect of disability on labor force participation. *Journal of Human Resources*, pages 361–395.